UNIVERSITY OF CALIFORNIA SAN DIEGO

Online Learning and Decision Making with Partial Information, a feedback perspective

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Electrical Engineering
(Machine Learning and Data Science)

by

Anshuka Rangi

Committee in charge:

Professor Massimo Franceschetti, Chair
Professor Sanjoy Dasgupta
Professor Alon Orlitsky
Professor Piya Pal
Professor Behrouz Touri

2021

The Dissertation of Anshuka Rangi is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To my parents

TABLE OF CONTENTS

Dissertation Approval Page iii

Dedication iv

Table of Contents v

List of Figures x

List of Tables xii

Acknowledgements xiii

Vita xvii

Abstract of the Dissertation xix

Chapter 1 Introduction 1
1.1 Introduction 1
1.1.1 Known dynamics and Unknown state 2
1.1.2 Unknown dynamics and Unknown state 3
1.1.3 Attacks and Security of Online systems 4
1.1.4 Non-stochastic Information Theory 6
1.2 Dissertation Overview 6

Chapter 2 Distributed Chernoff Test: Optimal Decision Systems over Network 12
2.1 Introduction 12
2.2 Related Work 14
2.3 Problem Formulation 16
2.3.1 Hypotheses Testing model 16
2.3.2 Network model 17
2.3.3 Communication model 17
2.3.4 Performance measure 18
2.3.5 Additional notation 19
2.4 Standard Chernoff Test 19
2.5 Decentralized Chernoff Test 21
2.5.1 Informal Discussion of DCT 23
2.6 Consensus-Based Chernoff Test 23
2.6.1 Informal Discussion of CCT 29
2.7 Performance analysis 29
2.7.1 Lower Bounds for a Sequential and an Adaptive test 29
2.7.2 Upper bounds for proposed DCT and CCT schemes 30
2.8 Numerical Results 36
2.9 Extension to Channels with Quantized Messages and Link Failures 40

2.9.1 Channels with Quantized Messages 41
2.9.2 Performance analysis for Channels with Quantized Messages 43
2.9.3 Channels with Quantized Messages and Erasures 46
2.9.4 Performance Analysis for Channels with Quantized Messages and Erasures 49
2.10 Conclusion 53
2.11 Acknowledgement 54
2.12 Appendix 55
2.12.1 Proof of Theorem 2 55
2.12.2 Proofs for DCT and CCT 58
2.12.3 Proofs of Miscellaneous Results 80

Chapter 3 Bounded Knapsack Bandits in crowdsourcing systems 89
3.1 Introduction 89
3.2 Problem Formulation 91
3.3 Related Work 94
3.4 Workers' Selection 96
3.5 Value Contributions of Workers 100
3.6 Performance Evaluation 103
3.7 Conclusion 108
3.8 Acknowledgment 109
3.9 Appendix 110
3.9.1 Proof of Theorem 11 110
3.9.2 Proof of Theorem 12 116
3.9.3 Proof of Theorem 13 117

Chapter 4 Unifying the Stochastic and the Adversarial Knapsack Bandits 119
4.1 Introduction 119
4.2 Contribution 121
4.2.1 Related Work 122
4.3 Problem Formulation 123
4.4 Adversarial BwK 125
4.5 One practical algorithm for both stochastic and adversarial BwK 129
4.6 BwK with unbounded cost 134
4.7 Conclusion 135
4.8 Acknowledgment 136
4.9 Appendix 137
4.9.1 Proof of Theorem 1 137
4.9.2 Proof of Theorem 3 141
4.9.3 Proof of Theorem 4 154

Chapter 5 Online learning with Feedback Graphs and Switching Costs 156
5.1 Introduction 156
5.1.1 Contributions 158
5.1.2 Related Work 159

5.2 Problem Formulation 160
5.3 Lower Bound in PI setting with SC 161
5.4 Algorithms in PI setting with SC 165
5.5 Performance Evaluation 171
5.6 Conclusion 173
5.7 Acknowledgement 174
5.8 Appendix 174
5.8.1 Proof of Theorem 1 174
5.8.2 Proof of Lemma 2 181
5.8.3 Proof of Theorem 3 182
5.8.4 Proof of Lemma 4 182
5.8.5 Proof of Theorem 5 183
5.8.6 Proof of Theorem 6 187

Chapter 6 Attacks and Security of Multi-Armed Bandits 190
6.1 Introduction 190
6.2 Contributions 192
6.3 Related Work 193
6.4 Preliminaries and Problem Statement 195
6.4.1 Poisoning Attacks on Stochastic Bandits 195
6.4.2 Remedy via Limited Reward Verification 196
6.5 Tight Characterization for the Cost of Poisoning Attack on Stochastic Bandits . 197
6.5.1 Upper Bound on the Contaminations 197
6.5.2 Matching Lower Bound on the Contaminations 199
6.6 Secure Upper Confidence Bound 202
6.7 Comparison of Attacker Models 208
6.8 Simulation Results 210
6.9 Conclusion 211
6.10 Acknowledgement 212
6.11 Appendix 213
6.11.1 Proof of Proposition 1 213
6.11.2 Attacks Based on Gap Estimation 214
6.11.3 Proof of Theorem 23 217
6.11.4 Proof of Corollary 23.1 223
6.11.5 Proof of Theorem 24 223
6.11.6 Proof of Theorem 25 236
6.11.7 Proof of Corollary 25.1 237

Chapter 7 Attacks in Episodic Reinforcement Learning 239
7.1 Introduction 239
7.1.1 Contribution 241
7.1.2 Related Work 242
7.2 Problem Formulation 243
7.3 Reward Poisoning Attacks in Unbounded Reward Setting 246

7.3.1 White-Box Attacks:a Warm-up . . . 246
7.3.2 Black-Box Attack: the more realistic setting . . . 248
7.4 Attacks in Bounded Reward Setting . . . 251
7.4.1 Insufficiency of (Only) Reward or Action Manipulation . . . 251
7.4.2 Efficient Attack by Combining Reward & Action Manipulation . . . 253
7.5 Conclusion and Future Directions . . . 255
7.6 Acknowledgement . . . 255
7.7 Appendix . . . 256
7.7.1 Proof of Theorem 27 . . . 256
7.7.2 Proof of Theorem 28 . . . 258
7.7.3 Proof of Theorem 29 . . . 263
7.7.4 White-box Attack in Bounded Reward Setting . . . 266
7.7.5 Proof of Theorem 30 . . . 270

Chapter 8 Learning-based attacks in Cyber-Physical Systems . . . 273
8.1 Introduction . . . 273
8.2 Related Work . . . 275
8.3 Problem Setup . . . 276
8.4 Learning based Attacks . . . 277
8.4.1 Performance Measures . . . 279
8.4.2 Main results . . . 280
8.5 Simulations . . . 285
8.6 Conclusions and Future Directions . . . 288
8.7 Acknowledgement . . . 288
8.8 Appendix . . . 289
8.8.1 Proof of Proposition 3 . . . 289
8.8.2 Proof of the Theorem 32 . . . 290
8.8.3 Proof of the Theorem 33 . . . 293
8.8.4 Proof of Theorem 34 . . . 294
8.8.5 Proof of Theorem 35 . . . 295
8.8.6 Proof of Theorem 36 . . . 296

Chapter 9 Non-Stochastic Information Theory . . . 297
9.1 Introduction . . . 297
9.2 Contributions . . . 300
9.3 Uncertain variables . . . 301
9.4 δ-Mutual information . . . 302
9.4.1 Uncertainty function . . . 302
9.4.2 Association and dissociation between UVs . . . 303
9.4.3 δ-mutual information . . . 308
9.5 (ϵ,δ)-Capacity . . . 316
9.6 (N, δ)-Capacity of General Channels . . . 328
9.7 Capacity of Stationary Memoryless Uncertain Channels . . . 331
9.7.1 Factorization of the Mutual Information . . . 339

9.7.2 Single letter expressions 346
9.8 Examples 358
9.8.1 Discussion 366
9.9 Applications 366
9.9.1 Error Correction in Adversarial Channels 367
9.9.2 Robustness of Neural Networks to Adversarial Attacks 371
9.9.3 Performance of Classification Systems 373
9.10 Conclusion 374
9.11 Acknowledgement 375
9.12 Appendix 375
9.12.1 Proof of Lemma 31 375
9.12.2 Proof of Theorem 45 377
9.12.3 Proof of Lemma 32 382
9.12.4 Auxiliary Results 388
9.12.5 Proof of 4 claims in Theorem 48 394
9.12.6 Taxicab symmetry of the mutual information 404

Bibliography 408

LIST OF FIGURES

Figure 1.1. Online decision making and learning systems 2

Figure 1.2. Online decision making and learning systems 4

Figure 2.1. Performance of DCT: risk vs. cost c for different number of sensors L . . . 36

Figure 2.2. Performance of DCT according to Theorem 3: risk vs. cost c for different number of sensors L 37

Figure 2.3. An example of sensor network with $L = 10$ nodes. 38

Figure 2.4. Performance of CCT for the ring with random attachments: risk vs. cost c for different number of sensors L 38

Figure 2.5. Performance of CCT according to Theorem 5 for the ring with random attachments: risk vs. cost c for different number of sensors L 38

Figure 2.6. Performance of CCT for the tree: risk vs. cost c for different number of sensors L 39

Figure 2.7. Performance of CCT according to Theorem 5 for the tree: risk vs. cost c for different number of sensors L 40

Figure 3.1. The first and second column of plots are corresponding to the classification error ϵ and number of tasks T performed by the workers respectively. a) T=50 and Set A workers b) T=50 and Set B workers c) T=100 and Set A workers 105

Figure 3.2. The first and second column of plots are corresponding to the classification error ϵ and number of tasks T performed by the workers respectively. a) T=50 and Set A workers b) T=50 and Set B workers c) T=100 and Set A workers 106

Figure 5.1. Performance evaluation of EXP3 SET and Threshold based EXP3 for K=25 172

Figure 5.2. Performance evaluation of Batch EXP3 and Threshold based EXP3 in MAB setting 173

Figure 6.1. Comparison between Secure-UCB, UCB and BARBAR 210

Figure 6.2. Performance of Secure-UCB 211

Figure 8.1. Exploration Phase. 277

Figure 8.2. Exploitation Phase. 277

Figure 8.3. Attacker's success rate versus L. 287

Figure 8.4. Attacker's success rate versus τ. 288

Figure 9.1. Illustration of disassociation between UVs. 304

Figure 9.2. Illustration of the possible time intervals for the walkers on the path. 306

Figure 9.3. The size of the equivocation set is inversely proportional to the amount of adversarial effort required to induce an error. 317

Figure 9.4. Illustration of the (ϵ, δ)-capacity in terms of packing ϵ-balls with maximum overlap δ. 319

Figure 9.5. Conditional ranges $[\![Y|x]\!]$ and $[\![X|y]\!]$ due to the ϵ-perturbation channel. . . 320

Figure 9.6. Uncertainty sets associated to three different codewords. Sets are not necessarily balls, they can be different across codewords, and also be composed of disconnected subsets. 329

Figure 9.7. Channel described in Example 3 360

Figure 9.8. The Hamming distance between any two overlapping codewords depends on the code parameters τn and δ_n. 369

LIST OF TABLES

Table 4.1. Contributions to the literature of BwK. 122

Table 5.1. Comparison of Threshold based EXP3 and EXP3.SC. 159

Table 7.1. Comparison of the attack cost in the episodic RL and MAB setting when the attacker has no information about the learning algorithm. 241

Table 9.1. Comparison of the Sufficient Conditions for the Existence of a Single-Letter expression 358

ACKNOWLEDGEMENTS

The principle of optimism under uncertainty has been the key to not only the technical contributions in this thesis, but also my journey as a PhD student over the past six years. However, it would not have been possible to stay optimistic, and progress in this journey without the support of several key people.

I would like to express my deepest gratitude to my advisor, Prof. Massimo Franceschetti, for his guidance throughout the duration of my PhD. He has been extremely patient with me during the initial years when I spent most of the time on my course work. He gave me complete freedom to explore areas of research that interest me, provided me feedback, and guided me through this journey. Massimo has taught me to be the critic of my own work. He has encouraged me to look for the analogous relationship between different problems, which has been instrumental in developing some key ideas during my PhD.

During my PhD, I also got an opportunity to work with professors Stefano Marano, Long Tran-Thanh and Haifeng Xu. I would like to thank them for the time they invested in me, which has helped me reach where I am today. Stefano provided me the positive reinforcement and the technical guidance that I needed during my initial years. Over the years, Long has guided me through multiple research ideas. He has taught me how to look at the simplest version of the problem and then generalize the solution. Haifeng has helped me strive for better results, and present them with simplicity and elegance. I am very grateful for all the technical and non-technical discussions I had with all of them.

I would like to express my thankfulness and appreciation to my dissertation committee members, Professor Sanjoy Dasgupta, Professor Alon Orlitsky, Professor Piya Pal and Professor Behrouz Touri for their valuable feedbacks. I would like to thank my lab-mates Vinnu Bhardwaj, Mohammad Khojasteh, Hamed Omidvar and Rohit Parasnis for sharing and discussing different ideas during the group meetings. I would also like to thank Rohit for the lunch-time discussions, and sharing interesting articles and books to read.

This brings me to acknowledging some of the most important people in my life, my

family.

This thesis would not have been possible without the love and support of my parents, Surya Parkash Rangi and Sunita Rani Rangi. My parents have been my backbone, and their contributions extend way beyond this journey. They have always been encouraging about my career decisions and life choices, have helped me chase my dreams, and had my back through difficult times. They have been my teachers for the longest duration, and have contributed to every positive change in my career so far. I am deeply indebted to my parents for their presence in every step of the way, and as the smallest token of appreciation, I dedicate this thesis to them.

My husband, Shekhar Kadyan, has been a constant source of unconditional love and support during this journey. Shekhar believed in my capabilities even when I doubted them, and has helped me endure through the ups and downs of this journey. His desire for me to excel has exceeded even my own, and he has consistently motivated me to work towards my goals. I am extremely fortunate to have him by my side during this journey, and look forward to our future together.

I am also grateful to my sister, Mimansa Rangi, for being a true friend, supporting me through thick and thin, and always being available to talk despite her busy schedule during her MBA. I really appreciate my brother, Jayant Rangi, for his love and care, providing feedback on my presentations, and helping me improve them. I am also thankful to both my grandmothers, *dadi* and *nani*, for their love, blessings and teachings, constant reminders to graduate, and patiently waiting for me to visit them while I was busy chasing my goals.

Although my parents were thousands of miles apart in India, I am extremely grateful to my uncle (or *chacha*), Jai Rangi and my aunt (or *chachi*), Lalita Rangi, who filled in the shoes of my parents at numerous moments during this journey. I am also thankful to my cousin, Himanshu Rangi, who encouraged and supported me during my PhD, and planned countless trips to San Diego for me. These people have been my family in US, and have shown unequivocal love and support, for which I am extremely grateful.

Chapter 2, in full, is a reprint of the material as it appears in Anshuka Rangi, Massimo

Franceschetti and Stefano Marano, "Distributed Chernoff Test: Optimal Decision Systems Over Networks", *IEEE Transactions on Information Theory*, vol. 67, pp. 2399 - 2425, April 2021, Anshuka Rangi, Massimo Franceschetti and Stefano Marano, "Consensus-Based Chernoff Test in Sensor Networks", *IEEE Conference on Decision and Control (CDC)*, December 2018, and Anshuka Rangi, Massimo Franceschetti and Stefano Marano, "Decentralized chernoff test in sensor networks", *IEEE International Symposium on Information Theory (ISIT)*, July 2018. The dissertation author was the primary investigator and author of these papers.

Chapter 3, in full, is a reprint of the material as it appears in Anshuka Rangi and Massimo Franceschetti, "Multi-armed bandit algorithms for crowdsourcing systems with online estimation of workers' ability", *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, July 2018. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, is a reprint of the material as it appears in Anshuka Rangi, Massimo Franceschetti and Long Tran-Thanh, "Unifying the Stochastic and the Adversarial Bandits with Knapsack", *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI)*, August 2019. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full, is a reprint of the material as it appears in Anshuka Rangi and Massimo Franceschetti, "Online learning with feedback graphs and switching costs", *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, April 2019. The dissertation author was the primary investigator and author of this paper.

Chapter 6, in part, contains material as it appears in Anshuka Rangi, Long Tran-Thanh, Haifeng Xu and Massimo Franceschetti, "Saving Stochastic Bandits from Poisoning Attacks via Limited Data Verification", *under preparation*. The dissertation author was the co-primary investigator and co-author of this paper.

Chapter 7, in part, contains material as it appears in Anshuka Rangi, Haifeng Xu, Long Tran-Thanh and Massimo Franceschetti, "Poisoning Attacks in Reinforcement Learning", *under*

preparation. The dissertation author was the co-primary investigator and co-author of this paper.

Chapter 8, in full, is a reprint of the material as it appears in Anshuka Rangi, Mohammad Javad Khojasteh and Massimo Franceschetti, "Learning-based attacks in Cyber-Physical Systems: Exploration, Detection, and Control Cost trade-offs", *Learning for Dynamics and Control*, June 2021. The dissertation author was the co-primary investigator and co-author of this paper.

Chapter 9, in part, is a reprint of the material as it appears in Anshuka Rangi and Massimo Franceschetti, "Non-stochastic Information Theory", *under preparation*, Anshuka Rangi and Massimo Franceschetti, "Towards a Non-Stochastic Information Theory", *IEEE International Symposium on Information Theory (ISIT)*, July 2019, and Anshuka Rangi and Massimo Franceschetti, "Channel Coding Theorems in Non-stochastic Information Theory", *IEEE International Symposium on Information Theory (ISIT)*, July 2021. The dissertation author was the primary investigator and author of this paper.

VITA

2009-2013 Bachelor of Technology in Electronics and Communication Engineering, Indian Institute of Technology Roorkee, India

2015-2018 Master of Science in Machine Learning and Data Science, University of California San Diego

2015-2021 Doctor of Philosophy in Machine Learning and Data Science, University of California San Diego

PUBLICATIONS

Anshuka Rangi, Massimo Franceschetti and Stefano Marano, "Distributed Chernoff Test: Optimal Decision Systems Over Networks", *IEEE Transactions on Information Theory*, vol. 67, pp. 2399 - 2425, April 2021.

Anshuka Rangi, Massimo Franceschetti and Stefano Marano, "Consensus-Based Chernoff Test in Sensor Networks", *IEEE Conference on Decision and Control (CDC)*, December 2018.

Anshuka Rangi, Massimo Franceschetti and Stefano Marano, "Decentralized chernoff test in sensor networks", *IEEE International Symposium on Information Theory (ISIT)*, July 2018.

Anshuka Rangi and Massimo Franceschetti, "Multi-armed bandit algorithms for crowdsourcing systems with online estimation of workers' ability", *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, July 2018.

Anshuka Rangi, Massimo Franceschetti and Long Tran-Thanh, "Unifying the Stochastic and the Adversarial Bandits with Knapsack", *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI)*, August 2019.

Anshuka Rangi and Massimo Franceschetti, "Online learning with feedback graphs and switching costs", *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, April 2019.

Anshuka Rangi, Long Tran-Thanh, Haifeng Xu and Massimo Franceschetti, "Saving Stochastic Bandits from Poisoning Attacks via Limited Data Verification", *under preparation*.

Anshuka Rangi, Haifeng Xu, Long Tran-Thanh and Massimo Franceschetti, "Poisoning Attacks in Reinforcement Learning", *under preparation*.

Anshuka Rangi, Mohammad Javad Khojasteh and Massimo Franceschetti, "Learning-based attacks in Cyber-Physical Systems: Exploration, Detection, and Control Cost trade-offs", *Learn-*

ing for Dynamics and Control, June 2021.

Anshuka Rangi and Massimo Franceschetti, "Towards a Non-Stochastic Information Theory", *IEEE International Symposium on Information Theory (ISIT)*, July 2019.

Anshuka Rangi and Massimo Franceschetti, "Channel Coding Theorems in Non-stochastic Information Theory", *IEEE International Symposium on Information Theory (ISIT)*, July 2021.

Anshuka Rangi and Massimo Franceschetti, "Non-stochastic Information Theory", *under preparation*.

ABSTRACT OF THE DISSERTATION

Online Learning and Decision Making with Partial Information, a feedback perspective

by

Anshuka Rangi

Doctor of Philosophy in Electrical Engineering
(Machine Learning and Data Science)

University of California San Diego, 2021

Professor Massimo Franceschetti, Chair

This dissertation considers a problem of online learning and online decision making where an agent or a group of agents aim to learn unknown parameters of interest. There are two key interacting components: agent and environment. The agent perform actions on the environment, these actions may or may not change the state of the environment, and the environment generates feedback based on the actions and its underlying state. The feedback is utilized by the agent to learn and improvise its decisions and actions, and optimize a certain objective.

In the first part of this dissertation, we consider different variants of the online learning and decision making systems. We propose optimal (or order-optimal) online learning algorithms

for these variants. We characterize the flow of information through feedback, and provide quantitative information measures that are key to optimal learning and decision making in these systems.

In the second part of this dissertation, we focus on the attacks and security of these online learning and decision making systems. Since the distributed nature of these systems is their Achilles' heel, making these systems secure requires an understanding of the regime where the systems can be attacked, as well as designing ways to mitigate these attacks. We study both of these aspects of the problem for stochastic Multi-Armed Bandits (MAB). We also study the former aspect of the problem, namely understanding the regime under which the system can be attacked, for Reinforcement Learning and Cyber Physical systems.

Finally, we lay the foundations of non-stochastic information theory. Classical information theory has little role in providing non-stochastic guarantees for online systems such as Cyber-Physical systems where occasional errors can quickly drive these systems out of control and lead to catastrophic failures. We propose a non-stochastic δ-mutual information to capture the worst case error guarantees, denoted by δ. We propose non-stochastic analogue of capacities which are studied in classical information theory. We also establish key results such as channel coding theorem and single letter characterization for the non-stochastic capacities.

Chapter 1

Introduction

1.1 Introduction

Modern learning and decision making systems such as recommendation systems, crowd-sourcing systems, and cyber physical systems are inherently *online*. These online systems are made of two key interacting components: agent and environment. In these systems, the agent perform actions on the environment, these actions may or may not change the state of the environment, and the environment generates feedback based on the actions and its underlying state. Namely, at a discrete time t, the agent performs action a_t, the state s_t of the environment evolves as $h(s_{t-1}, a_t)$, and the feedback signal $f(s_t, a_t)$ is observed by the agent (see Figure 1.1). The feedback corresponding corresponding to the actions is used by the agent to learn and improvise its decisions and actions, and optimize a certain objective. In these online systems, the agent faces uncertainity in decision making since either the state of the environment or behaviour of the environment to the agent's actions is unknown. The uncertainity in decision making leads to a well-studied *exploration* (searching the space of possible decisions) and *exploitation* (choosing the optimal decision based on the learned model) trade-off [16].

In this thesis, we will focus on different variants of these online learning and decision making systems. We propose optimal (or order-optimal) online learning algorithms for these variants. We characterize the flow of information through feedback, and provide quantitative information measures that are key to optimal learning and decision making in these systems.

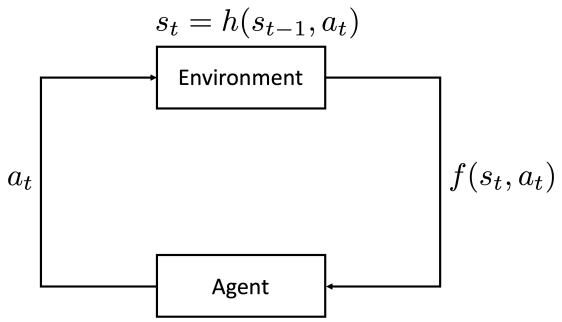


Figure 1.1. Online decision making and learning systems

These information measures capture either the expected or the worst-case behaviour of the information flow depending on the environment. We focus on the following aspects of these online systems based on the information possessed by the agent about the environment.

1.1.1 Known dynamics and Unknown state

In this setting, the state of the environment is unknown to the agent (or learner), and is fixed throughout the interaction of the agent with the environment, namely $s_t = s_{t-1} = s_0$ is fixed in Figure 1.1, and is unknown to the learner. However, the agent posses the knowledge of the dynamics (or behaviour) of the environment in each possible state, namely for all actions $a \in \mathcal{A}$ and all state $s \in \mathcal{S}$, the function $f(s, a)$ is known, where $\mathcal{A}$ and $\mathcal{S}$ denote all the possible actions and states, respectively. The objective of the agent is to identify the state of the environment among the possible states. This problem is also referred as *Hypothesis Testing* in the literature [45].

We study this problem in a network setting consisting of a group of agents connected to each other by communication links. Each agent interacts with the environment, and shares the information possessed by it with other agents over the network. The objective of the agents is to identify the state of the environment as soon as possible. More formally, the agents collectively minimize the risk, expressed by the expected cost required to reach a decision plus the expected cost of making a wrong decision.

In Chapter 2, we propose an "online" decision making scheme which is an extension of

classic Chernoff test. This scheme is asymptotically optimal in terms of risk for the networks with small diameters, and parsimonious in terms of communications in comparison to state-of-art schemes proposed in the literature. We show that the information measure characterizing the flow of information is the KL-divergence. In other words, the KL-divergence quantify the capabilities of the agents collectively to achieve their objective.

1.1.2 Unknown dynamics and Unknown state

In this setting, both the state and the dynamics of the environment are unknown to the agent, namely the current state s_t and the function $f(s_t, a_t)$ is unknown in Figure 1.1. Additionally, the state is fixed throughout the interaction of the agent with the environment, namely $s_t = s_{t-1} = s_0$. The objective of the agent is to minimize its expected regret, which is defined as the difference between the gain from the best fixed policy in the hindsight and the gain from agent's policy. This problem is commonly studied as *Multi-Armed Bandits* and *Online Learning* in the literature [16, 17].

We study this problem in a knapsack setting where the agent has an additional constraint over the resources needed to perform actions. We study this problem under two different feedback models: stochastic and adversarial. First, in the stochastic feedback model, the feedback corresponding to agent's action is drawn from a fixed underlying probability distribution function, which is unknown to the agent. Second, in the adversarial feedback model, the feedback corresponding to agent's action is assigned by an *oblivious* adversary. We propose two different order-optimal learning algorithms for both these feedback models. We also propose another novel algorithm which unifies both these feedback models by achieving almost order-optimal guarantees for both these models simultaneously. These algorithms and their regret guarantees are presented in Chapters 3 and 4.

We also study another variant of online learning where the feedback received from the environment can be modelled as a graph. In other words, the feedback from the environment provides information about the agent's action as well as a subset of actions which were not

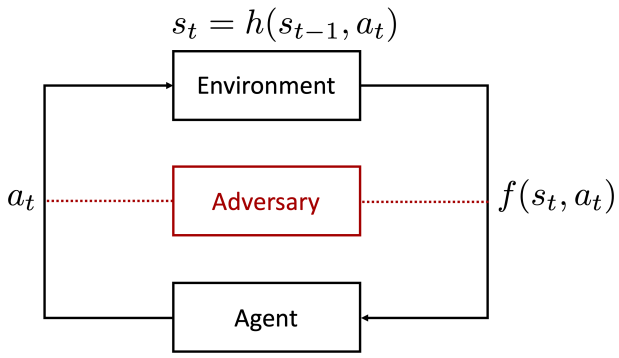


Figure 1.2. Online decision making and learning systems

performed. In the same spirit as the knapsack setting, the actions of the agent are constraint by adding a penalty (or cost) if the agent switches its action. We study this problem in the adversarial feedback setting in Chapter 5. We propose order-optimal algorithms which utilize the feedback efficiently. We introduce a new measure *independence sequence number* to characterize the flow of information in this setting .

In this section, we study all these variants for a single agent setting. However, these can also be extended to the network setting consisting of a group of agents connected to each other by communication links.

1.1.3 Attacks and Security of Online systems

The distributed nature of online learning systems is their Achilles' heel, as it is a source of vulnerability to third party attacks. For example, in web services decision making critically depends on reward collection, and this is prone to attacks that can impact observations and monitoring, delay or temper rewards, produce link failures, and generally modify or delete information through hijacking of communication links [2] [37]. Making these systems secure requires an understanding of the regime where the systems can be attacked, as well as designing ways to mitigate these attacks.

We study both of these aspects of the problem for stochastic Multi-Armed Bandits (MAB) setting in Chapter 6. We consider a data poisoning attack, also referred as man in the middle (MITM) attack. In this attack, there are three entities: the environment, the agent (or

MAB algorithm), and the attacker. In this setting, the attacker can eavesdrop and contaminate (or manipulate) the communication between the agent and the environment (see Figure 1.2). We establish the regime where any order-optimal MAB algorithm can be attacked. Formally, we provide order-optimal bounds on the amount of contamination required by the attacker to successfully attack the MAB algorithm. In the effort of developing secure ways to mitigate the attacks, we consider a *reward verification* model in which the agent can access verified (i.e. uncontaminated) rewards from the environment. This verified access can be implemented through a secure channel between the agent and the environment, or using auditing. Since verification is costly, the agent faces a tradeoff between its performance in terms of regret, and the number of times access to a verified reward occurs. Second, the agent needs to decide when to access a verified reward during the learning process. We design an order-optimal bandit algorithm which strategically plans the verification, and makes no assumptions on the attacker's strategy or capabilities.

We extend the study of the former aspect of the problem, namely understanding the regime under which the system can be attacked, from MAB setting to episodic Reinforcement Learning in Chapter 7. We study the regime where any order-optimal episodic RL algorithm can be attacked.

We also study the former aspect of the problem, namely understanding the regime under which the system can be attacked, in cyber physical systems in Chapter 8. We study learning based MITM attacks where the attacker has full access to the communication between the controller (or agent) and the plant (or environment), but the dynamics of the plant are unknown to the attacker. Thus, the attacker needs to learn about the plant in order to generate the fictitious signals to the controller that allow the attacker to remain undetected for the time needed to cause harm. On the other hand, the controller has perfect (or nearly perfect) knowledge of the dynamics of the plant and is actively looking out for an anomalous behaviour in the feedback signals from the plant. In this setting, both the attacker and the controller need to perform optimal online decision making in a feedback loop fashion. We study the trade-offs between the information acquired by the

attacker from observations, the detection capabilities of the controller, and the control cost.

1.1.4 Non-stochastic Information Theory

When Shannon laid the mathematical foundations of information theory he embraced a probabilistic approach [199]. Occasional violations from a specification are permitted, and cannot be avoided. This approach is well suited for consumer-oriented digital communication devices, where the occasional loss of data packets is not critical. In contrast, in the context of control of safety-critical online systems, error bounds must often be guaranteed at any time, not only on average. In this case, at each time step of the evolution of a dynamical system (or environment), the feedback is used by the agent to choose the next action, which is then fed back into the system. When this feedback loop is closed over a communication channel, occasional decoding errors can quickly drive the system out of control and lead to catastrophic failures. An example of such a safety critical online systems is cyber-physical systems (CPS) [115]. In this setting, classical information theory has little role in providing non-stochastic guarantees of meeting the control objectives. On the other hand, information in some sense *must be flowing across the network*, and this observation motivates the need for a meaningful theory of information in a non-stochastic setting.

In Chapter 8, we lay the foundations of non-stochastic information theory. We propose a non-stochastic δ-mutual information to capture the worst case error guarantees, denoted by δ. We propose non-stochastic analogue of capacities which are studied in classical information theory. We also establish key results such as channel coding theorem and single letter characterization for the non-stochastic capacities. Few applications of this work is discussed at length in Chapter 9.

1.2 Dissertation Overview

The rest of the dissertation is organized as follows.

In Chapter 2, we study "active" decision making over sensor networks where the sensors'

sequential probing actions are actively chosen by continuously learning from past observations. We consider two network settings: with and without central coordination. In the first case, the network nodes interact with each other through a central entity, which plays the role of a fusion center. In the second case, the network nodes interact in a fully distributed fashion. In both of these scenarios, we propose sequential and adaptive hypothesis tests extending the classic Chernoff test. We compare the performance of the proposed tests to the optimal sequential test. In the presence of a fusion center, our test achieves the same asymptotic optimality of the Chernoff test, minimizing the risk, expressed by the expected cost required to reach a decision plus the expected cost of making a wrong decision, when the observation cost per unit time tends to zero. The test is also asymptotically optimal in the higher moments of the time required to reach a decision. Additionally, the test is parsimonious in terms of communications, and the expected number of channel uses per network node tends to a small constant. In the distributed setup, our test achieves the same asymptotic optimality of Chernoff's test, up to a multiplicative constant in terms of both risk and the higher moments of the decision time. Additionally, the test is parsimonious in terms of communications in comparison to state-of-the-art schemes proposed in the literature. The analysis of these tests is also extended to account for message quantization and communication over channels with random erasures.

In Chapter 3, we study the setting of Bounded Knapsack Bandits with an application in Crowdsourcing Systems. Crowdsourcing systems have become a valuable solution for various organizations to outsource work on a temporary basis. Quality assurance in these systems remains a key issue due to the distributed setup of the crowdsourcing platforms and the absence of a priori information about the workers. Our work develops a notion of Limited-information Crowdsourcing Systems (LCS), where the task master can assign the work based on some knowledge of the workers' ability acquired over time. The key challenges in this new setup are determining an efficient workers' selection policy and estimating the abilities of the workers. To address the first challenge, we reduce the problem to an arm-limited, budget limited, multi-armed bandit (MAB) setup, also referred as Bounded Knapsack Bandits, and use the simplified

bounded KUBE (B-KUBE) algorithm as a solution. This algorithm has previously only been experimentally evaluated, and we provide provable performance guarantees, showing that it is order optimal, namely the expected regret of B-KUBE is O(log(B)) where B is the total budget of the task master. The second challenge is solved by formalizing the notion of workers' ability mathematically, and proposing a strategy for its estimation. We experimentally evaluate B-KUBE in conjunction with this strategy, showing that it outperforms other state-of- the-art MAB algorithms when applied in the same setting.

In Chapter 4, we investigate the adversarial Bandits with Knapsack (BwK), where a learner repeatedly chooses to perform an action, pays the corresponding cost, and receives a reward associated with the action. The learner is constrained by the maximum budget that can be spent to perform actions, and the rewards and the costs of the actions are assigned by an oblivious adversary. This problem has only been studied in the restricted setting where the reward of an action is greater than the cost of the action, while we provide a solution in the general setting. Namely, we propose EXP3.BwK, a novel algorithm that achieves order optimal regret. We also propose EXP3++. BwK, which is order optimal in the adversarial BwK setup, and incurs an almost optimal expected regret with an additional factor of in the stochastic BwK setup. Finally, we investigate the case of having large costs for the actions (ie, they are comparable to the budget size), and show that for the adversarial setting, achievable regret bounds can be significantly worse, compared to the case of having costs bounded by a constant, which is a common assumption within the BwK literature.

In Chapter 5, we study online learning when partial feedback information is provided following every action of the learning process, and the learner incurs switching costs for changing his actions. In this setting, the feedback information system can be represented by a graph, and previous works studied the expected regret of the learner in the case of a clique (Expert setup), or disconnected single loops (Multi-Armed Bandits). This work provides a lower bound on the expected regret in the Partial Information (PI) setting, namely for general feedback graphs–excluding the clique. Additionally, it shows that all algorithms that are optimal without

switching costs are necessarily sub-optimal in the presence of switching costs, which motivates the need to design new algorithms. We propose two new algorithms: Threshold Based EXP3 and EXP3.SC. For the two special cases of symmetric PI setting and MAB, the expected regret of both of these algorithms is order optimal in the duration of the learning process. Additionally, Threshold Based EXP3 is order optimal in the switching cost, whereas EXP3.SC is not. Finally, empirical evaluations show that Threshold Based EXP3 outperforms the previously proposed order-optimal algorithms EXP3 SET in the presence of switching costs, and Batch EXP3 in the MAB setting with switching costs.

In Chapter 6, we study bandit algorithms under data poisoning attacks in a bounded reward setting. We consider a strong attacker model in which the attacker can observe both the selected actions and their corresponding rewards, and can contaminate the rewards with additive noise. We show that *any* bandit algorithm with regret $O(\log T)$ can be forced to suffer a regret $\Omega(T)$ with an expected amount of contamination $O(\log T)$. This amount of contamination is also necessary, as we prove that there exists an $O(\log T)$ regret bandit algorithm, specifically the classical UCB, that requires $\Omega(\log T)$ amount of contamination to suffer regret $\Omega(T)$. To combat such poising attacks, our second main contribution is to propose a novel algorithm, Secure-UCB, which uses limited *verification* to access a limited number of uncontaminated rewards. We show that with $O(\log T)$ expected number of verifications, Secure-UCB can restore the order optimal $O(\log T)$ regret *irrespective of the amount of contamination* used by the attacker. Finally, we prove that for any bandit algorithm, this number of verifications $O(\log T)$ is necessary to recover the order-optimal regret. We can then conclude that Secure-UCB is order-optimal in terms of both the expected regret and the expected number of verifications, and can save stochastic bandits from any data poisoning attack.

In Chapter 7, we study poisoning attacks to manipulate any no-regret learning algorithm towards a targeted policy in episodic RL and examines different settings in which different kind of poisoning attacks, reward manipulation and action manipulation, could be damaging. We distinguish between two different settings: unbounded rewards and bounded rewards. In

unbounded rewards setting, we show that reward manipulation attacks are sufficient for an adversary to successfully manipulate any no-regret learning algorithm to follow any targeted policy using $\tilde{O}(\sqrt{T})$ amount of contamination, even without any knowledge of the Markov Decision Process (a.k.a., the *black-box* attacks). In bounded rewards setting, we first demonstrate that only reward manipulation or only action manipulation cannot lead to a successful attack, namely there exists a target policy and an MDP which cannot be attacked successfully by only reward manipulation or only action manipulation. Second, combining reward and action manipulation, we show that the adversary can manipulate any no-regret learning algorithm to follow any targeted policy with $\tilde{O}(\sqrt{T})$ attack cost, i.e., sum of amount of contamination and number of action manipulation, in the black-box attack setup. Our results reveal useful insights about what can or cannot be achieved by an adversary's poisoning attacks, and hopefully can spur more works on the design of robust RL algorithms.

In Chapter 8, we study the problem of learning-based attacks in linear systems, where the communication channel between the controller and the plant can be hijacked by a malicious attacker. We assume the attacker learns the dynamics of the system from observations, then overrides the controller's actuation signal, while mimicking legitimate operation by providing fictitious feedback about the sensor readings to the controller. On the other hand, the controller is on a lookout to detect the presence of the attacker and tries to enhance the detection performance by carefully crafting its control signals. We study the trade-offs between the information acquired by the attacker from observations, the detection capabilities of the controller, and the control cost. Specifically, we provide tight upper and lower bounds on the expected ϵ-deception time, namely the time required by the controller to make a decision regarding the presence of an attacker with confidence at least $(1 - \epsilon \log(1/\epsilon))$. We then show a probabilistic lower bound on the time that must be spent by the attacker learning the system, in order for the controller to have a given expected ϵ-deception time. We show that this bound is also order optimal, in the sense that if the attacker satisfies it, then there exists a learning algorithm with the given order expected deception time. Finally, we show a lower bound on the expected energy expenditure required to guarantee

detection with confidence at least $1 - \epsilon \log(1/\epsilon)$.

In Chapter 9, we introduce the δ-mutual information between uncertain variables as a generalization of Nair's non-stochastic information functional. Several properties of this new quantity are illustrated, and used to prove a channel coding theorem in a non-stochastic setting. Namely, it is shown that the largest δ-mutual information between a metric space and its ϵ-packing equals the (ϵ, δ)-capacity of the space. This notion of capacity generalizes the Kolmogorov ϵ-capacity to packing sets of overlap at most δ, and is a variation of a previous definition proposed by one of the authors. These results provide a framework for developing a non-stochastic information theory motivated by potential applications in control and learning theories. Compared to previous non-stochastic approaches, the theory admits the possibility of decoding errors as in Shannon's probabilistic setting, while retaining its worst-case non-stochastic character.

Chapter 2

Distributed Chernoff Test: Optimal Decision Systems over Network

2.1 Introduction

With the boom in the Internet of Things, sensor-network based solutions for inference systems have become increasingly popular [15, 130, 144]. This is mainly due to the decreasing cost of the sensors, their increasing computational capabilities, the availability of high-speed communication channels, and the redundancy provided by the distributed nature of the network [215]. Inference systems have two key functionalities: decision making (*viz.* hypothesis testing) and estimation. We focus on designing optimal tests for sensor networks in decision-making scenarios where the sensors actively choose their probing actions by continuously learning from past observations. Applications that fall in this framework include intrusion and target detection, and object classification and recognition [88, 138, 13, 123, 185].

Previous studies are broadly classified into two categories: fusion-center based and distributed setting. In the first case, all the nodes of the network are connected to a fusion center — and two operative modalities are considered. In the first modality, the network nodes simply deliver their observations to the fusion center, where the inference task is performed. In the second modality, the nodes exploit their computational capability to perform preliminary processing of the observations, and only a limited amount of information is delivered to the fusion center for making the final decision. This reduces the communication overhead, but

may also result in a loss of performance. In the distributed setup, network nodes are connected to each other via communication links, typically forming a sparse network, and there is no central processing unit. Thus, to perform an inference task, the network nodes need to perform computations locally, share their processed data with neighboring nodes, and collectively reach a decision. A natural question in both settings is what kind of local processing to perform at the nodes, and what fusion scheme to adopt at the fusion center or at the network nodes, in order to reduce the communication burden while keeping a high level of performance. In this work, we address this question and propose statistical tests for both settings.

Hypothesis tests can be broadly classified as sequential or non-sequential tests, as well as adaptive or non-adaptive tests. In a sequential test the number of observations needed to reach a decision is not fixed in advance, but depends on the realization of the observed data. The test proceeds to collect and process data until a decision with a prescribed level of reliability can be made, and an important performance figure — in addition to the probability of correct decision — is the average number of observations required to end the test. In an adaptive test, the sensors' probing actions are chosen on the basis of the collected data in an on-line, causal manner. Hence, the sensors learn from the past, and adapt their future probing actions in a closed-loop fashion. In this case, the sensors are said to be "active," in the sense that measurement observations are the consequence of the sensors' chosen probing actions. Our focus here is on sequential and adaptive tests.

We propose a Decentralized Chernoff Test (DCT) for the fusion center based setup, and a Consensus-based Chernoff Test (CCT) for the distributed setup. We provide bounds on the performance of the tests in terms of their risk, defined as the expected cost required to reach a decision plus the expected cost of making a wrong decision. We also provide converse results showing the best possible performance of *any* adaptive or non-adaptive sequential test over the network. We show that DCT is asymptotically optimal in terms of both the risk and the higher moments of the expected decision time, as the observation cost per unit time tends to zero. Additionally, DCT is parsimonious in terms of communication: when the observation cost per

unit time tends to zero, the expected number of messages sent per node tends to a small constant. Finally, we show that CCT also retains the asymptotic optimality of Chernoff's original solution, being order optimal up to a multiplicative constant, in terms of both risk and higher moments of decision time.

To ease the presentation, our initial analysis assumes ideal communication links carrying real-valued messages without errors. In a real network, messages are quantized into packets of a fixed length, and subject to random erasures at each transmission. In the second part of the paper, we extend our results to this scenario.

The rest of the paper is organized as follows: Section 2.2 discusses related work; Section 2.3 formulates the problem; Section 2.4 reviews the standard Chernoff test; Section 2.5 introduces the Decentralized Chernoff Test (DCT); Section 2.6 introduces the Consensus-based Chernoff Test (CCT); Section 2.7 presents theoretical results on DCT and CCT; Section 2.8 presents simulation results; Section 2.9 extends the analysis to quantized messages and erasure channels; Section 2.10 concludes the work. The proofs of all results appear in the Appendices.

2.2 Related Work

Sequential tests were first introduced by Wald in 1973 [221]. One of these tests, the Sequential Probability Ratio Test (SPRT) has been proven optimal for binary hypothesis testing in [222], and for multi-hypothesis testing in [57, 58]. The performance of sequential tests can be further improved by combining them with adaptive schemes. These schemes operate in closed-loop, adapting the choice of actions to past observations. In the case of sequential and adaptive tests, Chernoff provided the optimal test for binary composite hypotheses in [45]. Its asymptotic optimality for multi-hypothesis testing was proven in [164]; see also [68] and references therein for an application. Later, the sequentiality and adaptivity gains for different classes of tests were studied, and it was established that sequential adaptive tests outperform other classes of tests [155], and that the gains can vary from application to application [154, 156, 75, 174]. All

of these results were established in the case a single agent performs the test.

Different works discuss the extension to an ensemble of networked sensors independently making observations and coordinating to reach a decision [26, 219]. Different techniques for combining the information from different sensors at a fusion center are considered in [42, 215, 94, 134]. In this case, minimization of the risk, which depends on both the decision time and the reliability of the decision, requires joint optimization over both the node level computations and the fusion center operations. Key challenges of this optimization problem are discussed in [216], and asymptotically optimal sequential (non-adaptive) tests have been developed in [149, 225].

Previous works have not considered the performance of sequential, adaptive tests in a network setting. The DCT proposed here fills this gap for star networks, namely for networks in which each node is directly connected to a fusion center. On the other hand, the CCT proposed here considers networks having a general graph structure and no central entity. In this more general case, different non-sequential tests have been developed relying on gossip protocols for distributed computation [29, 165, 30, 48, 108, 31]. These protocols can be broadly classified into two categories: consensus protocols and running-consensus protocols. In consensus protocols, a distributed computation task is performed after the collection of all the measurements at the network nodes [29, 165, 48, 108]. Necessary and sufficient conditions for convergence are well studied, see e.g., [234]. In running-consensus protocols, the collection of the measurements from the environment and the computation task are performed simultaneously at the network nodes [30, 31]. Hypothesis testing schemes typically rely on consensus over "belief vectors." In this case, each network node holds a belief vector, whose elements represent the probability that a certain hypothesis is true, given all the information collected by the node. Different strategies are then used to transmit and combine the belief vectors over the network, leading to asymptotic learning of the correct hypothesis [93, 166, 161, 197, 59, 125]. For example, a strategy based on distributed dual averaging was proposed in [59], using an optimization algorithm from [196]. The work in [93] proposes usage of linear consensus strategies to combine the belief vectors, and [166] extends the results of [93] to the case of random time-varying networks. Other works

consider Bayesian strategies for updating and combining the belief vectors at the nodes [125]. In [125] the bounds on the asymptotic learning rate are presented in terms of KL-divergences of the beliefs at the different network nodes. Under the assumption that the log-likelihood ratio is bounded, finite-time analysis of the KL-divergence cost has been performed in [197]. Similar results have been obtained for networks modeled as time-varying graphs [161, 162].

Despite this huge literature, only limited attention has been given to distributed *sequential* hypothesis testing over general networks, which requires designing an appropriate stopping rule over the network and evaluating the corresponding expected decision time and performance in terms of risk. Recently, a sequential (non-adaptive) hypothesis test which is asymptotically optimal among non-adaptive tests has been proposed [131]. In the present work, we propose a sequential as well as *adaptive* hypothesis test in the distributed network setup. Unlike the previous literature, including [131], the proposed test does not perform consensus over the belief vector, and is parsimonious in terms of communication. The stopping criterion proposed in [131] is not applicable to our test. Our test is also asymptotically optimal among all adaptive or non-adaptive sequential tests, under a broad range of conditions. Finally, we point out that unlike our work, all of the above works do not consider the effect of quantization and erasures occuring over the communication links.

2.3 Problem Formulation

2.3.1 Hypotheses Testing model

We consider an ensemble $\mathcal{L} = \{1, 2, \ldots, L\}$ of sensor nodes engaged in a multi-hypothesis testing problem. The state of nature to be detected is one of M exhaustive and mutually exclusive hypotheses $\{h_i\}_{i\in[M]}$, where $[M] = \{1, 2, \ldots M\}$. Nodes are connected by bi-directional communication links to form a network. At each discrete time step n every node $\ell \in \mathcal{L}$ can select a probing action $u_{n,\ell} \in S$, where S is a fixed set of cardinality M. As a consequence of this action, the node observes the realization of a real-valued random variable

$Y_{n,\ell}$ whose distribution is $p_{i,\ell}^{u_{n,\ell}}$ and is known to node ℓ only. The node can then send one message over each of its incident links, and receive one message from each link. The probing actions and the messages sent at time n can be selected based on all past observations, actions taken, and messages sent and received up to time $n-1$. It follows that the observations at each node can be dependent across time. On the other hand, given the state of nature, we assume that the observations at different nodes are conditionally independent, but not necessarily identically distributed.

2.3.2 Network model

We consider two network setups.

1. *Star network.* In this case, the network is composed of the L sensors and of one special node acting as a fusion center. Each sensor is connected to the fusion center via a communication link, while there are no links between the sensors. This setup is used to introduce our Decentralized Chernoff Test (DCT).

2. *General network*. In this case, the network is represented by a connected graph $\mathcal{G}(\mathcal{L}, \mathcal{E})$, where $\mathcal{L}$ is the set of vertices, and the edges $\{(\ell, j)\} \in \mathcal{E}$, are such that $\ell, j \in \mathcal{L}$, we have $\ell \neq j$. Communication and information processing tasks are fully distributed and there is no fusion center. This setup is used to introduce our Consensus-based Chernoff Test (CCT).

2.3.3 Communication model

We first assume an ideal communication model, where at each time step every node can send and receive a vector composed of C real-values over each of its incident links. The messages sent are received instantaneously and without error. This synchronous model of communication with no queueing delay and real vector channels has been widely used in the literature of detection and estimation, see e.g. [149, 225, 29, 165, 30, 48, 108, 31, 93, 161, 197, 59, 125, 234, 131].

We then refine the communication model by taking into account that in a real packet-switched network, links can only carry a finite number of bits at each transmission, rather than real numbers. In this case, if there is a communication link connecting nodes ℓ and j, then we assume that at each time step node ℓ can transmit a packet of C bits to node j and at the same time step node j can transmit a packet of C bits to node ℓ. This accounts for quantization of the real data in the previous model. In information-theoretic terms, every link behaves in each direction as a noiseless channel of finite capacity C bits/transmission. As in the previous model, every packet transmission occurs synchronously in one time step, and there is no queuing delay. Although less popular than the previous one, this refined model has been considered in the context of quantized consensus in [160], and in the context of estimation and detection in [149, 225, 214, 233, 215, 147].

Finally, we further extend the communication model by considering random packet erasures. We assume that at any time step any link in the network can fail independently with probability ϵ. When a link fails, packets travelling on both directions of the link are received as "erasures." In information-theoretic terms, every link behaves in this case in both directions as a C-bit erasure channel without feedback, having capacity $(1-\epsilon)C$ bits/transmission. As in the previous case, transmissions are synchronous, and there is no queuing delay. A related model, where links can fail at random times but carry real numbers rather than quantized packets has been used to study consensus in [101] and estimation and detection in [39, 40, 99, 197, 159, 244, 223, 168, 90, 100].

2.3.4 Performance measure

Our objective is to design a scheme to select at each step the nodes' probing actions and the messages to transmit, to eventually decide the state of nature with sufficiently high reliability. To quantify the performance of the proposed scheme, we let N be the random time at which all the nodes have reached the same decision and halt the test. We consider both the expectation and the higher moments of this stopping time. Following [45], we also consider the risk, expressed

as the sum of the expected cost required to reach a decision and the expected cost of making a wrong decision. Namely, under the true hypothesis $H^* = h_i$, we let the risk $\mathbb{R}_i^\delta$ of a test δ be

$$\mathbb{R}_i^\delta = c\,\mathbb{E}_i^\delta[N] + \omega_i\,\mathbb{P}_i^\delta(\hat{H} \neq h_i), \tag{2.1}$$

where c is the observation cost per unit time, $\hat{H}$ is the final decision, $\mathbb{E}_i$ and $\mathbb{P}_i$ are the expectation and the probability operators computed under $H^* = h_i$, and ω_i is the cost of a wrong decision. As in [45], we evaluate the risk for all $i \in [M]$, as $c \to 0$.

2.3.5 Additional notation

We write log for natural logarithms, unless otherwise indicated. For the general network case, we denote by $d^{\mathcal{G}}$ the diameter of the network, which is the maximum shortest hop-distance between any pair of nodes of $\mathcal{G}(\mathcal{L}, \mathcal{E})$. We denote by $h^{\mathcal{G}}$ the shortest height of all possible spanning trees of $\mathcal{G}(\mathcal{L}, \mathcal{E})$. Since the network is connected, $d^{\mathcal{G}}$ and $h^{\mathcal{G}}$ are both finite. For all $\ell \in [L]$, $u \in S$ and $i, j \in [M]$, the KL-divergence between hypotheses h_i and h_j is denoted by $D(p_{i,\ell}^u || p_{j,\ell}^u)$, and is assumed to be finite over the entire action set S. We also assume that for all $\ell \in [L]$ and $i, j \in [M]$, there exists an action $u \in S$ such that $D(p_{i,\ell}^u || p_{j,\ell}^u) > 0$. This assumption entails little loss of generality, rules out trivialities, and is commonly adopted in the literature, see e.g., [45]. For all $\ell \in [L]$, $u \in S$ and $i, j \in [M]$, we assume $\mathbb{E}[\log(p_{i,\ell}^u(Y))/\log(p_{j,\ell}^u(Y))]^2 < \infty$. If $v_1 = [v_{1,1}, \dots v_{k,1}]$ and $v_2 = [v_{1,2}, \dots v_{k,2}]$ are two vectors of same dimension, then $v_1 \preceq v_2$ implies that for all $i \in [k]$, $v_{i,1} \leq v_{i,2}$. Finally, we indicate with $|v_1|$ the vector of absolute values of the entries of v_1.

2.4 Standard Chernoff Test

We start by describing the Standard Chernoff Test (SCT) for a single sensor ℓ attempting to detect the true hypothesis H^*, having no interactions with other sensors in the network [45]. For all $n > 1$ we let $y_\ell^n = \{y_{1,\ell}, \dots, y_{n-1,\ell}\}$, where $y_{i,\ell}$ denotes the realization of the observation

collected at time step i, and let $u_\ell^n = \{u_{1,\ell}, \ldots, u_{n-1,\ell}\}$, where $u_{i,\ell}$ denotes the realization of the action made at step i. For $n = 1$ we initialize the set of previous actions $u_\ell^n = \emptyset$ and previous observations $y_\ell^n = \emptyset$, and let all posterior probabilities be the same, namely $\mathbb{P}(H^* = h_i | y_\ell^n, u_\ell^n) = 1/M$.

At every step $n \geq 1$, the test proceeds as follows:

1) A temporary decision is made, based on the maximum posterior probability of the hypotheses, given the past observations and actions of the sensor. Ties are resolved at random. This temporary decision is in favor of $h_{i_n^*}$ if

$$i_n^* = \arg\max_{i \in [M]} \mathbb{P}(H^* = h_i | y_\ell^n, u_\ell^n). \tag{2.2}$$

2) A new action $u_{n,\ell}$ is randomly chosen among the elements of the action set S, according to the Probability Mass Function (PMF)

$$Q_{i_n^*}^\ell = \arg\max_{q \in \mathcal{Q}} \min_{j \in M_{i_n^*}} \sum_{u \in [M]} q(u) D(p_{i_n^*,\ell}^u || p_{j,\ell}^u), \tag{2.3}$$

where $\mathcal{Q}$ denotes the set of all the possible PMFs over the $[M]$ actions, and $M_{i_n^*} = [M] \setminus \{i_n^*\}$.

3) As a consequence of this action, a new observation $y_{n,\ell}$ is collected, and for all $i \in [M]$ the posterior probabilities $\mathbb{P}(H^* = h_i | y_\ell^{n+1}, u_\ell^{n+1})$ are updated.

4) The test stops if the worst case log-likelihood ratio crosses a prescribed fixed threshold γ, namely if

$$\log \frac{\mathbb{P}(H^* = h_{i_n^*,} | y_\ell^{n+1}, u_\ell^{n+1})}{\max_{j \neq i_n^*} \mathbb{P}(H^* = h_j | y_\ell^{n+1}, u_\ell^{n+1})} \geq \gamma, \tag{2.4}$$

If the test stops, then the final decision is $h_{i_n^*}$, otherwise n is incremented by one and the procedure continues from 1).

2.5 Decentralized Chernoff Test

We now extend the SCT to a DCT in the star network configuration. We start by noticing that in the SCT the quantity

$$v_{i,\ell} = \max_{q \in \mathcal{Q}} \min_{j \neq i} \sum_{u \in [M]} q(u) D(p^u_{i,\ell} || p^u_{j,\ell}), \tag{2.5}$$

is a measure of the capability of node ℓ to detect hypothesis h_i (see [45] for a discussion), and plays a critical role for the selection of the action in (2.3) that is performed at each step and is adapted to the current belief. In a network setting, the quantity

$$I(i) = \sum_{\ell=1}^{L} v_{i,\ell}, \tag{2.6}$$

represents a measure of the cumulative capability of the network to detect hypothesis h_i and can be used for the selection of the threshold of each node in a coordinated fashion to optimize the expected decision time. Accordingly, in DCT, the fusion center collects $v_{i,\ell}$ for all $i \in [M]$ and $\ell \in [L]$, computes $I(i)$ for all $i \in [M]$, and distributes this result to all the nodes to enable their threshold selection. The nodes then perform SCTs in parallel, until they all reach the same decision and terminate the test. The three phases of the test are as follows:

Initialization phase

1. Without performing any probing action, each node ℓ sends the vector $v_\ell = [v_{1,\ell}, \ldots, v_{M,\ell}]$ to the fusion center.

2. The fusion center sends the cumulative capability vector $I = [I(1), \ldots, I(M)]$ back to each node, and upon reception, each node ℓ computes the vector $\rho_\ell = [\rho_{1,\ell}, \ldots, \rho_{M,\ell}]$ representing its fraction of network detection capability, namely for all $i \in [M]$, we have

$$\rho_{i,\ell} = v_{i,\ell} / I(i). \tag{2.7}$$

Test phase

Proceeding in parallel, every node ℓ performs a SCT using the threshold

$$\gamma = \rho_{i_n^*,\ell} \, |\log c|. \tag{2.8}$$

This threshold depends on both the current estimate of the hypothesis and the node identity, while it was a constant in (2.4). If the log-likelihood ratio in (2.4) exceeds the threshold, node ℓ sends its preference for $h_{i_n^*}$ to the fusion center and continues to run the test. Hence, rather than using it as a stopping condition, the threshold is used here as a triggering condition for the communication of a preference by node ℓ to the fusion center.

Stopping phase

When the preferences expressed by all the nodes are the same, the fusion center sends a halting message to all the nodes, who stop the test and declare the final decision.

The proposed DCT only requires the communication of the messages in the initialization phase, the local preferences from the nodes during the test phase, and the halting message in the stopping phase. We show below that, while maintaining the same asymptotic optimality of the Chernoff test, the oscillations in the local preferences of the nodes in the test phase vanish as $c \to 0$ and, if $C \geq M$, each sensor tends to use the channel on average at most four times: two in the initialization phase, one (on average) to communicate the local preference, and one to receive the halting message. In the case $C < M$, the test retains its asymptotic optimality, although the expected number of channel uses per node increases from four to a constant that is at most $2(M+1)$, since in this case multiple transmissions are needed to communicate each vector in the initialization phase.

2.5.1 Informal Discussion of DCT

The key idea behind the proposed DCT is to first determine the individual capabilities of the nodes for detecting the hypotheses. These capabilities are captured by the vector v_ℓ, whose i^{th} element is a measure of node ℓ capability to detect the hypothesis h_i. The fusion center gathers this information, and utilizes it to control the threshold at each node through the vector $\rho_{i,\ell}$. In this context, $I(i)$ is the measure of the cumulative detection capability of the network for hypothesis h_i and $\rho_{i,\ell}$ represents the fraction of this capability contributed by node ℓ for hypothesis h_i. To minimize the expected time to reach a decision, it is desirable to determine the threshold for each node ℓ such that all the nodes require roughly the same time to reach the triggering condition in (2.8). This is achieved by dividing the task of hypothesis testing among the nodes based on their speed of performing the task, so that all the nodes finish their share of the task at roughly the same time.

2.6 Consensus-Based Chernoff Test

We now describe CCT in a general network setup, without a fusion center. The main idea is to generalize the DCT to a fully distributed setting. CCT employs a consensus protocol to agree on the cumulative capability of the network to detect each hypothesis, performs individual SCTs, and then employs another consensus protocol to finalize the decision. To ease the presentation of CCT, similar to [93, 161, 197, 59, 125], we now assume that $C \geq M$, so that consensus can be performed by exchanging real vector messages of size M at every time step. In the case $C < M$ the test proceeds along the same lines, but performing vector communications of size M now requires multiple time-steps, and the test completion time must be scaled accordingly. The three phases of the test are as follows:

Initialization Phase

The nodes use a distributed protocol to compute the vector $I = [I(1), \ldots, I(M)]$. Using consensus, they compute the arithmetic mean I/L, and then compute I using their knowledge of L. For all $\ell \in [L]$, we let the initial estimate for I/L at every node be $\hat{I}_\ell^0 = [v_{1,\ell}, \ldots, v_{M,\ell}]$, which can be computed locally using (2.5). Then, every node ℓ runs the following consensus protocol by iteratively exchanging messages without performing any probing action: for $n \geq 0$,

$$\hat{I}_\ell^{n+1} = w_{\ell,\ell}\hat{I}_\ell^n + \sum_{j \in \mathcal{N}_\ell} w_{\ell,j}\hat{I}_j^n, \tag{2.9}$$

where $\hat{I}_\ell^n = [\hat{I}_\ell^n(1), \ldots, \hat{I}_\ell^n(M)]$ is an estimate of I/L at node ℓ and at time n, $w_{\ell,j}$ is the weight assigned by node ℓ to the estimate received from node j, and $\mathcal{N}_\ell = \{j | \{\ell, j\} \in \mathcal{E}\}$ is the set of neighbors of node ℓ in $\mathcal{G}(\mathcal{L}, \mathcal{E})$. We now rewrite (2.9) in the matrix form as

$$\hat{I}^{n+1} = W\hat{I}^n, \tag{2.10}$$

where $\hat{I}^n$ is an $L \times M$ matrix whose ℓth row is $\hat{I}_\ell^n$ and W is an $L \times L$ matrix whose elements satisfy

$$0 < w_{\ell,j} < 1 \text{ if } j \in \mathcal{N}_\ell \cup \{\ell\}, \text{ otherwise } w_{\ell,j} = 0. \tag{2.11}$$

The following theorem presents the necessary and sufficient conditions for the consensus protocol (2.10) to converge to I/L, as $n \to \infty$.

Theorem 1. [234, Theorem 1]. *The consensus protocol (2.10) converges to I/L as $n \to \infty$ if and only if*

$$\boldsymbol{1}_{L\times 1}^T W = \boldsymbol{1}_{L\times 1}^T, \tag{2.12}$$

$$W\boldsymbol{1}_{L\times 1} = \boldsymbol{1}_{L\times 1}, \tag{2.13}$$

and

$$R\left(W - \frac{\mathbf{1}_{L\times 1}\mathbf{1}_{1\times L}}{L}\right) < 1, \tag{2.14}$$

where $R(\cdot)$ denotes the spectral radius of a matrix, and $\mathbf{1}_{A\times B}$ is a $A \times B$ matrix of all ones. Additionally, the rate of convergence is proportional to the spectral radius in the left-hand side of (2.14).

Based on the above theorem, the computation of the weights in the matrix W can be formulated as a convex optimization problem minimizing the spectral radius in (2.14), subject to (2.11), (2.12) and (2.13), and can be determined using standard techniques [234]. Hence, in the following we assume that, in addition to (2.11), the matrix W verifies the conditions stated in Theorem 1.

Although the consensus protocol converges to the correct value I/L as $n \to \infty$, the initialization phase must terminate in finite time and guarantee that consensus has been reached in a suitable approximate fashion.

To characterize approximate consensus, we define a *local* c-consensus status if for all $\ell \in [L]$ and $j \in \mathcal{N}_\ell$, we have

$$|\hat{I}_\ell^n - \hat{I}_j^n| \preceq \frac{c}{L^2}\mathbf{1}_{1\times M}. \tag{2.15}$$

We also define a *global* c-consensus status if for all $\ell, j \in [L]$, we have

$$|\hat{I}_\ell^n - \hat{I}_j^n| \preceq \frac{c}{L}\mathbf{1}_{1\times M}. \tag{2.16}$$

Since the diameter $d^{\mathcal{G}} \leq L$, it should be clear that local c-consensus implies *global* c-consensus.

We employ a stopping rule for the initialization phase that guarantees global c-consensus, and is illustrated in Algorithm 1. A similar rule has been previously studied in [235]. In Algorithm 1, the variable r_ℓ^n indicates the number of time steps since node ℓ is in *local* c-consensus, namely satisfies (2.15). The variable z_ℓ^n is responsible for the percolation of the consensus information across the network. If at any node ℓ we have $z_\ell^n > L + 1$, then the network has reached global

Algorithm 1. Initialization Phase of CCT

Initialize $n = 0$, and for all $\ell \in [L]$, $\hat{I}_\ell^n$, $r_\ell^n = 0$ and $z_\ell^n = 0$
while True do
  For all $\ell \in [L]$, broadcast local information $\hat{I}_\ell^{(n)}$ and z_ℓ^n.
  Update the local cumulative capability using (2.9).
  if $n \geq 1$ **then**
    $z_\ell^n = \min\{r_\ell^{n-1}, \min_{j \in \mathcal{N}_\ell \cup \{\ell\}} z_j^{n-1}\} + 1$
  end if
  if $z_\ell^n > L + 1$ or $m^{(1)} = 1$ is received **then**
    $\hat{I}_\ell^n \leftarrow L\hat{I}_\ell^n$
    Sensor ℓ broadcasts $m^{(1)} = 1$ and stops updating.
    Break While;
  end if
  if $\max_{j \in \mathcal{N}_\ell} |\hat{I}_\ell^{(n)} - \hat{I}_j^{(n)}| \preceq c\mathbf{1}_{1\times M}/L^2$ **then**
    $r_\ell^n = r_\ell^{n-1} + 1$
  else
    $r_\ell^n = 0$
  end if
  $n = n + 1$
end while

c-consensus and node ℓ sends a termination message $m^{(1)} = 1$ to its neighbors, where the superscript indicates that this is the termination message of the initialization phase. When a node receives a termination message, it halts the protocol, it scales the final estimate by L, namely

$$\hat{I}_\ell^n \leftarrow L\hat{I}_\ell^n, \tag{2.17}$$

and forwards the termination message to its neighbors. It follows that all the nodes receive a termination message at most $d^{\mathcal{G}}$ time steps after the first termination message has been sent, and at the end of the initialization phase for all $\ell, j \in [L]$, we have

$$|\hat{I}_\ell^n - \hat{I}_j^n| \preceq c\mathbf{1}_{1\times M}. \tag{2.18}$$

In the following phases, we let $\hat{I}_\ell$ denote the estimate of vector I at node ℓ at the end of the initialization phase.

Algorithm 2. Test Phase of CCT

For all $i \in [M]$ and $\ell \in [L]$, $n = 0$; $\hat{H}_\ell^n = \text{NULL}$
Input: Termination message of stopping phase, i.e., $m^{(3)}$
while Final decision is not made, namely $m^{(3)} \neq 1$ **do**
  For all $\ell \in [L]$, perform SCT with $\gamma = \hat{\rho}_{i_n^*,\ell} \, |\log c|$
  If $\hat{H}_\ell^n \neq$NULL, then broadcast $\hat{H}_\ell^n$
  $n = n + 1$
end while

Algorithm 3. Stopping Phase of CCT

For all $\ell \in [L]$, initialize $n = 0$; $d_{\ell,n} = x_\ell^n = 0$, $m^{(3)} = 0$;
while TRUE **do**
  if $m^{(3)} = 1$ is received from neighbor j **then**
    Set the final decision, i.e., $\hat{H}_\ell^n = \hat{H}_j^{n-1}$
    Broadcast $m^{(3)}$ and $\hat{H}_\ell^n$.
    Break;
  end if
  For all $\ell \in [L]$, update x_ℓ^n according to (2.21).
  For all $\ell \in [L]$, update d_ℓ^n according to (2.20).
  if $d_\ell^N > L + 1$ **then**
    $m^{(3)} = 1$
    For all $\ell \in [L]$, broadcast $m^{(3)}$ and $\hat{H}_\ell^n$.
  else
    For all $\ell \in [L]$, broadcast d_ℓ^n and $\hat{H}_\ell^n$.
  end if
  $n = n + 1$
end while

Test Phase

This phase is illustrated in Algorithm 2 and begins following the termination of the initialization phase, namely after receiving $m^{(1)} = 1$. Every node ℓ performs a SCT using the threshold

$$\gamma = \hat{\rho}_{i_n^*,\ell} \, |\log c|, \tag{2.19}$$

where $\hat{\rho}_{i_n^*,\ell}^n = v_{i_n^*,\ell} / \hat{I}_\ell(i_n^*)$. If the log-likelihood in (2.4) exceeds the threshold, then node ℓ updates its local preference $\hat{H}_\ell^n$ in favor of the hypothesis $h_{i_n^*}$; otherwise, it sets its local preference to NULL. Similar to DCT, the node ℓ communicates its preference $\hat{H}_\ell^n$, if any, to its neighbors (instead than to the fusion center) and continues to run the test. Hence, rather than using it as a

stopping condition, the threshold is used here as a triggering condition for the communication of the preference by node ℓ to its neighbors in $\mathcal{N}_\ell$.

Stopping Phase

This phase is illustrated in Algorithm 3, and runs in parallel with the test phase. This phase detects if all the network nodes have reached the same preference, and halts the test if the preferences are the same. At every time step $n \geq 1$, every node $\ell \in [L]$ sends d_ℓ^n to its neighbors which is defined as

$$d_\ell^n = \min \left\{ \min_{j \in \mathcal{N}_\ell \cup \{\ell\}} d_j^{n-1}, x_\ell^{n-1} \right\} + 1, \tag{2.20}$$

where $d_\ell^0 = 0$, $x_\ell^0 = 0$, and

$$x_\ell^n = \begin{cases} x_\ell^{n-1} + 1 & \text{if } \forall j \in \mathcal{N}_\ell, \hat{H}_\ell^n = \hat{H}_j^n, \hat{H}_\ell^n = \hat{H}_\ell^{n-1}, \text{ and } \hat{H}_\ell^n \neq \text{NULL}, \\ 1 & \text{if } \forall j \in \mathcal{N}_\ell, \hat{H}_\ell^n = \hat{H}_j^n, \hat{H}_\ell^n \neq \hat{H}_\ell^{n-1}, \text{ and } \hat{H}_\ell^n \neq \text{NULL}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.21}$$

The rationale of (2.20) and (2.21) is as follows. Suppose $x_\ell^n = k$. Then, for the past k time steps the local preference of the neighbors of node ℓ was the same as the local preference $\hat{H}_\ell^n$ of node ℓ. The value of d_ℓ^n is responsible for the percolation of this information along the network. Using (2.21), if node $j \in \mathcal{N}_\ell$ does not report any local preference, then the value $x_j^n = 0$ is received by the neighbors of j. If at any node ℓ we have $d_\ell^N > L + 1$, then there exists a time $k \leq N$ at which the local decisions of all the nodes are the same, namely $\min_{j \in [L]} x_j^k \geq 1$ (see Lemma 3 in Appendix 2.12.2). This node ℓ sends the final decision $\hat{H}_\ell^N$ and the termination message $m^{(3)} = 1$ to its neighbors, where $m^{(3)} = 1$ represents the termination message for the stopping phase. When a node receives the termination message and the final decision $\hat{H}_\ell^N$, it halts the test and forwards $m^{(3)}$ along with $\hat{H}_\ell^N$ to its neighbors. It follows that all nodes receives the termination message and the final decision at most $d^{\mathcal{G}}$ time steps after the first termination message of the stopping phase has been sent.

2.6.1 Informal Discussion of CCT

As in DCT, the key idea behind CCT is to first determine the individual capabilities of the nodes for detecting the hypotheses. These capabilities are captured by the vector v_ℓ, whose i^{th} element is a measure of node's ℓ capability to detect the hypothesis h_i. However, in contrast to DCT, there is no central entity to facilitate the sharing of this information among different nodes, and a consensus algorithm is used — in the first phase of CCT — to gain global knowledge at each node of the capabilities of all the other nodes. If the consensus algorithm stops at time N, then $\hat{\rho}_{i,\ell}^N$ denotes the estimated fraction of the capability contributed by node ℓ for hypothesis h_i. To minimize the expected time to reach a decision, it is desirable to determine this threshold for each node ℓ such that all the nodes require roughly the same time to reach the triggering condition in (2.8). This is achieved by dividing the task of hypothesis testing among the nodes based on their speed of performing the task, so that all the nodes finish their share of the task roughly at the same time. The decision phase is a distributed stopping criterion for the Chernoff test, and ensures that the nodes stop the test as they reach the same decisions.

2.7 Performance analysis

We now present the performance analysis of our tests. The proofs of all theorems are deferred to the Appendices.

2.7.1 Lower Bounds for a Sequential and an Adaptive test

In this section, we present lower bounds on two different performance measures, namely risk and decision time, for *any* sequential and adaptive test. The superscript δ is appended to quantities that refer to a generic test and N indicates the time required to take a decision.

Theorem 2. (Converse) *For any hypothesis testing scheme δ operating over a network as*

described in Section 2.3, we have that for all $i \in [M]$, if the probability of missed detection is

$$\mathbb{P}_i^{\delta}(\hat{H} \neq h_i) = O(c\,|\log c|), \quad \textit{as } c \to 0, \tag{2.22}$$

then for all integers $r \geq 1$, we have

$$\mathbb{E}_i^{\delta}[N^r] \;\; \geq \;\; \left((1+o(1))\frac{|\log c|}{I(i)}\right)^r, \; \textit{as } c \to 0. \tag{2.23}$$

Using (2.23) with $r = 1$, we also have

$$\mathbb{R}_i^{\delta} \;\; \geq \;\; (1+o(1))\frac{c\,|\log c|}{I(i)}, \quad \textit{as } c \to 0. \tag{2.24}$$

The lower bounds provided by Theorem 2 hold for any scheme operating in our problem formulation setting, in both a star network or general network configuration. In the case the network is composed of a single node and $r = 1$, these results recover Chernoff's original results [45].

2.7.2 Upper bounds for proposed DCT and CCT schemes

We now provide upper bounds on the performance of our schemes, starting with DCT. In the following theorems, the superscript $\mathcal{D}$ refers to the DCT. Part (i) of Theorem 3 states that the probability of making a wrong decision can be made as small as desired by an appropriate choice of the observation cost c. Part (ii) provides an upper bound on the expected time to reach the final decision, and part (iii) bounds the risk as an immediate consequence of parts (i) and (ii). Finally, part (iv) presents an upper bound on the higher moments of the decision time of DCT.

Theorem 3. (Direct). *The following statements hold:*

(i) *For all $c \in (0,1)$ and $i \in [M]$, the probability that DCT makes an incorrect decision is*

$$\mathbb{P}_i^{\mathcal{D}}(\hat{H} \neq h_i) \leq \min\{(M-1)c, 1\}. \tag{2.25}$$

(ii) *For all* $\ell \in [L]$, $i, j \in [M]$ *and* $u \in S$, *if* $\mathbb{E}\big[\log p^u_{i,\ell}(Y)/p^u_{j,\ell}(Y)\big]^2 < \infty$, *then we have*

$$\mathbb{E}_i^{\mathcal{D}}[N] \leq (1 + o(1))\frac{|\log c|}{I(i)}, \quad \textit{as } c \to 0. \tag{2.26}$$

(iii) *Combining* (i) *and* (ii), *we have*

$$\mathbb{R}_i^{\mathcal{D}} \leq (1 + o(1))\frac{c\,|\log c|}{I(i)}, \quad \textit{as } c \to 0. \tag{2.27}$$

(iv) *For all* $\ell \in [L]$, $i, j \in [M]$, $u \in S$ *and all integers* $r \geq 2$, *if*

$$\mathbb{E}\big[|\log p^u_{i,\ell}(Y)/p^u_{j,\ell}(Y)|^{r+1}\big] < \infty, \tag{2.28}$$

then we have

$$\mathbb{E}_i^{\mathcal{D}}[N^r] \leq \left((1 + o(1))\frac{c\,|\log c|}{I(i)}\right)^r, \ \textit{as } c \to 0. \tag{2.29}$$

In the above theorem, the bound on the expected decision time in (ii) requires the second moment of the log-likelihood ratio to be finite. Likewise, for all $r \geq 2$, the bound on the r^{th} moment of the decision time requires the $r + 1^{st}$ moment of the log-likelihood ratio to be finite.

The next result is a consequence of Theorems 2 and 3. It shows the asymptotic optimality of DCT, and presents the expected communication overhead, as $c \to 0$.

Theorem 4. *For any hypothesis testing scheme* δ *operating over a network as described in Section 2.3, we have*

(i) *For all* $\ell \in [L]$, $i, j \in [M]$ *and* $u \in S$, *if* $\mathbb{E}\big[\log p^u_{i,\ell}(Y)/p^u_{j,\ell}(Y)\big]^2 < \infty$, *then we have*

$$\lim_{c \to 0} \frac{\mathbb{E}_i^{\mathcal{D}}[N]}{\mathbb{E}_i^{\delta}[N]} \leq 1. \tag{2.30}$$

Additionally,

$$\lim_{c\to 0} \frac{\mathbb{R}_i^{\mathcal{D}}}{\mathbb{R}_i^{\delta}} \;\leq\; 1. \tag{2.31}$$

(ii) *For all* $\ell \in [L]$*,* $i, j \in [M]$*,* $u \in S$ *and all integers* $r \geq 2$*, if*

$$\mathbb{E}\big[|\log p_{i,\ell}^{u}(Y)/p_{j,\ell}^{u}(Y)|^{r+1}\big] < \infty, \tag{2.32}$$

then we have

$$\lim_{c\to 0} \frac{\mathbb{E}_i^{\mathcal{D}}[N^r]}{\mathbb{E}_i^{\delta}[N^r]} \leq 1. \tag{2.33}$$

(iii) *Assuming* $C \geq M$*, and letting the communication overhead* C_O *be the number of channel usages by each node, we have*

$$\lim_{c\to 0} \mathbb{E}_i^{\mathcal{D}}[C_O] = 4. \tag{2.34}$$

Combining Theorem 3 and Theorem 4, it follows that DCT is asymptotically optimal in terms of stopping time and risk, as the observation cost tends to zero. This asymptotic optimality, expressed by (2.30), (2.31), and (2.33), holds for all values of C, although in the case $C < M$ the expected number of channel uses per node in (2.34) increases from four to a constant that is at most $2(M + 1)$, due to multiple transmissions required to communicate each vector in the initialization phase. We also point out that the performance of DCT depends only on the cumulative capability $I(i)$ of the network to detect hypothesis h_i, and is independent of how the capabilities $v_{i,\ell}$ are distributed over the network. If two networks have the same cumulative capabilities, then the expected decision time will be the same for both of them. These results hold irrespective of the number of nodes in the network.

We now provide upper bounds on the performance of CCT. We make use of the following well known lemma:

Lemma 1. [54, Proposition 1]. *For any connected graph $\mathcal{G}(\mathcal{L}, \mathcal{E})$ with weights assigned to the edges satisfying* (2.11)*, we have that*

$$0 < \eta(W^{h^{\mathcal{G}}}) < 1, \tag{2.35}$$

where

$$\eta(W) = \min_{i \neq j} \sum_{k=1}^{L} \min\{w_{i,k}, w_{j,k}\}, \tag{2.36}$$

is the ergodic coefficient of the weight matrix W.

In the following theorems, the superscript $\mathcal{C}$ refers to the CCT. Part (i) of Theorem 5 states that the probability of making a wrong decision can be made as small as desired by an appropriate choice of c. Part (ii) provides an upper bound on the expected time to reach the final decision, and part (iii) bounds the risk as an immediate consequence of parts (i) and (ii). Finally, part (iv) presents an upper bound on the higher moments of the decision time of CCT.

Theorem 5. (Direct). *Assuming $C \geq M$, the following statements hold:*
(i) *For all $c \in (0, 1)$ and $i \in [M]$, the probability that CCT makes an incorrect decision is*

$$\mathbb{P}_i^{\mathcal{C}}(\hat{H} \neq h_i) \leq \min\left\{(M-1)c^{\frac{1}{1+c/I(i)}}, 1\right\}. \tag{2.37}$$

(ii) *For all $\ell \in [L]$, $i, j \in [M]$ and $u \in S$, if $\mathbb{E}\big[\log p_{i,\ell}^u(Y)/p_{j,\ell}^u(Y)\big]^2 < \infty$, then we have*

$$\mathbb{E}_i^{\mathcal{C}}[N] \leq (1 + o(1))\left(\frac{h^{\mathcal{G}}|\log(c/\max_{j\in[L]} I(j))|}{|\log\left(1 - \eta(W^{h^{\mathcal{G}}})\right)|} + \frac{|\log c|}{I(i) - c}\right), \tag{2.38}$$

as $c \to 0$.

(iii) *Combining (i) and (ii), we have*

$$\mathbb{R}_i^{\mathcal{C}} \leq (1 + o(1))\left(\frac{h^{\mathcal{G}} c|\log(c/\max_{j\in[L]} I(j))|}{|\log(1 - \eta(W^{h^{\mathcal{G}}}))|} + \frac{c|\log c|}{I(i) - c}\right), \tag{2.39}$$

as $c \to 0$.

(iv) *For all* $\ell \in [L]$, $i, j \in [M]$, $u \in S$ *and all integers* $r \geq 2$, *if*

$$\mathbb{E}\big[|\log p^u_{i,\ell}(Y)/p^u_{j,\ell}(Y)|^{r+1}\big] < \infty, \tag{2.40}$$

then we have

$$\mathbb{E}^{\mathcal{C}}_i[N^r] \leq \left((1+o(1))\left(\frac{h^{\mathcal{G}}|\log(c/\max_{j\in[L]} I(j))|}{|\log(1-\eta(W^{h^{\mathcal{G}}}))|}+\frac{|\log c|}{I(i)-c}\right)\right)^r, \tag{2.41}$$

as $c \to 0$.

The following result is a consequence of Theorems 2 and 5, and shows that CCT is asymptotically optimal, up to a constant factor, as the observation cost tends to zero.

Theorem 6. *For any hypothesis testing scheme* δ *operating over a network as described in Section 2.3 and assuming* $C \geq M$, *we have:*

(i) *For all* $\ell \in [L]$, $i, j \in [M]$ *and* $u \in S$, *if* $\mathbb{E}\big[\log p^u_{i,\ell}(Y)/p^u_{j,\ell}(Y)\big]^2 < \infty$, *then*

$$\lim_{c\to 0} \frac{\mathbb{E}^{\mathcal{C}}_i[N]}{\mathbb{E}^{\delta}_i[N]} \leq \left(\frac{I(i)h^{\mathcal{G}}|\log(1/\max_{j\in[L]} I(j))|}{|\log(1-\eta(W^{h^{\mathcal{G}}}))|}+1\right). \tag{2.42}$$

Additionally,

$$\lim_{c\to 0} \frac{\mathbb{R}^{\mathcal{C}}_i}{\mathbb{R}^{\delta}_i} \;\leq\; \left(\frac{I(i)h^{\mathcal{G}}|\log(1/\max_{j\in[L]} I(j))|}{|\log(1-\eta(W^{h^{\mathcal{G}}}))|}+1\right). \tag{2.43}$$

(ii) *For all* $\ell \in [L]$, $i, j \in [M]$, $u \in S$ *and all integers* $r \geq 2$, *if*

$$\mathbb{E}\big[|\log p^u_{i,\ell}(Y)/p^u_{j,\ell}(Y)|^{r+1}\big] < \infty, \tag{2.44}$$

then we have

$$\lim_{c\to 0}\frac{\mathbb{E}_i^{\mathcal{C}}[N^r]}{\mathbb{E}_i^{\delta}[N^r]} \;\leq\; \left(\frac{I(i)h^{\mathcal{G}}|\log(1/\max_{j\in[L]} I(j))|}{|\log(1-\eta(W^{h^{\mathcal{G}}}))|}+1\right)^r. \tag{2.45}$$

While Theorems 5 and 6 provide bounds for the case $C \geq M$, it should be clear from their proof that when $C < M$ CCT is still asymptotically optimal up to a constant factor, as $c \to 0$. In this case, the right-hand sides of (2.42) and (2.43) are simply scaled by an additional factor that is upper bounded by M, due to the multiple transmissions required to complete each vector transmission. Similarly, the right-hand side of (2.45) is scaled by a factor upper bounded by M^r.

The decision time of CCT, refer (2.38) and (2.41), depends on two terms: A_1 and A_2, where

$$A_1 = \frac{h^{\mathcal{G}}|\log(c/\max_{j\in[L]} I(j))|}{|\log\left(1-\eta(W^{h^{\mathcal{G}}})\right)|}, \tag{2.46}$$

$$A_2 = \frac{|\log c|}{I(i)-c}. \tag{2.47}$$

Here, A_1 corresponds to the expected time of the initialization phase. Since this phase performs consensus over the network, this time depends on the network parameters $h^{\mathcal{G}}$ and matrix W. Similarly, A_2 corresponds to the expected time of the test phase, where the Chernoff test is performed independently at all the nodes. This time is independent of the network parameters. Finally, since the decision phase of CCT begins only after the termination of the initialization phase and is dependent on the test phase, the expected decision time of CCT depends on $A_1 + A_2$. Thus, in Theorem 6, the ratio of the performance parameters of CCT and of the optimal test converges to the constant $1 + I(i)h^{\mathcal{G}}|\log(1/\max_{j\in[L]} I(j))|/|\log(1-\eta(W^{h^{\mathcal{G}}}))|$. It follows that the gap between the performance parameters of CCT and the optimal test is given by the quantity $I(i)h^{\mathcal{G}}|\log(1/\max_{j\in[L]} I(j))|/|\log(1-\eta(W^{h^{\mathcal{G}}}))|$, and as the expected time of initialization phase decreases this gap decreases.

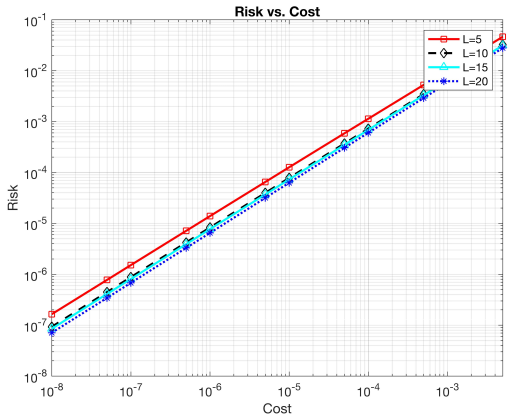


Figure 2.1. Performance of DCT: risk vs. cost c for different number of sensors L

As a final remark, we point out that the star network configuration is a special case of the distributed setup. In this case, the cumulative capability vector I can be estimated (with no error) in two time steps at all the nodes, namely for $n = 2$ and $\ell, j \in [L]$, the equivalent of (2.16) is

$$|\hat{I}_\ell^n - \hat{I}_j^n| \preceq 0\mathbf{1}_{1\times M}, \tag{2.48}$$

and is independent of the parameter c. In the regime of vanishing cost $c \to 0$, we have that $A_1 + A_2 = 2 + A_2$, which implies the asymptotic optimality of DCT.

2.8 Numerical Results

In this section, we evaluate the performance of both DCT and CCT by simulations, and compare the results to the theoretical bounds presented in the previous section. The performance of these tests is evaluated for different sizes of networks. In our experiments, the number of hypotheses is $M = 3$. The probability distribution $p_{i,\ell}^u$ is Bernoulli with parameter p, which is selected uniformly at random from $(0, 1/3)$,$(1/3, 2/3)$ and $(2/3, 1)$ for $i = 1, 2$ and 3 respectively.

Figure 2.1 shows the risk of DCT in a fusion center based setup, as obtained by simulations. Figure 2.2 shows the corresponding value of the risk, as predicted by Theorem 3. The risk decreases as the observation cost c decreases. This is because the threshold in the triggering

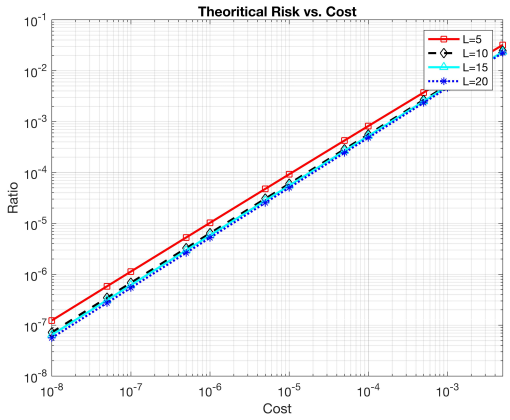


Figure 2.2. Performance of DCT according to Theorem 3: risk vs. cost c for different number of sensors L

condition (2.8) increases, which ensures that the nodes have a greater confidence about their local decision. On the other hand, the risk decreases by increasing the number of sensors L. This is because the cumulative capability of the network to detect the hypothesis, defined in (2.6), increases with L, and the task of hypothesis testing is divided among a larger number of sensors. Hence, the final decision can be reached more quickly, and this decreases the risk. The trends are in agreement with the theoretical results obtained for DCT.

Our simulations also confirm the prediction that, on the average, only four channel usages are required, per single sensor, see (2.34) in Theorem 4. The results of these simulations are not reported here for the sake of brevity. We only mention that, on rare occasions, for individual realizations it may happen that the number of channel usages is substantially larger than four —a manifestation of the long-run phenomenon [169, p. 110]. In practice, this can be remedied by resorting to a truncated version of the sequential test, for which the maximum number of probing actions is fixed, see [207, 206] and references therein for a discussion, and see [148] for a simple implementation of truncation. A precise analysis of DCT using truncated tests is out of the scope of the present paper.

The performance of CCT is evaluated for two network configurations. In the first configuration, given the number of network nodes L, $\lceil L/2 \rceil$ sensors are connected to form a

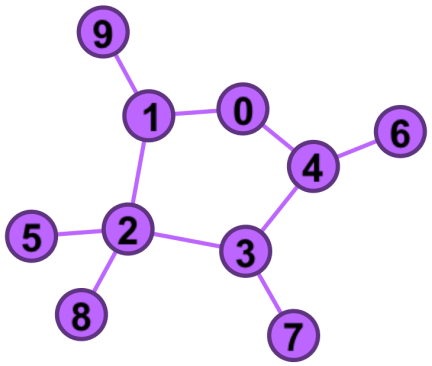


Figure 2.3. An example of sensor network with $L = 10$ nodes.

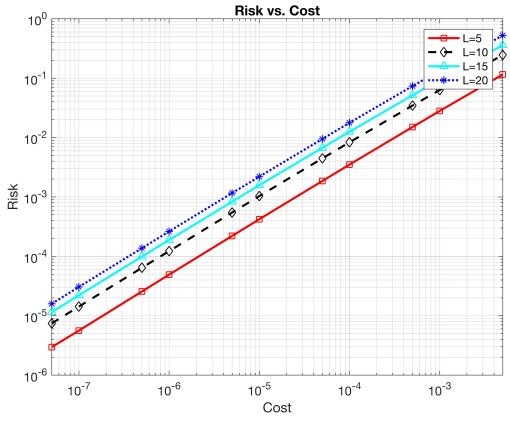


Figure 2.4. Performance of CCT for the ring with random attachments: risk vs. cost c for different number of sensors L

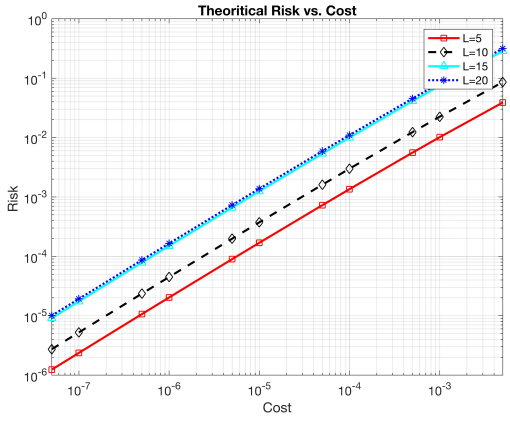


Figure 2.5. Performance of CCT according to Theorem 5 for the ring with random attachments: risk vs. cost c for different number of sensors L

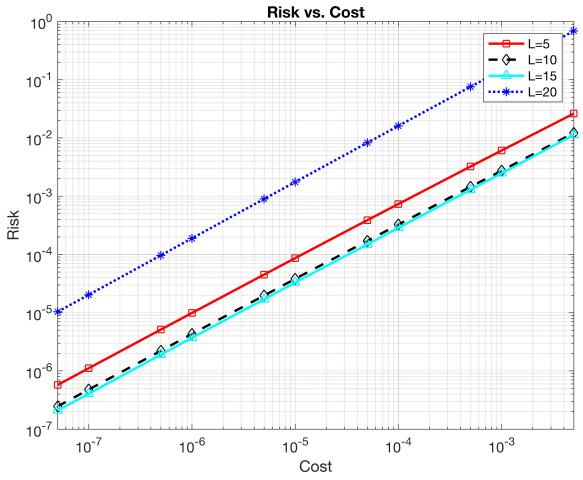


Figure 2.6. Performance of CCT for the tree: risk vs. cost c for different number of sensors L

ring, and the remaining sensors are randomly connected to the sensors in the ring. An example of network with $L = 10$ is shown in Figure 2.3. In this case, the spanning height of the tree is linear in L. In the second configuration, given the number of network nodes L, the nodes are connected to form a binary tree. In this case, the spanning height of the tree is $O(\log_2 L)$.

Figure 2.4 shows the performance of CCT for the ring with random attachments, obtained by computer simulations. Figure 2.5 shows the value of risk according to Theorem 5. Like in the case of DCT, the risk of CCT decreases as the observation cost c decreases. Instead, the behavior as function of L is different. Unlike DCT, the risk of CCT increases by increasing the number of network nodes L. This effect can be explained by observing that in CCT there is a trade-off between the time required by the initialization phase and the time required by the test phase. For the considered network $\mathcal{G}$ and consensus matrix W, as the number of nodes L increases, the consensus scheme in the initialization phase will require more time in comparison to the test phase. Additionally, the time required by the test phase decreases with L, for the same reasons as in the DCT case. Figures 2.4 and 2.5 show that the consensus between the sensors in the first phase of CCT becomes the dominating factor in the decision time. This is in agreement with the theoretical bounds provided in Theorem 5.

Figures 2.6 and 2.7 show the performance of CCT for the tree configuration, via simula-

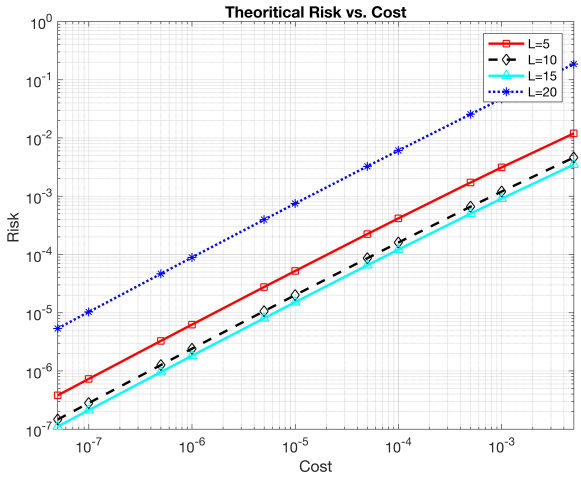


Figure 2.7. Performance of CCT according to Theorem 5 for the tree: risk vs. cost c for different number of sensors L

tions and using the theoretical predictions of Theorem 5, respectively. The risk of CCT decreases as c decreases. Unlike the ring configuration with random attachments, the risk decreases by increasing L until $L = 15$, and then increases. In this setup, for the initial values of L, the time required by the test phase is larger than the time for the initialization phase, hence, it is the dominating factor in the decision time of CCT. On the contrary, for $L = 20$, the time of the initialization phase becomes dominant, which leads to the increase in the risk with L. Finally, comparing Figures 2.6 and 2.7, we see that the theoretical values of the risk are close to the results of numerical simulations.

2.9 Extension to Channels with Quantized Messages and Link Failures

In the previous sections, we have assumed a communication model carrying real numbers over ideal links, without errors. This models a situation where transmission are finely quantized and adequately protected against errors. We now wish to explicitly take into account the effect of data quantization, and of link failures leading to packet erasures.

2.9.1 Channels with Quantized Messages

We start by considering channels supporting quantized messages, rather than real numbers, as described in Section 2.3. We extend our previous results by describing the key changes to both DCT and CCT formulations.

DCT with quantized messages

In the initialization phase, the vectors v_ℓ and I need to be quantized using C bits before transmission. Accordingly, at the sensor nodes we construct the quantized vector $\lfloor v_\ell \rfloor = [\lfloor v_{1,\ell} \rfloor, \ldots, \lfloor v_{M,\ell} \rfloor]$ and at the fusion center we construct the corresponding vector

$$\lfloor I \rfloor = [\lfloor I(1) \rfloor, \ldots, \lfloor I(M) \rfloor]. \tag{2.49}$$

Using (2.6), for all $i \in [M]$ and $\ell \in [L]$, we have that $v_{i,\ell} \leq I(i)$. It follows that to construct the first vector we can divide the interval $[0, \max_i I(i)]$ uniformly into Q sub-intervals, where $Q = 2^{C/M}$, and let $\lfloor v_{i,\ell} \rfloor$ be the nearest value among the Q quantization levels smaller than $v_{i,\ell}$. In this way, the difference between any two contiguous quantization levels for $v_{i,\ell}$ is

$$\Delta\Big(\max_i I(i), Q\Big) = \frac{\max_i I(i)}{Q}. \tag{2.50}$$

The quantized vector $\lfloor v_\ell \rfloor = [\lfloor v_{1,\ell} \rfloor, \ldots, \lfloor v_{M,\ell} \rfloor]$ is then sent by each node to the fusion center using $M \log_2 Q = C$ bits in one transmission. On the other hand, for the second vector we let, for all $i \in [M]$

$$\lfloor I(i) \rfloor = \sum_{\ell=1}^{L} \lfloor v_{i,\ell} \rfloor. \tag{2.51}$$

Since $\lfloor v_{i,\ell} \rfloor$ lies in the interval $[0, \max_i I(i)]$ and $\sum_{\ell=1}^{L} v_{i,\ell} = I(i)$, then $\lfloor I(i) \rfloor$ also corresponds to a quantization level of the interval $[0, \max_i I(i)]$ when this is uniformly divided into Q sub-intervals. It follows that the fusion center can send the vector $\lfloor I \rfloor$ to each node using C bits in one transmission.

Upon reception of $\lfloor I \rfloor$ from the fusion center, every node ℓ computes a vector $\rho_\ell = [\rho_{1,\ell}, \ldots, \rho_{M,\ell}]$, where for all $i \in [M]$

$$\rho_{i,\ell} = \frac{v_{i,\ell}}{\sum_{\tilde{\ell}=1}^{L} \lfloor v_{i,\tilde{\ell}} \rfloor} = \frac{v_{i,\ell}}{\lfloor I(i) \rfloor}, \tag{2.52}$$

and uses it in the test phase for the determination of the threshold in (2.8). In the test phase, each local preference can be communicated using $\log_2 M$ bits and in the stopping phase, the halting message can be communicated using a single bit.

CCT with quantized messages

In the initialization phase, we need to send z_ℓ^n and $\hat{I}_\ell^n$ over the channel at each transmission using C bits. Since the initialization phase terminates when $z_\ell^n > L+1$ (see Algorithm 1), it follows that at most $\log_2(L+2)$ bits are needed to communicate z_ℓ^n. The remaining $\tilde{C} = C - \log_2(L+2)$ bits can then used to communicate the vector $\hat{I}_\ell^n$. Similar to DCT, we divide the interval $[0, \max_i I(i)]$ uniformly into $\tilde{Q} = 2^{\tilde{C}/M}$ sub-intervals so that the difference between any two adjacent quantization levels is

$$\Delta\Big(\max_i I(i), \tilde{Q}\Big) = \frac{\max_i I(i)}{\tilde{Q}}. \tag{2.53}$$

We let the initial estimate $\hat{I}_\ell^0 = [\lfloor v_{1,\ell} \rfloor, \ldots, \lfloor v_{M,\ell} \rfloor]$, where $\lfloor v_{i,\ell} \rfloor$ is the nearest lower value among the $\tilde{Q}$ quantization levels representing $v_{i,\ell}$. The consensus protocol is then modified as follows

$$\hat{I}_\ell^{n+1} = \Big\lfloor w_{\ell,\ell} \hat{I}_\ell^n + \sum_{j \in \mathcal{N}_\ell} w_{\ell,j} \hat{I}_j^n \Big\rfloor. \tag{2.54}$$

It follows that every node ℓ performs a convex combination of the quantized self-estimate $\hat{I}_\ell^n$ and the quantized estimates $\{\hat{I}_j^n\}_{j \in \mathcal{N}_\ell}$ from its neighbors and the updated estimate $\hat{I}_\ell^{n+1}$ is a quantized version of this convex combination. The stopping rule of the initialization phase remains the same as stated in Algorithm 1. In the following phases, we let $\lfloor \hat{I}_\ell \rfloor = [\lfloor \hat{I}_\ell(1) \rfloor, \ldots, \lfloor \hat{I}_\ell(M) \rfloor]$

denote the estimate of the vector I using (2.54) at node ℓ at the end of the initialization phase.

In the test phase of CCT, the SCT is performed locally using the result of the consensus algorithm to select the threshold, namely $\gamma = \hat{\rho}_{i_n^*,\ell}|\log c|$ and $\hat{\rho}_{i_n^*,\ell} = v_{i_n^*,\ell}/\lfloor \hat{I}_\ell(i_n^*) \rfloor$. Finally, in the stopping phase of CCT, presented in Algorithm 3, the variable d_ℓ^n and the local decision $\hat{H}_\ell^n$ are communicated over the channel. Since the stopping phase terminates when $d_\ell^n > L+1$, no more than $\log_2(L+2)$ bits are needed to communicate d_ℓ^n. The local preference $\hat{H}_\ell^n$ can be communicated by $\log_2 M$ bits.

2.9.2 Performance analysis for Channels with Quantized Messages

In this section, we extend the results in Theorem 3 and Theorem 5 to channels with quantized messages.

Theorem 7. (Direct). *Letting*

$$f(Q) = \frac{L \max_i I(i)}{Q}, \tag{2.55}$$

and assuming C is sufficiently large such that for all $i \in [M]$, we have $f(Q) < I(i)$, the following statements hold for DCT:

(i) *For all $c \in (0,1)$ and $i \in [M]$, the probability that the DCT takes an incorrect decision is*

$$\mathbb{P}_i^{\mathcal{D}}(\hat{H} \neq h_i) \leq \min\{(M-1)c, 1\}. \tag{2.56}$$

(ii) *For all $\ell \in [L]$, $i,j \in [M]$ and $u \in S$, if $\mathbb{E}\big[\log p_{i,\ell}^u(Y)/p_{j,\ell}^u(Y)\big]^2 < \infty$, then the expected decision time is*

$$\mathbb{E}_i^{\mathcal{D}}[N] \leq (1+o(1))\frac{|\log c|}{I(i) - f(Q)}, \quad \text{as } c \to 0. \tag{2.57}$$

(iii) *Combining (i) and (ii), the risk defined in (2.1) is*

$$\mathbb{R}_i^{\mathcal{D}} \leq (1+o(1))\frac{c\,|\log c|}{I(i) - f(Q)}, \quad \text{as } c \to 0. \tag{2.58}$$

(iv) *For all* $\ell \in [L]$, $i, j \in [M]$, $u \in S$ *and all integers* $r \geq 2$, *if*

$$\mathbb{E}\big[|\log p^u_{i,\ell}(Y)/p^u_{j,\ell}(Y)|^{r+1}\big] < \infty, \tag{2.59}$$

then the r^{th} *moment of the decision time* N *is*

$$\mathbb{E}^{\mathcal{D}}_i[N^r] \leq \left((1+o(1))\frac{c\,|\log c|}{I(i)-f(Q)}\right)^r, \text{ as } c \to 0. \tag{2.60}$$

By Theorem 7, it follows that the performance of DCT depends on the number of quantization levels through the function $f(Q)$. As $Q \to \infty$, we have that $f(Q) \to 0$ and the results of Theorem 3 are recovered. As $Q \to \infty$, real numbers can be communicated perfectly over the channels, hence DCT incurs no loss of asymptotic performance. We can then view $f(Q)$ as quantifying the loss in the performance of DCT due to quantization. This is also evident by combining (2.50) and (2.51), which show that the quantization error $|I(i) - \lfloor I(i) \rfloor|$ is at most $f(Q)$. By assuming that $f(Q) < I(i)$, our theorem statement ensures that this quantization error is smaller than $I(i)$. Since $Q = 2^{C/M}$, this constraint can be satisfied by having C sufficiently large.

Next, we consider the CCT case. We make the following assumptions that are commonly adopted in the literature of consensus over channels with quantized messages.

Assumption 1. [160, Assumption 1] *The matrix* W *is doubly stochastic, namely (2.12) and (2.13) holds, with positive diagonal entries. In addition, there exists a constant* $\alpha > 0$ *such that if* $w_{i,j} > 0$, *then* $w_{i,j} > \alpha$.

The double stochastic assumption on the weight matrix W guarantees that the average of the sensor values remains the same at each consensus iteration. The second part of Assumption 1 ensures that each sensor gives a non-negligible weight to its values and to the values of its neighbors at each time.

Assumption 2. [160, Assumption 4] *For all* ℓ *and* i, *we have that* $v_{i,\ell}$ *is a multiple of* $M/\tilde{Q}$.

The above assumption states that the values of vector $\hat{I}_\ell^0$ are already quantized, namely $\hat{I}_\ell^0 = [\lfloor v_{1,\ell} \rfloor, \ldots, \lfloor v_{M,\ell} \rfloor] = [v_{1,\ell}, \ldots, v_{M,\ell}]$.

Theorem 8. (Direct). *Let*

$$g(Q,c,\alpha) = \frac{L}{Q}\Bigg(\frac{2L^2}{\alpha}\log(\min(Q^2, L^4/c^2)\max_j I^2(j)) + 1 + h^{\mathcal{G}}\left(\frac{-\log(d^{\mathcal{G}})}{\log\left(1-\eta(W^{h^{\mathcal{G}}})\right)}+1\right) + L + 1\Bigg). \tag{2.61}$$

Assume that C is sufficiently large such that for all $i \in [M]$, we have $g(Q,c,\alpha) < I(i)$ and $C > \log_2(L+2) + \log_2 M$, and Assumptions 1 and 2 hold. Then, the following statements hold for CCT:

(i) *For all $c \in (0,1)$ and $i \in [M]$, the probability that CCT takes an incorrect decision is*

$$\mathbb{P}_i^{\mathcal{C}}(\hat{H} \neq h_i) \leq \min\left\{(M-1)c^{\frac{I(i)}{I(i)+g(\tilde{Q},c,\alpha)}}, 1\right\}. \tag{2.62}$$

(ii) *For all $\ell \in [L]$, $i,j \in [M]$ and $u \in S$, if $\mathbb{E}\big[\log p_{i,\ell}^u(Y)/p_{j,\ell}^u(Y)\big]^2 < \infty$, then the expected decision time is*

$$\mathbb{E}_i^{\mathcal{C}}[N] \leq (1+o(1))\left(\tilde{Q}g(\tilde{Q},c,\alpha) + \frac{|\log c|}{I(i) - g(\tilde{Q},c,\alpha)}\right), \tag{2.63}$$

as $c \to 0$.

(iii) *Combining (i) and (ii), the risk defined in (2.1) is*

$$\mathbb{R}_i^{\mathcal{C}} \leq (1+o(1))\left(\frac{\tilde{Q}g(\tilde{Q},c,\alpha)}{c} + \frac{1}{I(i) - g(\tilde{Q},c,\alpha)}\right)c\,|\log c|, \tag{2.64}$$

as $c \to 0$.

(iv) *For all* $\ell \in [L]$, $i, j \in [M]$, $u \in S$ *and all integers* $r \geq 2$, *if*

$$\mathbb{E}\big[|\log p^u_{i,\ell}(Y)/p^u_{j,\ell}(Y)|^{r+1}\big] < \infty, \tag{2.65}$$

then the r^{th} *moment of the expected decision time is*

$$\mathbb{E}^{\mathcal{C}}_i[N^r] \leq \left((1+o(1))\left(\tilde{Q}g(\tilde{Q},c,\alpha) + \frac{|\log c|}{I(i) - g(\tilde{Q},c,\alpha)}\right)\right)^r, \tag{2.66}$$

as $c \to 0$.

By Theorem 8, it follows that the performance of CCT depends on the number $\tilde{Q}$ of quantization levels through the function $g(\tilde{Q},c,\alpha)$. As $Q \to \infty$, $\tilde{Q} \to \infty$ and $g(\tilde{Q},c,\alpha) \to 0$. The time required by the initialization phase is given by $\tilde{Q}g(\tilde{Q},c,\alpha) = O(|\log(c)|)$ as $Q \to \infty$, which is of the same order as the quantity $h^{\mathcal{G}}|\log(c/\max_{j\in[L]} I(j))|/\log\left(1-\eta(W^{h^{\mathcal{G}}})\right) = O(|\log(c)|)$ appearing in Theorem 5. As $Q \to \infty$, Theorem 8 recovers the same optimality of CCT expressed by Theorem 6. In conclusion, $g(\tilde{Q},c,\alpha)$ quantifies, in terms of the relevant system parameters, the loss in asymptotic performance of CCT due to quantization. In this case, the error for $|I(i) - \lfloor \hat{I}_\ell(i) \rfloor|$ is at most $g(\tilde{Q},c,\alpha)$ and our theorem assumes that this error is smaller than $I(i)$. Since $\tilde{Q} = 2^{\tilde{C}/M}$, this constraint can be satisfied by having C sufficiently large. The additional capacity constraint $C > \log_2(L+2) + \log_2 M$ in the statement of the theorem is due to the transmission of d^n_ℓ and the local preference $\hat{H}^n_\ell$.

2.9.3 Channels with Quantized Messages and Erasures

In this section, we consider both quantized channels and ϵ-random packet erasures, as described in Section 2.3. We extend our previous results by describing key changes to both DCT and CCT.

DCT with Quantization and Erasures

In the initialization phase each node ℓ communicates the vector $\lfloor v_\ell \rfloor$ to the fusion center using a packet of C bits. The expected time for successful transmission of the packet is $1/(1-\epsilon)$. After receiving the vector $\lfloor v_\ell \rfloor$ from all the nodes, the fusion center communicates the vector $\lfloor I \rfloor = [\lfloor I(1) \rfloor, \ldots, \lfloor I(M) \rfloor]$ back to each node ℓ, which requires an expected time $1/(1-\epsilon)$ as well.

In the test phase, each local preference is communicated using a packet of $\log_2 M$ bits to the fusion center, also with an expected time $1/(1-\epsilon)$.

The final decision $\hat{H}$ at the fusion center is made in favor of hypothesis h_i when the local decisions received from all the network nodes are in favor of the hypothesis h_i. Given the local decision h_i is reached at all the nodes, the expected time for reaching the final decision $\hat{H}$ is $1/(1-\epsilon)^L$, as it is required that all the links are simultaneously active. Upon taking the final decision, the fusion center sends a halting message to each node ℓ.

CCT with Quantization and Erasures

In this case, at each time step n, we consider the time-varying graph $\mathcal{G}(\mathcal{L}, \mathcal{E}(n))$, where $\mathcal{E}(n) \subseteq \mathcal{E}$ denotes the set of communication links where a packet can be sent successfully.

In the initialization phase of CCT, since the graph is time-varying, the weight matrix $W = W(n)$ also varies over time. This matrix can be expressed as [101]

$$W(n) = U_{L\times L} - \beta \bar{L}(n), \tag{2.67}$$

where β is a design parameter, $U_{L\times L}$ is the identity matrix of dimension $L \times L$, $\bar{L}(n)$ is the $L \times L$

dimensional Laplacian matrix of $\mathcal{G}(\mathcal{L}, \mathcal{E}(n))$ [101], with entries:

$$\bar{l}_{i,j}(n) = \begin{cases} \sum_{j' \neq i} \mathbf{1}((i,j') \in \mathcal{E}(n)) & \text{if } i = j, \\ -1 & \text{if } (i,j) \in \mathcal{E}(n), \\ 0 & \text{otherwise,} \end{cases} \tag{2.68}$$

where $\mathbf{1}(\cdot)$ denotes the indicator function. Each node i can compute locally the values $\bar{l}_{i,j}(n)$, based on whether a packet is received from node j at time n. Since $\bar{l}_{i,j}(n) = \bar{l}_{j,i}(n)$, it follows that $W(n)$ is a symmetric matrix, where [101]

$$w_{i,j}(n) = \begin{cases} 1 - \beta \sum_{j' \neq i} \mathbf{1}((i,j') \in \mathcal{E}(n)) & \text{if } i = j, \\ \beta & \text{if } (i,j) \in \mathcal{E}(n), \\ 0 & \text{otherwise.} \end{cases} \tag{2.69}$$

Then, as in (2.54), node ℓ updates its quantized estimate at time step n as

$$\hat{I}_\ell^{n+1} = \left\lfloor w_{\ell,\ell}(n) \hat{I}_\ell^n + \sum_{j \in \mathcal{N}_\ell} w_{\ell,j}(n) \hat{I}_j^n \right\rfloor. \tag{2.70}$$

Whenever links are active, the information communicated over the channels is of the same form as that over channels with quantized messages. The stopping rule of this phase remains the same as stated in Algorithm 1. In the following phases, we let $\lfloor \hat{I}_\ell^\epsilon \rfloor = [\lfloor \hat{I}_\ell^\epsilon(1) \rfloor, \ldots, \lfloor \hat{I}_\ell^\epsilon(M) \rfloor]$ denote the estimate of vector I using (2.70) at node ℓ at the end of the initialization phase in this channel model.

In the test phase of CCT, the SCT is performed locally using the result of the consensus algorithm to select the threshold, namely $\gamma = \hat{\rho}_{i_n^*,\ell}^\epsilon |\log c|$ and $\hat{\rho}_{i_n^*,\ell}^\epsilon = v_{i_n^*,\ell} / \lfloor \hat{I}_\ell^\epsilon(i_n^*) \rfloor$.

Finally, in the stopping phase of CCT, presented in Algorithm 3, the variable d_ℓ^n and the local decision $\hat{H}_\ell^n$ are communicated over channel by $\log_2(L+2) + \log_2 M$ bits. Of course,

these communications are successful only when the link between the nodes is active.

2.9.4 Performance Analysis for Channels with Quantized Messages and Erasures

In this section, we extend the results of Theorem 7 and Theorem 8 to channels with quantized messages and erasures.

Theorem 9. *In the presence of channel with quantized messages and ϵ-random packet erasures, Theorem 7 holds unmodified.*

Intuitively, the reason why the results of Theorem 7 hold unmodified is as follows. Link failures only delay the communication of the quantized information over the channel, which impacts the decision time. Note that the expected time for communication of $\lfloor v_\ell \rfloor$ from all the nodes is at most $L/(1-\epsilon)$, as is the expected time to communicate the response vector to all the nodes. Given the same local decision is reached at the nodes, the expected time to reach the final decision is $1/(1-\epsilon)^L$. Likewise, the expected time to communicate the halting message is $L/(1-\epsilon)$. All these delays introduced by the ϵ-erasure channel are finite and independent of c, and are embodied in the terms $o(1)$ appearing in the statement of Theorem 9.

Next, we give a lemma needed to provide the performance guarantees of CCT.

Lemma 2. *For all n and $0 < \beta < 1/(2|\mathcal{E}|)$, the following holds:*
(i) $W(n)$ is a doubly stochastic matrix, namely (2.12) and (2.13) holds.
(ii) For all $i, j \in [L]$, if $w_{i,j}(n) > 0$, then we have $w_{i,j}(n) > \min(1 - \mathcal{D}(\mathcal{G})\beta, \beta)$, where $\mathcal{D}(\mathcal{G}) = \max_s \sum_{j \neq s} \mathbf{1}((j,s) \in \mathcal{E})$ is the maximum node degree in the graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$.
(iii) The spectral radius verifies

$$R\left(W(n) - \frac{\boldsymbol{1}_{L\times 1}\boldsymbol{1}_{1\times L}}{L}\right) < 1. \tag{2.71}$$

Theorem 10. (Direct). *Let* $\epsilon < 1/|\mathcal{E}|$*,*

$$h(Q, c, \alpha, \epsilon) = \frac{g(Q, c, \alpha)(2 - |\mathcal{E}|\epsilon)}{(1 - |\mathcal{E}|\epsilon)^2}, \tag{2.72}$$

$$q(Q, c, \alpha, \epsilon) = \frac{Qg(Q, c, \alpha)}{L(2 - |\mathcal{E}|\epsilon)}, \tag{2.73}$$

and $0 < \beta < 1/(2|\mathcal{E}|)$*. Assume that* C *is sufficiently large such that for all* $i \in [M]$*, we have* $h(\tilde{Q}, c, \min(1 - \mathcal{D}(\mathcal{G})\beta, \beta), \epsilon) < I(i)$ *and* $C > \log_2(L + 2) + \log_2 M$*, and Assumption 2 holds. Then the following statements hold for CCT:*

(i) *For all* $c \in (0, \sqrt{(1 - |\mathcal{E}|\epsilon)/2})$ *and* $i \in [M]$*, the probability that CCT takes an incorrect decision is*

$$\begin{aligned}\mathbb{P}_i^{\mathcal{C}}(\hat{H} \neq h_i) \leq \min\{&(M-1)(1 - \exp(-2q(\tilde{Q}, c, \min(1 - \mathcal{D}(\mathcal{G})\beta, \beta), \epsilon)) \\ &c^{I(i)/(I(i)+h(\tilde{Q},c,\min(1-\mathcal{D}(\mathcal{G})\beta,\beta),\epsilon))} \\ &+ \exp(-2q(\tilde{Q}, c, \min(1 - \mathcal{D}(\mathcal{G})\beta, \beta), \epsilon), 1\}.\end{aligned} \tag{2.74}$$

(ii) *For all* $\ell \in [L]$*,* $i, j \in [M]$ *and* $u \in S$*, if* $\mathbb{E}\big[\log p_{i,\ell}^u(Y)/p_{j,\ell}^u(Y)\big]^2 < \infty$*, then the expected decision time is*

$$\begin{aligned}\mathbb{E}_i^{\mathcal{C}}[N \big| \{\lfloor \hat{I}_\ell^\epsilon \rfloor\}_{\ell \in [L]}] &\leq (1 + o(1))\Bigg(\tilde{Q}h(\tilde{Q}, c, \min(1 - \mathcal{D}(\mathcal{G})\beta, \beta), \epsilon) \\ &\qquad + \frac{|\log c|}{\min_{\ell \in [L]} \lfloor \hat{I}_\ell^\epsilon(i) \rfloor}\Bigg) \qquad (2.75) \\ &\leq (1 + o(1))\Bigg(\tilde{Q}h(\tilde{Q}, c, \min(1 - \mathcal{D}(\mathcal{G})\beta, \beta), \epsilon) \\ &\qquad + \frac{|\log c|}{I(i) - h(\tilde{Q}, c, \min(1 - \mathcal{D}(\mathcal{G})\beta, \beta), \epsilon)}\Bigg), \qquad (2.76)\end{aligned}$$

with probability at least

$$1 - \exp\left(-2q(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon)\right), \text{ as } c \to 0. \tag{2.77}$$

(iii) *Combining* (i) *and* (ii)*, the risk is*

$$\mathbb{R}_i^{\mathcal{C}} \leq (1 + o(1))\left(\frac{\tilde{Q}h(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon)}{c} + \frac{1}{I(i) - h(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon)}\right) c|\log c|, \tag{2.78}$$

with probability at least

$$1 - \exp\left(-2q(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon)\right), \text{ as } c \to 0. \tag{2.79}$$

(iv) *For all* $\ell \in [L]$*,* $i, j \in [M]$*,* $u \in S$ *and all integers* $r \geq 2$*, if*

$$\mathbb{E}\left[|\log p_{i,\ell}^u(Y)/p_{j,\ell}^u(Y)|^{r+1}\right] < \infty, \tag{2.80}$$

then the r^{th} *moment of the expected decision time is*

$$\mathbb{E}_i^{\mathcal{C}}[N^r \big| \{\lfloor \hat{I}_\ell^\epsilon \rfloor\}_{\ell \in [L]}] \leq \left((1 + o(1))\left(\tilde{Q}h(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon) + \frac{|\log c|}{I(i) - h(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon)}\right)\right)^r, \tag{2.81}$$

with probability at least

$$1 - \exp\left(-2q(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon)\right), \text{ as } c \to 0. \tag{2.82}$$

We point out that when estimating the vector $\lfloor \hat{I}_\ell^\epsilon \rfloor$ in the initialization phase of CCT, the ϵ-random erasure model introduces additional randomness. For this reason, (2.75) represents the conditional expected decision time given $\{\lfloor \hat{I}_\ell^\epsilon \rfloor\}_{\ell \in [L]}$. To obtain (2.76), we use the fact that for all $\ell \in [L]$, we have that the random variable

$$\begin{aligned} \lfloor \hat{I}_\ell^\epsilon(i) \rfloor \in [I(i) - h(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon), \\ I(i) + h(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon)], \end{aligned} \tag{2.83}$$

with probability at least

$$1 - \exp\left(-2q(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon)\right), \tag{2.84}$$

shown in (2.169) in Appendix 2.12.2.

In Theorem 10, the performance guarantees are provided with high probability, and this probability depends on the number of quantization levels and on the packet erasures through $q(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon)$. As $c \to 0$ and $Q \to \infty$ (in arbitrary order), we have that $q(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon) \to \infty$ and $1 - \exp\left(-2q(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon)\right)$ converges to one. Additionally, the performance of CCT also depends on $h(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon)$. As $Q \to \infty$, we have $g(\tilde{Q}, c, \alpha) \to 0$, which implies $h(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon) \to 0$. Finally, the time required to complete the initialization phase is given by the quantity

$$\tilde{Q} h(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon) = O(|\log(c)|), \tag{2.85}$$

as $Q \to \infty$.

As $Q \to \infty$, Theorem 10 recovers the same optimality of CCT expressed in Theorem 6. The quantity $h(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon)$ quantifies the loss in performance of CCT due to both quantization and random packet erasures. In this case, since $\lfloor \hat{I}_\ell^\epsilon \rfloor$ is a random variable,

the error for $|I(i) - \lfloor \hat{I}_\ell^\epsilon(i) \rfloor|$ is at most $h(\tilde{Q}, c, \min\,(1 - \mathcal{D}(\mathcal{G})\beta, \beta), \epsilon)$ with high probability, and our theorem assumes that this error is smaller than $I(i)$. Since $\tilde{Q} = 2^{\tilde{C}/M}$, this constraint can be satisfied by having C sufficiently large. The additional capacity constraint $C > \log_2(L+2) + \log_2 M$ in the statement of the theorem is due to the transmission of d_ℓ^n and the local preference $\hat{H}_\ell^n$.

2.10 Conclusion

Networked sensor systems are becoming increasingly popular for inference problems due to their improved robusteness, scalability, versatility, and performance. Initial implementations were based on inexpensive small sensors, with extremely limited hardware/software capabilities. Progressively, these devices acquired more and more functionalities, and are nowadays capable of active sensing, namely they can adapt the probing signal on the basis of previous measurements, in order to optimize their sensing capability. Thus, individual sensors have become intelligent devices which continuously learn from the past and can decide their future actions, in closed-loop adaptive scheme.

We considered two network configurations of these "intelligent" sensors: a star network configuration with a fusion center, and a general network configuration that is fully distributed. In the first configuration, the fusion center coordinates the actions of the remote nodes, and takes the final decision. The second configuration does not have a central coordination, and all the processing takes place at the nodes: they actively collect measurements, exchange information with immediate neighbors, and collectively take a decision.

For the first configuration we proposed a sequential adaptive decision system — referred to as DCT — which operates in three phases. First, there is a round of communication between the fusion center and the remote nodes, needed to define the relative capability of each node to detect the hypotheses. This capability is then used to divide the decision task among the nodes. Each node begins to continuously sense the environment, and makes the central entity aware

about decisions that are locally believed to be sufficiently reliable. The final decision is taken by the fusion center on the basis of these *local suggestions* about the true hypothesis.

We provided a theoretical analysis of detection performance and expected time to reach a decision. We show that the test is asymptotically optimal in terms of detection performance (risk), as the observation cost per unit time tends to zero.

For the second configuration, we exploit ideas from the DCT implementation, combined with gossip protocols that use consensus techniques, to design a fully distributed adaptive sequential decision system, which is referred to as CCT. Our approach is markedly different from those usually exploited in the literature, where real-valued belief vectors are continuously exchanged over the network to reach consensus.

Our CCT works in three phases. In the first phase, a consensus about the relative capability of the nodes to detect the state of nature is achieved by means of gossip protocols with local information exchange. In the second phase, nodes implement the Chernoff test and, once all the network nodes reach their preference, the final decision is reached in a distributed way in the third phase of operation, by diffusing messages across the network that percolate the information of whether the other sensors have terminated their share of task. We prove the asymptotic optimality of CCT, up to a multiplicative factor in terms of both risk and higher moments of the decision time.

2.11 Acknowledgement

Chapter 2, in full, is a reprint of the material as it appears in Anshuka Rangi, Massimo Franceschetti and Stefano Marano, "Distributed Chernoff Test: Optimal Decision Systems Over Networks", *IEEE Transactions on Information Theory*, vol. 67, pp. 2399 - 2425, April 2021, Anshuka Rangi, Massimo Franceschetti and Stefano Marano, "Consensus-Based Chernoff Test in Sensor Networks", *IEEE Conference on Decision and Control (CDC)*, December 2018, and Anshuka Rangi, Massimo Franceschetti and Stefano Marano, "Decentralized chernoff test in

sensor networks", *IEEE International Symposium on Information Theory (ISIT)*, July 2018. The dissertation author was the primary investigator and author of these papers.

2.12 Appendix

2.12.1 Proof of Theorem 2

Proof. Let $H^* = h_i$ be the true hypothesis. The proof of Theorem 2 consists of two parts. First, for all $0 < \epsilon < 1$, we show that for the probability of error to be close to zero, the log-likelihood ratio between h_i and all $h_m \neq h_i$, should be greater than $-(1-\epsilon)\log c$ with high probability as $c \to 0$. Namely, the inequality

$$S^N(h_i, h_m) = \sum_{\ell=1}^{L}\sum_{k=1}^{N} \log \frac{p_{i,\ell}^{u_{k,\ell}}(y_{k,\ell})}{p_{m,\ell}^{u_{k,\ell}}(y_{k,\ell})} \geq -(1-\epsilon)\log c, \tag{2.86}$$

must hold with high probability, as $c \to 0$. Second, we show that for all $0 < \epsilon < 1$ and $n < -(1-\epsilon)\log c / I(i)$, it is unlikely that such inequality is satisfied for some hypothesis $h_m \neq h_i$.

We start by defining two sets of hypotheses $\mathcal{H}_0' = \{h_i\}$ and $\mathcal{H}_1' = \{h_m\}_{m\neq i}$. By (2.22), both type I and type II error probabilities of the hypothesis test $\mathcal{H}_0'$ vs. $\mathcal{H}_1'$ are $O(-c\log c)$. Thus, by [45, Lemma 4], for all hypotheses $h_m \neq h_i$ and $0 < \epsilon < 1$, we have

$$\mathbb{P}_i\Big(S^N(h_i, h_m) \leq -(1-\epsilon)\log c\Big) = \mathcal{O}(-c^{\epsilon}\log c). \tag{2.87}$$

Therefore, as $c \to 0$, the probability in (2.87) tends to 0, which concludes the first part of the proof.

Now, we show that for all $\epsilon > 0$, we have

$$\lim_{n'\to\infty} \mathbb{P}_i\left(\max_{1\leq n\leq n'} \min_{m\neq i} S^n(h_i, h_m) \geq n'(I(i)+\epsilon)\right) = 0. \tag{2.88}$$

We have

$$S^n(h_i, h_m) = \sum_{\ell=1}^{L}\sum_{k=1}^{n}\left(\log\frac{p_{i,\ell}^{u_{k,\ell}}(y_{k,\ell})}{p_{m,\ell}^{u_{k,\ell}}(y_{k,\ell})} - D(p_{i,\ell}^{u_{k,\ell}}||p_{m,\ell}^{u_{k,\ell}})\right) + \sum_{\ell=1}^{L}\sum_{k=1}^{n} D(p_{i,\ell}^{u_{k,\ell}}||p_{m,\ell}^{u_{k,\ell}})$$
$$= M_1^n + M_2^n, \tag{2.89}$$

where

$$M_1^n = \sum_{\ell=1}^{L}\sum_{k=1}^{n}\left(\log\frac{p_{i,\ell}^{u_{k,\ell}}(y_{k,\ell})}{p_{m,\ell}^{u_{k,\ell}}(y_{k,\ell})} - D(p_{i,\ell}^{u_{k,\ell}}||p_{m,\ell}^{u_{k,\ell}})\right), \tag{2.90}$$

is a martingale with mean 0, and

$$M_2^n = \sum_{\ell=1}^{L}\sum_{k=1}^{n} D(p_{i,\ell}^{u_{k,\ell}}||p_{m,\ell}^{u_{k,\ell}}). \tag{2.91}$$

Then, for all $1 \leq n \leq n'$, we have

$$\begin{aligned}\min_{m\neq i} M_2^n &= \min_{m\neq i}\sum_{\ell=1}^{L}\sum_{k=1}^{n} D(p_{i,\ell}^{u_{k,\ell}}||p_{m,\ell}^{u_{k,\ell}}) \\ &\overset{(a)}{\leq} \sum_{\ell=1}^{L}\sum_{k=1}^{n} v_{i,\ell} \\ &\overset{(b)}{=} nI(i) \\ &\overset{(c)}{\leq} n'I(i),\end{aligned} \tag{2.92}$$

where (a) follows from the definition of $v_{i,\ell}$ in (2.5), (b) follows from the definition of $I(i)$ in (2.6), and (c) follows from $n \leq n'$. Now, using (2.92), if the event in (2.88) occurs for a fixed n_1, namely

$$\min_{m\neq i}(M_1^{n_1} + M_2^{n_1}) \geq n'\big(I(i) + \epsilon\big), \tag{2.93}$$

then there exists a hypothesis h_m such that $M_1^{n_1} \geq n'\epsilon$. Thus, there exists a constant $K' > 0$

such that the probability in (2.88) becomes

$$\begin{aligned}
&\mathbb{P}_i\Big(\max_{1\le n\le n'}\min_{m\neq i} S^n(h_i,h_m)\ge n'(I(i)+\epsilon)\Big)\\
&\le \sum_{m\neq i}\mathbb{P}_i\Big(\max_{1\le n\le n'} M_1^n\ge n'\epsilon\Big)\\
&\overset{(a)}{\le} \frac{(M-1)K'}{n'\epsilon^2},
\end{aligned}\tag{2.94}$$

where (a) follows from the fact M_1^n is a martingale with mean zero and using the Doob Kolmogorov extension of Chebyshev's inequality [56]. Thus, (2.88) follows. As discussed in [45, Theorem 2], for $n_0=-(1-\epsilon)\log c/(I(i)+\epsilon)$, we have

$$\begin{aligned}
\mathbb{P}_i(N\le n_0) &\le \mathbb{P}_i\Big(N\le n_0 \text{ and } \forall m\neq i: S^N(h_i,h_m)\ge n_0(I(i)+\epsilon)\Big)\\
&\quad+\mathbb{P}_i\Big(\exists m\neq i: S^N(h_i,h_m)\le n_0(I(i)+\epsilon)\Big)\\
&\le \mathbb{P}_i\Big(\max_{1\le n\le n_0}\min_{m\neq i} S^n(h_i,h_m)\ge n_0(I(i)+\epsilon)\Big)\\
&\quad+\mathbb{P}_i\Big(\exists m\neq i: S^N(h_i,h_m)\le n_0(I(i)+\epsilon)\Big).
\end{aligned}\tag{2.95}$$

The first and the second terms at the right-hand side of (2.95) approach zero by (2.88) and (2.87) respectively. Now, using (2.95), we also have

$$\mathbb{P}_i(N^r\le n_0^r)=\mathbb{P}_i(N\le n_0)\to 0,\tag{2.96}$$

as $c\to 0$. (2.23) follows upon observing that as $c\to 0$, $\mathbb{E}_i[N^r]\ge n_0^r$ which is

$$\mathbb{E}_i[N^r]\ge\left((1+o(1))\frac{|\log c|}{I(i)}\right)^r.$$

The proof of (2.24) is straightforward. □

2.12.2 Proofs for DCT and CCT

Proof of Theorem 3

Proof. To prove Theorem 3, we need some additional notation. Let $A_{n,j}$ be the set of sample paths where the decision made by the fusion center is in favor of h_j at the n^{th} step, and we indicate a single sample path as $\{(u_1^n, y_1^n) \dots (u_L^n, y_L^n)\}$. We indicate by $A_{n,j,\ell}$ the set of sample paths in $A_{n,j}$ corresponding to the ℓ^{th} node. Finally, we define

$$\begin{aligned} N_{i,\ell} &= \inf\left\{ n : \log \frac{\mathbb{P}(H^* = h_{i_n^*} | y_\ell^{n+1}, u_\ell^{n+1})}{\max_{j \neq i_n^*} \mathbb{P}(H^* = h_j | y_\ell^{n+1}, u_\ell^{n+1})} \geq \rho_{i,\ell} \left| \log c \right| \right\} \\ &= \inf\left\{ n : \sum_{k=1}^{n} \log \frac{p_{i,\ell}^{u_{k,\ell}}(y_{k,\ell})}{\max_{j \neq i} p_{j,\ell}^{u_{k,\ell}}(y_{k,\ell})} \geq \rho_{i,\ell} \left| \log c \right| \right\}. \end{aligned}$$

The proof consists of two parts. First, we write $\mathbb{P}_i^{\mathcal{D}}(\hat{H} \neq h_i)$ as the probability of a countable union of disjoint sets of sample paths. An upper bound on this probability then follows from an upper bound on the probability of these disjoint sets, in conjunction with the union bound. Second, we upper bound $\mathbb{E}_i^{\mathcal{D}}[N]$ by the sum of the expected time required to reach the threshold in (2.8) at node ℓ for $H^* = h_i$, and the expected delay between the time of reaching the threshold and the time when the final decision is taken in favor of hypothesis h_i at the fusion center. We then show that these expectations are the same at all the nodes, so that (2.26) follows.

Consider the probability $\mathbb{P}_i^{\mathcal{D}}(\hat{H} = h_j)$. This is the same as the probability of the countable

union of disjoint sets $A_{n,j}$. Thus, for all $j \neq i$, we have

$$\begin{aligned}
\mathbb{P}_i^{\mathcal{D}}(A_{n,j}) &= \int_{A_{n,j}} \prod_{\ell=1}^{L} \prod_{k=1}^{n} p_{i,\ell}^{u_{k,\ell}}(y_{k,\ell})\, dy_{1,\ell}(u_{1,\ell}) \dots\ dy_{n,\ell}(u_{n,\ell}) \\
&\overset{(a)}{=} \prod_{\ell=1}^{L} \int_{A_{n,j,\ell}} \prod_{k=1}^{n} p_{i,\ell}^{u_{k,\ell}}(y_{k,\ell})\, dy_{1,\ell}(u_{1,\ell}) ..... dy_{n,\ell}(u_{n,\ell}) \\
&\overset{(b)}{\leq} \prod_{\ell=1}^{L} \int_{A_{n,j,\ell}} c^{\rho_{j,\ell}} \prod_{k=1}^{n} p_{j,\ell}^{u_{k,\ell}}(y_{k,\ell})\, dy_{1,\ell}(u_{1,\ell}) \dots\ dy_{n,\ell}(u_{n,\ell}) \\
&\overset{(c)}{=} c \prod_{\ell=1}^{L} \int_{A_{n,j,\ell}} \prod_{k=1}^{n} p_{j,\ell}^{u_{k,\ell}}(y_{k,\ell})\, dy_{1,\ell}(u_{1,\ell}) \dots\ dy_{n,\ell}(u_{n,\ell}) \\
&= c \prod_{\ell=1}^{L} \mathbb{P}_j^{\mathcal{D}}(\hat{H} = h_j \text{ at sample } n \text{ at } \ell^{th} \text{ sensor}) \\
&= c\, \mathbb{P}_j^{\mathcal{D}}(\hat{H} = h_j \text{ at sample } n), \qquad (2.97)
\end{aligned}$$

where (a) follows from the definition of $A_{n,j,\ell}$; (b) follows from the definition of $N_{i,\ell}$; (c) follows from $\sum_{\ell=1}^{L} \rho_{j,\ell} = 1$. Now, we can bound $\mathbb{P}_i^{\mathcal{D}}(\hat{H} \neq h_i)$ as follows

$$\begin{aligned}
\mathbb{P}_i^{\mathcal{D}}(\hat{H} \neq h_i) &= \sum_{j \neq i} \mathbb{P}_i^{\mathcal{D}}(\hat{H} = h_j) = \sum_{j \neq i} \sum_{n=1}^{\infty} \mathbb{P}_i^{\mathcal{D}}(A_{n,j}) \\
&\leq \sum_{j \neq i} \sum_{n=1}^{\infty} c\, \mathbb{P}_j^{\mathcal{D}}(\hat{H} = h_j \text{ at sample } n) \\
&= \sum_{j \neq i} c\, \mathbb{P}_j^{\mathcal{D}}(\hat{H} = h_j) \leq c\,(M-1), \qquad (2.98)
\end{aligned}$$

where the first inequality follows from (2.97). This proves part (i) of the theorem.

Let us now define

$$\tau(N_{i,\ell}) = \sup \Big\{ n \ : \sum_{k=N_{i,\ell}+1}^{N_{i,\ell}+n} \log \frac{p_{i,\ell}^{u_{k,\ell}}(y_{k,\ell})}{\max_{j \neq i} p_{j,\ell}^{u_{k,\ell}}(y_{k,\ell})} < 0 \Big\}.$$

The condition in (2.4) is satisfied for threshold in (2.8) at the ℓ^{th} node for all $n > N_{i,\ell} + \tau(N_{i,\ell})$,

yielding

$$\begin{aligned} N &\leq \max_{1\leq \ell \leq L} (N_{i,\ell} + \tau(N_{i,\ell}) + 3(M+1)) \\ &\leq \max_{1\leq \ell \leq L} N_{i,\ell} + \sum_{\ell=1}^{L} \tau(N_{i,\ell}) + 3(M+1), \end{aligned} \tag{2.99}$$

where if $C \geq M$, then three time steps are needed to communicate v_ℓ, I and the halting message; otherwise at most $3(M+1)$ time steps are needed to communicate this information.

Taking the expectation of both sides, we have

$$\mathbb{E}_i^{\mathcal{D}}[N] \leq \mathbb{E}_i \left[\max_{1\leq \ell \leq L} N_{i,\ell} \right] + \sum_{\ell=1}^{L} \mathbb{E}_i[\tau(N_{i,\ell})] + 3(M+1). \tag{2.100}$$

We now bound the terms on the right-hand side of (2.100). Since each node performs the Chernoff test individually, for all $\ell \in [L]$ and $i \in [M]$, there exist two constants $K_{i,\ell} > 0$ and $b_{i,\ell} > 0$ such that for all $\epsilon > 0$ and $n \geq (1+\epsilon)|\log(c)|/I(i)$, we have [45, Lemma 2]

$$\mathbb{P}_i(N_{i,\ell} \geq n) \leq K_{i,\ell} e^{-b_{i,\ell} n}. \tag{2.101}$$

Thus, we have

$$\mathbb{E}_i[N_{i,\ell}] = (1+o(1)) \frac{|\log c|}{I(i)}, \tag{2.102}$$

which is independent of ℓ. Using (2.101), for all $\epsilon > 0$ and $n \geq (1+\epsilon)|\log(c)|/I(i)$, we also have that

$$\begin{aligned} \mathbb{P}_i\left(\max_{1\leq \ell \leq L} N_{i,\ell} \geq n \right) &\leq \sum_{\ell=1}^{L} \mathbb{P}_i(N_{i,\ell} \geq n) \\ &\leq L K_i e^{-b_i n}, \end{aligned} \tag{2.103}$$

where $K_i = \max_\ell K_{i,\ell}$ and $b_i = \min_\ell b_{i,\ell}$. For all $r \geq 1$, we have the bound on the r^{th} moment

of $\max_{1\leq \ell \leq L} N_{i,\ell}$,i.e.

$$\mathbb{E}_i\left[\left(\max_{1\leq \ell \leq L} N_{i,\ell}\right)^r\right] \leq \left((1+o(1))\frac{|\log c|}{I(i)}\right)^r. \tag{2.104}$$

Now, we bound the higher moments of $\tau(N_{i,\ell})$. Let N^* be the time instance such that for all $n \geq N^*$, the local decision $\hat{H}_\ell$ at node ℓ is h^*, i.e., $\hat{H}_\ell = h^*$. Using [45, Lemma 1], there exists $K > 0$ and $b > 0$ such that

$$\mathbb{P}_i(N^* \geq n) \leq k \exp(-bn), \tag{2.105}$$

which implies $\mathbb{P}_i(N^* < \infty) = 1$. Then, node ℓ following time N^* selects the actions in an i.i.d. fashion according to the probability mass function given by (2.3).

Let $G_{n,\ell}$ be the joint cumulative distribution function of the variables $(y_{n,\ell}, u_{n,\ell})$ at round n and node ℓ for the Chernoff test. Also, let F_ℓ be the joint cumulative distribution function of $(y_{n,\ell}, u_{n,\ell})$ under the true hypothesis h^* when the actions are selected according to $Q^\ell_{h^*}$ (see (2.3)) at each round at sensor ℓ. Then, for all $n > N^*$, we have $G_{n,\ell} = F_\ell$. Since $\mathbb{P}_i(N^* < \infty) = 1$, it follows that the distribution $G_{n,\ell}$ converges to F_ℓ.

Given that for all n, $(y_n, u_n) \sim F_\ell$ are i.i.d. random variables, we have that

$$\mathbb{E}_i\Big[\log(p^{u_{i,\ell}}_{i,\ell}(y_{k,\ell})/\max_{j\neq i} p^{u_{i,\ell}}_{j,\ell}(y_{k,\ell}))\Big] = v_{i,\ell} > 0. \tag{2.106}$$

Additionally, using (2.106), finiteness of the $r+1^{st}$ moment of log-likelihood ratio for $r \geq 1$, and by Corollary 10.1, Lemma 5 and (2.202) in Appendix 2.12.3, we have that

$$\mathbb{E}_i[\tau(N_{i,\ell}))^r] < \infty, \tag{2.107}$$

where the expectation is with respect to F_ℓ.

We now note that (2.107) also holds when the expectation is with respect to $G_{n,\ell}$. To show

this claim, first observe that $\mathbb{E}_i[\tau(N_{i,\ell}))^r]$ is upper bounded by the two terms at the right-hand side of (2.182) in Corollary 10.1. The first term is bounded, since the KL-divergence between any two probability measures is finite. The second term can be split into two summations, one for $1 \leq n \leq N^*$, and the other for $n \geq N^* + 1$. The first summation is finite since $N^* < \infty$ a.s., and the probability is at most one. By using Lemma 5 in Appendix 2.12.3 and $G_{n,\ell} = F_\ell$, the second summation is also finite. It follows that (2.107) holds for the SCT.

Since $\mathbb{E}[\log(p_{i,\ell}^{u_{i,\ell}}(y_{k,\ell})/\max_{j\neq i} p_{j,\ell}^{u_{i,\ell}}(y_{k,\ell}))]^2$ is finite, using (2.107), $\mathbb{E}_i[\tau(N_{i,\ell})]$ on the right-hand side of (2.100) is finite and independent of c. Now, combining equation (2.100), (2.104) and the finiteness of $\mathbb{E}_i[\tau(N_{i,\ell})]$, as $c \to 0$, we get (2.26). Thus, part (ii) of the theorem is proved.

Now,

$$\mathbb{E}_i^{\mathcal{D}}[N^r] \leq \mathbb{E}_i\Big[\Big(\max_{1\leq\ell\leq L} N_{i,\ell} + \sum_{\ell\in[L]} \tau(N_{i,\ell}) + 1\Big)^r\Big]. \tag{2.108}$$

The moments of $\sum_{\ell\in[L]} \tau(N_{i,\ell})$ are finite and independent of c. Hence, the dominant term, dependent on c, in the expansion of the right-hand side of (2.108) is given only by $\max_{1\leq\ell\leq L} N_{i,\ell}$. Using (2.107) and (2.104), it follows that as $c \to 0$, we have

$$\mathbb{E}_i^{\mathcal{D}}[N^r] \leq \left((1 + o(1))\frac{|\log c|}{I(i)}\right)^r, \tag{2.109}$$

which proves part (iv) of the theorem. □

Proof of Theorem 4

Proof. Combining Theorems 2 and 3, we have that (2.30), (2.31) and (2.33) follow immediately. We then turn to the proof of (2.34).

For all $\ell \in [L]$, given that hypothesis h_i is true, we have that as $c \to 0$, the probability of

incorrect detection tends to zero. It follows that $\hat{H} = h_i$ and

$$\begin{aligned}\mathbb{E}_i^{\mathcal{D}}[N] &= (1 + o(1))\frac{|\log c|}{I(i)} \\ &= \mathbb{E}_i^{\mathcal{D}}[N_{i,\ell}],\end{aligned} \tag{2.110}$$

where the last equality follows from (2.102). Thus, as $c \to 0$, all the nodes reach the same local decision, on average, at the same time, and the average number of messages that each node sends to the fusion center to communicate this local decision is one. It follows that, as $c \to 0$, the total expected communication overhead is four: two in the initialization phase, one to communicate the local decision, and one to receive the halting message. □

Proof of Theorem 5

Proof. Let $B_{n,j}$ be the set of sample paths where the final decision $\hat{H}$ is initiated in favor of h_j at the n^{th} step, and we indicate a single sample path as $\{(u_1^n, y_1^n) \dots (u_L^n, y_L^n)\}$. We indicate by $B_{n,j,\ell}$ the set of sample paths in $B_{n,j}$ corresponding to the ℓ^{th} node. N^c denotes the time taken to terminate the initialization phase of CCT. Now, we define the two times associated with the test phase of CCT:

$$\begin{aligned}T_{i,\ell} &= \inf\left\{n : \log\frac{\mathbb{P}(H^* = h_{i_n^*}|y_\ell^{n+1}, u_\ell^{n+1})}{\max_{j\neq i_n^*}\mathbb{P}(H^* = h_j|y_\ell^{n+1}, u_\ell^{n+1})} \geq \hat{\rho}_{i,\ell}^{(N^c)}\,|\log c|\right\} \\ &= \inf\left\{n : \sum_{k=1}^{n}\log\frac{p_{i,\ell}^{u_{k,\ell}}(y_{k,\ell})}{\max_{j\neq i}p_{j,\ell}^{u_{k,\ell}}(y_{k,\ell})} \geq \hat{\rho}_{i,\ell}^{(N^c)}\,|\log c|\right\},\end{aligned}$$

and

$$\tau(T_{i,\ell}) = \sup\left\{n \;:\; \sum_{k=T_{i,\ell}+1}^{T_{i,\ell}+n}\log\frac{p_{i,\ell}^{u_{k,\ell}}(y_{k,\ell})}{\max_{j\neq i}p_{j,\ell}^{u_{k,\ell}}(y_{k,\ell})} < 0\right\}.$$

The proof consists of two parts. First, we write $\mathbb{P}_i^{\mathcal{C}}(\hat{H} \neq h_i)$ as the probability of a countable union of disjoint sets of sample paths. An upper bound on this probability then follows from an upper bound on the probability of these disjoint sets, in conjunction with the union

bound. Second, $\mathbb{E}_i^{\mathcal{C}}[N]$ is dependent on the time required to reach and detect the consensus during the initialization phase, the time required to reach the threshold in (2.19) in the test phase, and the time required to reach and detect that the nodes have reached a common preference about a hypothesis in the stopping phase. The stopping time N can be bounded as

$$N \leq N^c + \max_{1\leq \ell \leq L}(T_{i,\ell} + \tau(T_{i,\ell})) + N^s, \tag{2.111}$$

where N^s is the time taken to detect the common preference about the hypothesis in the stopping phase of CCT.

Consider the probability $\mathbb{P}_i^{\mathcal{C}}(\hat{H} = h_j)$. This is the same as the probability of the countable union of disjoint sets $B_{n,j}$. Thus, for $j \neq i$, we can write

$$\begin{aligned}
\mathbb{P}_i^{\mathcal{C}}(B_{n,j}) &= \int_{B_{n,j}} \prod_{\ell=1}^{L}\prod_{k=1}^{n} p_{i,\ell}^{u_{k,\ell}}(y_{k,\ell})\, dy_{1,\ell}(u_{1,\ell}) \dots\ dy_{n,\ell}(u_{n,\ell}) \\
&\overset{(a)}{=} \prod_{\ell=1}^{L} \int_{B_{n,j,\ell}} \prod_{k=1}^{n} p_{i,\ell}^{u_{k,\ell}}(y_{k,\ell})\, dy_{1,\ell}(u_{1,\ell}) ..... dy_{n,\ell}(u_{n,\ell}) \\
&\overset{(b)}{\leq} \prod_{\ell=1}^{L} \int_{B_{n,j,\ell}} c^{\hat{\rho}_{j,\ell}^{(n)}} \prod_{k=1}^{n} p_{j,\ell}^{u_{k,\ell}}(y_{k,\ell})\, dy_{1,\ell}(u_{1,\ell}) \dots\ dy_{n,\ell}(u_{n,\ell}) \\
&\overset{(c)}{\leq} c^{I(i)/(I(i)+c)} \prod_{\ell=1}^{L} \int_{B_{n,j,\ell}} \prod_{k=1}^{n} p_{j,\ell}^{u_{k,\ell}}(y_{k,\ell})\, dy_{1,\ell}(u_{1,\ell}) \dots\ dy_{n,\ell}(u_{n,\ell}) \\
&= c^{I(i)/(I(i)+c)} \prod_{\ell=1}^{L} \mathbb{P}_j^{\mathcal{C}}(\hat{H} = h_j \text{ at sample } n \text{ at } \ell^{th} \text{ sensor}) \\
&= c^{I(i)/(I(i)+c)} \mathbb{P}_j^{\mathcal{C}}(\hat{H} = h_j \text{ at sample } n),
\end{aligned} \tag{2.112}$$

where (a) follows from the definition of $B_{n,j,\ell}$; (b) follows from the definition of $T_{i,\ell}$; (c) follows from the facts that using Theorem 1 and (2.18), we have

$$I(i)/(I(i)+c) \leq \sum_{\ell=1}^{L} \hat{\rho}_{j,\ell}^{(n)} \leq I(i)/(I(i)-c), \tag{2.113}$$

and $c < 1$. Now, we can bound $\mathbb{P}_i^{\mathcal{C}}(\hat{H} \neq h_i)$ as follows

$$\begin{aligned}\mathbb{P}_i^{\mathcal{C}}(\hat{H} \neq h_i) &= \sum_{j \neq i} \mathbb{P}_i^{\mathcal{C}}(\hat{H} = h_j) = \sum_{j \neq i} \sum_{n=1}^{\infty} \mathbb{P}_i^{\mathcal{C}}(B_{n,j}) \\ &\leq \sum_{j \neq i} \sum_{n=1}^{\infty} c^{I(i)/(I(i)+c)}\, \mathbb{P}_j^{\mathcal{C}}(\hat{H} = h_j \text{ at sample } n) \\ &= c^{I(i)/(I(i)+c)}\,(M-1), \end{aligned} \tag{2.114}$$

where the inequality in the chain follows by (2.112). This proves part (i) of the theorem.

Let us bound the time N^c required to terminate the initialization phase. Since matrix W in (2.10) is row stochastic using (2.13) and the graph $\mathcal{G}(\mathcal{N}, \mathcal{E})$ is connected, the ergodic coefficient $\eta(W) \in (0, 1)$ using Lemma 1. It follows from [54] that for all $k, n \in \mathbb{N}$ and $\ell, j \in [L]$, we have

$$e_{\ell,j}^{k+n} \preceq \big(1 - \eta(W^n)\big) e_{\ell,j}^{k}, \tag{2.115}$$

where $e_{\ell,j}^k = |\hat{I}_\ell^k - \hat{I}_j^k|$. Now, if the initialization phase reaches uniformly local c-consensus at time instance k_0, then using (2.18), for all $\ell, j \in [L]$, we have

$$e_{\ell,j}^{k_0} \preceq c\mathbf{1}_{1\times M}. \tag{2.116}$$

Thus, there exists $k' \in \mathbb{N}$ such that $h^{\mathcal{G}} k' \leq k_0 \leq h^{\mathcal{G}}(k'+1)$. Using (2.115), for all $\ell, j \in [L]$, we have

$$\begin{aligned} e_{\ell,j}^{k_0} &\preceq e_{\ell,j}^{h^{\mathcal{G}}k'} \\ &\overset{(a)}{\preceq} \Big(1 - \eta\big(W^{h^{\mathcal{G}}}\big)\Big)^{k'} e_{\ell,j}^{0} \\ &\overset{(b)}{\preceq} \Big(1 - \eta\big(W^{h^{\mathcal{G}}}\big)\Big)^{k'} I, \end{aligned} \tag{2.117}$$

where (a) follows from $\hat{I}^{h^{\mathcal{G}}k'} = W^{h^{\mathcal{G}}} \hat{I}^{h^{\mathcal{G}}(k'-1)}$ and Lemma 1, and (b) follows from the fact that

for all $\ell, j \in [L]$, we have $e^0_{\ell,j} \preceq I$. Since for all $\ell, j \in [L]$, we have $e^{k_0}_{\ell,j} \preceq c\mathbf{1}_{1\times M}$, using (2.117), we have

$$\left(1 - \eta(W^{h^{\mathcal{G}}})\right)^{k'} I \preceq c, \tag{2.118}$$

and

$$k^{'} \leq \frac{\log(c/\max_{j\in[L]} I(j))}{\log\left(1 - \eta(W^{h^{\mathcal{G}}})\right)}. \tag{2.119}$$

Since $k_0 \leq h^{\mathcal{G}}(k'+1)$, we have

$$k_0 \leq h^{\mathcal{G}}\left(\frac{\log(c/\max_{j\in[L]} I(j))}{\log\left(1 - \eta(W^{h^{\mathcal{G}}})\right)} + 1\right). \tag{2.120}$$

Now, let k_d be the time to detect the local c-consensus. From [235], we have

$$k_d \leq h^{\mathcal{G}}\left(\frac{-\log(d^{\mathcal{G}})}{\log\left(1 - \eta(W^{h^{\mathcal{G}}})\right)} + 1\right) + L + 1. \tag{2.121}$$

Now, the time N^c for initialization phase is bounded as follows

$$\begin{aligned} N^c &\leq k_0 + k_d \\ &\leq h^{\mathcal{G}}\left(\frac{\log(c/\max_{j\in[L]} I(j))}{\log\left(1 - \eta(W^{h^{\mathcal{G}}})\right)} + 1\right) \\ &\quad + h^{\mathcal{G}}\left(\frac{-\log(d^{\mathcal{G}})}{\log\left(1 - \eta(W^{h^{\mathcal{G}}})\right)} + 1\right) + L + 1. \end{aligned} \tag{2.122}$$

The expected time of the test phase of CCT is at most $\mathbb{E}_i\left[\max_{1\leq\ell\leq L}(T_{i,\ell} + \tau(T_{i,\ell}))\right]$. Combining (2.104), (2.107), and the fact that $|\hat{I}_\ell(i) - I(i)| \leq c$ using Theorem 1 and (2.18), as $c \to 0$, we have

$$\mathbb{E}_i\left[\max_{1\leq\ell\leq L}(T_{i,\ell} + \tau(T_{i,\ell}))\right] \leq (1 + o(1))\frac{|\log c|}{I(i) - c}. \tag{2.123}$$

Now, we compute the time for the decision phase of CCT. The network will reach the

final decision for all $n > \max_{1\leq\ell\leq L} \tau(T_{i,\ell}) + k_r$, where k_r is the time taken by the termination message $m_i^{(3)}$ to reach every node after its initiation at any node. Thus, the time N^s of the decision phase is bounded above as

$$N^s \leq \max_{1\leq\ell\leq L} \tau(T_{i,\ell}) + k_r.$$

Therefore, we have

$$\mathbb{E}_i[N^s] \leq \sum_{\ell=1}^{L} \mathbb{E}_i[\tau(N_{i,\ell})] + \mathbb{E}_i[k_r]. \tag{2.124}$$

Using (2.107), the term $\mathbb{E}_i[\tau(T_{i,\ell})]$ at the right-hand side of (2.124) is finite and independent of c. Additionally, $k_r < d^{\mathcal{G}} + 1$. Thus, $\mathbb{E}_i[N^s]$ is finite and independent of c.

Combining equations (2.122), (2.123), and the finiteness of $\mathbb{E}_i[N^s]$, we get that (2.38) holds as $c \to 0$, proving part (ii) of the theorem.

Now we derive the bounds for the higher moments of the decision time N. We have

$$\begin{aligned} N &\leq N^c + \max_{1\leq\ell\leq L} (T_{i,\ell} + \tau(T_{i,\ell})) + N^s \\ &\leq N^c + \max_{1\leq\ell\leq L} (T_{i,\ell}) + 2 \max_{1\leq\ell\leq L} \tau(T_{i,\ell}) + k_r \\ &\leq N^c + \max_{1\leq\ell\leq L} (T_{i,\ell}) + 2 \sum_{\ell\in[L]} \tau(T_{i,\ell}) + k_r. \end{aligned} \tag{2.125}$$

Now, we present the bound on the r^{th} moment of each term in the right-hand side of (2.125). Using (2.122), N^c is bounded above by a constant. As $c \to 0$, we have

$$(N^c)^r \leq \left((1 + o(1)) \frac{h^{\mathcal{G}} \log(c/\max_{j\in[L]} I(j))}{\log(1 - \eta(W^{h^{\mathcal{G}}})} \right)^r. \tag{2.126}$$

Using (2.104) and the fact that $|\hat{I}_\ell(i) - I(i)| \leq c$, we have

$$\mathbb{E}_i \left[\max_{1\leq\ell\leq L} T_{i,\ell}^r \right] = \left((1 + o(1)) \frac{|\log c|}{I(i) - c} \right)^r. \tag{2.127}$$

Using (2.107), the higher moments of the third term in the right-hand side of (2.125) are finite and independent of c by definition of $\tau(T_{i,\ell})$. Additionally, $k_r \leq L+1 < \infty$. Now,

$$\mathbb{E}_i^{\mathcal{C}}[N^r] \leq \mathbb{E}_i\Big[N^c + \max_{1\leq \ell \leq L}(T_{i,\ell}) + 2\sum_{\ell\in[L]}\tau(T_{i,\ell}) + k_r\Big]^r. \tag{2.128}$$

The moments of $\sum_{\ell\in[L]}\tau(T_{i,\ell}) + k_r$ are finite and independent of c. The dominant terms, dependent on c, in the expansion of the right-hand side of (2.128) depend only on $N^c + \max_{1\leq\ell\leq L}(T_{i,\ell})$. Therefore, as $c \to 0$, we have

$$\mathbb{E}_i^{\mathcal{C}}[N^r] \leq \left((1+o(1))\left(\frac{h^{\mathcal{G}}\log(c/\max_{j\in[L]}I(j))}{\log(1-\eta(W^{h^{\mathcal{G}}}))} + \frac{|\log c|}{I(i)-c}\right)\right)^r, \tag{2.129}$$

which proves part (iv) of the theorem. □

Proof of Theorem 6

Proof. Combining Theorems 2 and 5, we have (2.42), (2.43) and (2.45) follow immediately. □

Decision Phase of CCT

Lemma 3. *If $d_\ell^N > L+1$, then there exists a time $k \leq N$ at which the local decision of all the nodes are the same, i.e.,* $\min_{j\in[L]} x_j^k \geq 1$. *This decision is the same as the local decision $\hat{H}_\ell^N$ of node ℓ at time N.*

Proof. At time N and node ℓ, if $d_\ell^N > L+1$, then for all $k \in \mathcal{N}_\ell$, we have $d_k^{N-1} > L$ and $x_k^{N-2} \geq L$. If the shortest distance between the node ℓ and j is $s_{\ell,j}$, then we have

$$d_j^{N-s_{\ell,j}} > L - s_{\ell,j} + 1. \tag{2.130}$$

Thus, for all $j \in [L]$, we have

$$d_j^{N-d^{\mathcal{G}}} > d_j^{N-s_{\ell,j}} + s_{\ell,j} - d^{\mathcal{G}} > 1, \tag{2.131}$$

since $s_{\ell,j} \leq d^{\mathcal{G}} \leq L$. This implies that for all $j \in [L]$, we have

$$x_j^{N-d^{\mathcal{G}}-1} \geq 1. \tag{2.132}$$

Thus, the first statement of the claim follows.

Now, we prove the second statement by contradiction. For all $j \in [L]$, let the decision at time $N - d^{\mathcal{G}} - 1$ be $\hat{H}_j^{N-d^{\mathcal{G}}-1} = h'$ which is different from $\hat{H}_\ell^N$. At sensor ℓ, let the decision change from h' to $\hat{H}_\ell^N$ at time n. Then,

$$N - d^{\mathcal{G}} - 1 < n \leq N. \tag{2.133}$$

Therefore, $x_\ell^n = 1$, which implies

$$d_\ell^{n+1} \leq 2. \tag{2.134}$$

Now,

$$\begin{aligned} d_\ell^N &\leq d_\ell^{n+1} + N - n - 1 \\ &\leq 2 + N - n - 1 \\ &< 2 + d^{\mathcal{G}} \\ &\leq 2 + L. \end{aligned} \tag{2.135}$$

However, $d_\ell^N \geq L + 2$ by the statement of the Lemma. Hence, by contradiction, we conclude that the second statement of our claim holds. □

Proof of Theorem 7

Proof. The proof of the theorem is exactly along the same lines as the proof of Theorem 3. The key difference lies in the computation of the constant $\rho_{i,\ell}$. Due to quantization into Q

sub-intervals, we have

$$v_{i,\ell} - \Delta(\max_i I(i), Q) \leq \lfloor v_{i,\ell} \rfloor \leq v_{i,\ell}, \tag{2.136}$$

which implies

$$v_{i,\ell} - f(Q)/L \leq \lfloor v_{i,\ell} \rfloor \leq v_{i,\ell}. \tag{2.137}$$

Using (2.137), we have that $\lfloor I(i) \rfloor$ is

$$I(i) - f(Q) \leq \lfloor I(i) \rfloor \leq I(i), \tag{2.138}$$

which implies that $\rho_{i,\ell}$ in (2.52) verifies

$$\frac{v_{i,\ell}}{I(i)} \leq \rho_{i,\ell} \leq \frac{v_{i,\ell}}{I(i) - f(Q)}. \tag{2.139}$$

For part (i), using the lower bound from (2.139) in (2.97), we have

$$\begin{aligned} \mathbb{P}_i^{\mathcal{D}}(A_{n,j}) &\leq c^{\sum_\ell v_{i,\ell}/I(i)}\, \mathbb{P}_j^{\mathcal{D}}(\hat{H} = h_j \text{ at sample } n) \\ &= c\, \mathbb{P}_j^{\mathcal{D}}(\hat{H} = h_j \text{ at sample } n). \end{aligned} \tag{2.140}$$

Now, the result in part (i) follows similar to (2.98).

For part (ii), (iii) and (iv), since $C > \log_2 M$, the local decisions can be communicated at each time instance. Using (2.138) and the assumption that $f(Q) \leq I(i)$, for all $r \geq 1$, similar to (2.104), we have

$$\begin{aligned} \mathbb{E}_i \left[\Big(\max_{1 \leq \ell \leq L} N_{i,\ell} \Big)^r \right] &\leq \left((1 + o(1)) \frac{|\log c|}{\lfloor I(i) \rfloor} \right)^r \\ &\leq \left((1 + o(1)) \frac{|\log c|}{I(i) - f(Q)} \right)^r. \end{aligned} \tag{2.141}$$

Now, similar to (2.109), we have

$$\mathbb{E}_i^{\mathcal{D}}[N^r] \leq \left((1+o(1))\frac{|\log c|}{I(i)-f(Q)}\right)^r. \tag{2.142}$$

Hence, part (ii), (iii) and (iv) follows. □

Proof of Theorem 8

Proof. The proof of the theorem is along the same lines as Theorem 5. The key difference lies in the computation of the constant $\hat{\rho}_{i,\ell}$.

Firstly, we will upper bound and lower bound $\hat{\rho}_{i,\ell}$ in terms of $I(i)$ and $g(\tilde{Q}, c, \alpha)$. Since Assumptions 1 and 2 hold, and the graph $\mathcal{G}$ is strongly connected, using [160, Proposition 5], the time k_0 to reach *local* c-consensus is

$$k_0 \leq 2\frac{L^2}{\alpha}\log(\min(\tilde{Q}^2, L^4/c^2)\max_j I^2(j)) + 1. \tag{2.143}$$

Using $C > \log_2(L+2)$ and (2.121), time k_d to detect the consensus is

$$k_d \leq h^{\mathcal{G}}\left(\frac{-\log(d^{\mathcal{G}})}{\log\left(1-\eta(W^{h^{\mathcal{G}}})\right)}+1\right)+L+1. \tag{2.144}$$

Now, using [160, Proposition 7] and the fact that the average decreases by at most $1/\tilde{Q}$ in each iteration of consensus, for all $i \in [M]$ and $\ell \in [L]$, the error in estimation $\lfloor \hat{I}_\ell(i) \rfloor$ at the end of initialization phase is at most

$$\begin{aligned}|\lfloor \hat{I}_\ell(i) \rfloor - I(i)| \leq \frac{L}{\tilde{Q}}(k_0+k_d) &\leq \frac{L}{\tilde{Q}}\left(2\frac{L^2}{\alpha}\log(\min(\tilde{Q}^2, L^4/c^2)\max_j I^2(j))+1\right)\\ &\quad + \frac{L}{\tilde{Q}}\left(h^{\mathcal{G}}\left(\frac{-\log(d^{\mathcal{G}})}{\log\left(1-\eta(W^{h^{\mathcal{G}}})\right)}+1\right)+L+1\right)\\ &= g(\tilde{Q}, c, \alpha).\end{aligned} \tag{2.145}$$

This implies

$$\frac{v_{i,\ell}}{I(i)+g(\tilde{Q},c,\alpha)} \leq \hat{\rho}_{i,\ell} \leq \frac{v_{i,\ell}}{I(i)-g(\tilde{Q},c,\alpha)}. \tag{2.146}$$

Thus, for part (i), using the lower bound from (2.146) in (2.112), we have

$$\mathbb{P}_i^{\mathcal{C}}(B_{n,j}) \leq c^{I(i)/(I(i)+g(\tilde{Q},c,\alpha))}\mathbb{P}_j^{\mathcal{C}}(\hat{H}=h_j \text{ at sample } n). \tag{2.147}$$

The result in part (i) follows similar to (2.114).

For part (ii), (iii) and (iv), the time required in the initialization phase is at most k_0+k_d and can be bounded using (2.143) and (2.144). For test phase, using (2.146) and the assumption that $g(\tilde{Q},c,\alpha) < I(i)$, for all $r \geq 1$, similar to (2.127), we have

$$\begin{aligned}\mathbb{E}_i\left[\max_{1\leq\ell\leq L} T_{i,\ell}^r\right] &= \left((1+o(1))\frac{|\log(c)|}{\min_{\ell\in[L]}\hat{I}_\ell(i)}\right)^r \\ &\leq \left((1+o(1))\frac{|\log(c)|}{I(i)-g(\tilde{Q},c,\alpha)}\right)^r. \end{aligned}\tag{2.148}$$

For decision phase, since $C > \log_2(L+2)+\log_2 M$, the local decisions and d_ℓ^n can be communicated at each time instance. Hence, the time N^s of decision phase is finite from Theorem 5. Similar to (2.129), the result follows by combining the time for all the three phases of CCT. □

Proof of Theorem 9

Proof. For part (i), since the vectors v_ℓ and I are communicated using Q levels of quantization, the proof is exactly same as that of part (i) in Theorem 7.

For part (ii), (iii) and (iv), the additional delays in comparison to the setting in Theorem 7 are the time to communicate the vectors $\lfloor v_\ell \rfloor$ to the fusion center, the time to communicate vector $\lfloor I \rfloor$ to the nodes, and time to make a final decision given the same preferences about the hypothesis are reached at the nodes. Since each link is active with probability $1-\epsilon$, the expected

time to communicate the vectors $\lfloor v_\ell \rfloor$ and $\lfloor I \rfloor$ is at most

$$\frac{2L}{1-\epsilon}. \tag{2.149}$$

Given that all the local preferences are reached at the nodes, i.e., $n > \max_i \tau(N_{i,\ell})$, the probability that all these preferences are received at the same time instances at the fusion center is $(1-\epsilon)^L$, which corresponds to all the links being active at the same time. The expected decision time following $n > \max_i \tau(N_{i,\ell})$ is

$$\frac{1}{(1-\epsilon)^L}. \tag{2.150}$$

Combining the delays in (2.149) and (2.150), and the results in Theorem 7, the statement of the theorem follows. □

Proof of Lemma 2

Proof. For part (i), for all $\ell \in [L]$, we have

$$\begin{aligned}\sum_{j=1}^{L} w_{\ell,j}(n) &= w_{\ell,\ell}(n) + \sum_{j\neq\ell} w_{\ell,j}(n) \\ &= 1 - \beta \sum_{j\neq\ell} \mathbf{1}((j,\ell) \in \mathcal{E}(n)) + \beta \sum_{j\neq\ell} \mathbf{1}((j,\ell) \in \mathcal{E}(n)) \\ &= 1. \end{aligned} \tag{2.151}$$

Since $w_{i,j}(n) = w_{j,i}(n)$, we have

$$\sum_{\ell=1}^{L} w_{\ell,j}(n) = 1. \tag{2.152}$$

Hence, $W(n)$ is doubly stochastic.

For part (ii), for all $(i,j) \in \mathcal{E}(n)$, we have

$$w_{i,j}(n) \geq \min\left(\beta, 1 - \beta \sum_{\ell\neq i} \mathbf{1}((i,\ell) \in \mathcal{E}(n))\right). \tag{2.153}$$

Thus, for all $(i,j) \in \mathcal{E}(n)$, we have

$$w_{i,j}(n) > \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \tag{2.154}$$

where $\mathcal{D}(\mathcal{G}) = \max_s \sum_{j \neq s} \mathbf{1}((j,s) \in \mathcal{E})$.

For part (iii), note that the eigenvalues of $\bar{L}(n)$ are non-negative and recall that the sum of the diagonal elements of $\bar{L}(n)$ is equal to the sum of the its eigenvalues [87]. Let λ is an eigenvalue of $\bar{L}(n)$. Then, we have

$$\lambda \leq 2|\mathcal{E}|, \tag{2.155}$$

because $|\mathcal{E}(n)| \leq |\mathcal{E}|$. The eigenvalues of $W(n)$ are of the form $1 - \beta\lambda$. Since $0 < \lambda \leq 2|\mathcal{E}|$, for all $0 < \beta < 1/(2|\mathcal{E}|)$, we have

$$0 < 1 - \beta\lambda < 1, \tag{2.156}$$

which implies $R(W(n)) < 1$. To show (2.71), let $\bar{\lambda}$ be an eigenvalue of $W(n) - (\mathbf{1}_{L\times 1}\mathbf{1}_{1\times L})/L$ and not an eigenvalue of $W(n)$. We have

$$\begin{aligned}
&\det\left(\bar{\lambda}U_{L\times L} - W(n) + \frac{\mathbf{1}_{L\times 1}\mathbf{1}_{1\times L}}{L}\right) \\
&\overset{(a)}{=} \det(\bar{\lambda}U_{L\times L} - W(n))\left(1 + \frac{\mathbf{1}_{1\times L}(\bar{\lambda}U_{L\times L} - W(n))^{-1}\mathbf{1}_{L\times 1}}{L}\right) \\
&\overset{(b)}{=} \det(\bar{\lambda}U_{L\times L} - W(n))\left(1 + \frac{\mathbf{1}_{1\times L}\mathbf{1}_{L\times 1}}{(\bar{\lambda}-1)L}\right) \\
&= \det(\bar{\lambda}U_{L\times L} - W(n))\left(1 + \frac{1}{(\bar{\lambda}-1)}\right),
\end{aligned} \tag{2.157}$$

where $\det(\cdot)$ denotes the determinant of a matrix, (a) follows from the fact that $(\bar{\lambda}U_{L\times L} - W(n))$ is non-singular because $\bar{\lambda}$ is not an eigenvalue of $W(n)$, and exploits the matrix determinant lemma [87], namely if A is a non-singular matrix of dimension $L \times L$ and c and d are column

vectors of dimension $L \times 1$, then

$$\det(A + cd^T) = \det(A)(1 + d^T A^{-1} c), \tag{2.158}$$

(b) follows from the fact that $(\bar{\lambda} U_{L\times L} - W(n))$ is non-singular and doubly stochastic, which implies

$$\begin{aligned}(\bar{\lambda} U_{L\times L} - W(n))\mathbf{1}_{L\times 1} &= \bar{\lambda}\mathbf{1}_{L\times 1} - \mathbf{1}_{L\times 1} \\ &= (\bar{\lambda} - 1)\mathbf{1}_{L\times 1}. \end{aligned} \tag{2.159}$$

Since $\bar{\lambda}$ is an eigenvalue of $W(n) - (\mathbf{1}_{L\times 1}\mathbf{1}_{1\times L})/L$, we have

$$\left(1 + \frac{1}{(\bar{\lambda} - 1)}\right) = 0, \tag{2.160}$$

which implies $\bar{\lambda} = 0$. Combining the facts that $\bar{\lambda} < 1$ and $R(W(n)) < 1$, the claim in (iii) follows. □

Proof of Theorem 10

Proof. The proof of the theorem is along the same lines as the proof of Theorem 8. The key difference is that, unlike $\hat{\rho}_{i,\ell}$, in this case $\hat{\rho}^{\epsilon}_{i,\ell}$ is a random variable, and the randomness is introduced by the time-varying configuration of the network due to ϵ-random packet erasures.

We derive the upper and lower bound on $\hat{\rho}^{\epsilon}_{i,\ell}$ with high probability in terms of $I(i)$ and $h(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon)$. First, for all n, we establish that $W(n)$ satisfies Assumption 1. Second, for all n, we show that the resulting graph $\mathcal{G}(\mathcal{V}, \mathcal{E}(n))$ is strongly connected with probability at least $1 - |\mathcal{E}|\epsilon$. Using these two results, similar to Theorem 8, we bound the time to reach consensus $K_0 + K_d$, which is now a random variable (see (2.143) and (2.144)), and the estimation error (see (2.145)). The rest of the proof is similar to that of Theorem 8.

For all n, given an edge $e \in \mathcal{E}$, the probability that $e \notin \mathcal{E}(n)$ is ϵ since the link failures

are independent and identically distributed across time and independent of other links. Thus, the probability that the graph $\mathcal{G}(\mathcal{V}, \mathcal{E}(n))$ is strongly connected is

$$\begin{aligned}\mathbb{P}(\mathcal{G}(\mathcal{V}, \mathcal{E}(n)) \text{ is strongly connected}) &\geq \mathbb{P}(\text{For all } e \in \mathcal{E}, \text{ we have } e \in \mathcal{E}(n)) \\ &\geq (1-\epsilon)^{|\mathcal{E}|} \\ &\geq 1 - |\mathcal{E}|\epsilon. \end{aligned} \tag{2.161}$$

Since Assumptions 1 and 2 hold, and the graph $\mathcal{G}(\mathcal{V}, \mathcal{E}(n))$ is strongly connected with probability at least $1 - |\mathcal{E}|\epsilon$, using Lemma 2,[160, Proposition 5] and (2.143), the number of time steps satisfying the property that $\mathcal{G}(\mathcal{V}, \mathcal{E}(n))$ is strongly connected and that are required to converge to *local* c-consensus is at most

$$\frac{2L^2 \log(\min(\tilde{Q}^2, L^4/c^2) \max_j I^2(j))}{\min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right)} + 1. \tag{2.162}$$

This along with (2.161) implies that $\mathbb{E}[K_0]$ to reach *local* c-consensus is

$$\mathbb{E}[K_0] \leq \frac{1}{(1-|\mathcal{E}|\epsilon)} \left(\frac{2L^2 \log(\min(\tilde{Q}^2, L^4/c^2) \max_j I^2(j))}{\min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right)} + 1 \right). \tag{2.163}$$

Now, similar to (2.144), we have that $\mathbb{E}[K_d]$ to detect consensus is

$$\mathbb{E}[K_d] \leq \frac{1}{(1 - |\mathcal{E}|\epsilon)} \left(h^{\mathcal{G}} \left(\frac{-\log(d^{\mathcal{G}})}{\log\left(1 - \eta(W^{h^{\mathcal{G}}})\right)} + 1 \right) + L + 1 \right). \tag{2.164}$$

To obtain the high probability bound on the estimation error of the vector I, let us introduce a sequence of Bernoulli i.i.d random variables $\{Z_n\}_{n=1}^{\infty}$ with probability of success $\mathbb{P}(Z_n = 1) = 1 - |\mathcal{E}|\epsilon$. Then, with probability one, we have

$$K_0 \leq \min \left\{ n \geq 1 : \sum_{k=1}^{n} Z_k > \frac{2L^2 \log(\min(\tilde{Q}^2, L^4/c^2) \max_j I^2(j))}{\min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right)} + 1 \right\}. \tag{2.165}$$

Let $\delta = 1/(1-|\mathcal{E}|\epsilon)$, and

$$N_0 = \frac{1}{(1-|\mathcal{E}|\epsilon)}\left(\frac{2L^2\log(\min(\tilde{Q}^2, L^4/c^2)\max_j I^2(j))}{\min(1-\mathcal{D}(\mathcal{G})\beta, \beta)} + 1\right).$$

Using Hoeffding's inequality [85], we have

$$\begin{aligned}
\mathbb{P}\Big(K_0 \geq N_0(1+\delta)\Big) &\overset{(a)}{\leq} \mathbb{P}\Big(\sum_{n=1}^{N_0(1+\delta)} Z_n \leq \frac{2L^2\log(\min(\tilde{Q}^2, L^4/c^2)\max_j I^2(j))}{\min(1-\mathcal{D}(\mathcal{G})\beta, \beta)} + 1\Big) \\
&= \mathbb{P}\Big(\sum_{n=1}^{N_0(1+\delta)} Z_n - N_0(1+\delta)(1-|\mathcal{E}|\epsilon) \leq -N_0(1+\delta)(1-|\mathcal{E}|\epsilon) \\
&\qquad \frac{2L^2\log(\min(\tilde{Q}^2, L^4/c^2)\max_j I^2(j))}{\min(1-\mathcal{D}(\mathcal{G})\beta, \beta)} + 1\Big) \\
&\overset{(b)}{=} \mathbb{P}\Big(\sum_{n=1}^{N_0(1+\delta)} Z_n - N_0(1+\delta)(1-|\mathcal{E}|\epsilon) \leq -N_0\delta(1-|\mathcal{E}|\epsilon)\Big) \\
&\leq \exp\left(-2\delta^2(1-|\mathcal{E}|\epsilon)^2 N_0(1+\delta)/(1+\delta)^2\right) \\
&= \exp\left(-2\delta^2(1-|\mathcal{E}|\epsilon)^2 N_0/(1+\delta)\right) \\
&= \exp\left(-2(1-|\mathcal{E}|\epsilon)N_0/(2-|\mathcal{E}|\epsilon)\right), \qquad (2.166)
\end{aligned}$$

where (a) follows from (2.165), and (b) follows from the definition of N_0.

Similarly, we can show that for $\delta = 1/(1-|\mathcal{E}|\epsilon)$ and

$$N_0' = \frac{1}{(1-|\mathcal{E}|\epsilon)}\left(h^{\mathcal{G}}\left(\frac{-\log(d^{\mathcal{G}})}{\log\left(1-\eta(W^{h^{\mathcal{G}}})\right)} + 1\right) + L + 1\right),$$

we have

$$\mathbb{P}(K_d > N_0'(1+\delta)) \leq \exp\left(-2(1-|\mathcal{E}|\epsilon)N_0'/(2-|\mathcal{E}|\epsilon)\right). \qquad (2.167)$$

Thus, similar to (2.145), using (2.166) and (2.167), we have that with probability one, the error

in the estimation of $\lfloor \hat{I}_\ell^\epsilon(i) \rfloor$ at the end of initialization phase is

$$|\lfloor \hat{I}_\ell^\epsilon(i) \rfloor - I(i)| \leq \frac{L}{\tilde{Q}}(K_0 + K_d). \tag{2.168}$$

This implies that using (2.166) and (2.167) , we have

$$\begin{aligned}
|\lfloor \hat{I}_\ell^\epsilon(i) \rfloor - I(i)| &\leq \frac{L}{\tilde{Q}}(K_0 + K_d) \\
&\leq \frac{L(1+\delta)}{\tilde{Q}(1-|\mathcal{E}|\epsilon)} \Bigg(\frac{2L^2 \log(\min(\tilde{Q}^2, L^4/c^2) \max_j I^2(j))}{\min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right)} + 1 \\
&\qquad + h^{\mathcal{G}} \left(\frac{-\log(d^{\mathcal{G}})}{\log\left(1 - \eta(W^{h^{\mathcal{G}}})\right)} + 1 \right) + L + 1 \Bigg) \\
&= g(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right))(2 - |\mathcal{E}|\epsilon)/(1 - |\mathcal{E}|\epsilon)^2 \\
&= h(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon),
\end{aligned} \tag{2.169}$$

with probability at least

$$\begin{aligned}
&1 - \exp\left(-\frac{2}{(2-|\mathcal{E}|\epsilon)} \frac{2L^2 \log(\min(\tilde{Q}^2, L^4/c^2) \max_j I^2(j))}{\min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right)} \right) \\
&\exp\left(-\frac{2}{(2-|\mathcal{E}|\epsilon)} \left(h^{\mathcal{G}} \left(\frac{-\log(d^{\mathcal{G}})}{\log\left(1 - \eta(W^{h^{\mathcal{G}}})\right)} + 1 \right) + L + 2 \right) \right) \\
&= 1 - \exp\left(-2\tilde{Q} g(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right))/L(2 - |\mathcal{E}|\epsilon)\right) \\
&= 1 - \exp(-2q(\tilde{Q}, c, \min\left(1 - \mathcal{D}(\mathcal{G})\beta, \beta\right), \epsilon),
\end{aligned} \tag{2.170}$$

since K_0 and K_d are independent.

Now, for part (i), using the lower bound from (2.169) in (2.112), we have $\mathbb{P}_i^{\mathcal{C}}(B_{n,j})$ is at

most

$$
\begin{aligned}
&(1-\exp(-2q(\tilde{Q},c,\min{(1-\mathcal{D}(\mathcal{G})\beta,\beta)},\epsilon))\\
&c^{I(i)/(I(i)+h(\tilde{Q},c,\min{(1-\mathcal{D}(\mathcal{G})\beta,\beta)},\epsilon))}\mathbb{P}_j^{\mathcal{C}}(\hat{H}=h_j \text{at sample} n)\\
&+\exp(-2q(\tilde{Q},c,\min{(1-\mathcal{D}(\mathcal{G})\beta,\beta)},\epsilon)\mathbb{P}_j^{\mathcal{C}}(\hat{H}=h_j \text{ at sample } n). \qquad (2.171)
\end{aligned}
$$

The result in part (i) follows similar to (2.114).

Consider next parts (ii), (iii) and (iv). For the consensus phase, the expected time $\mathbb{E}[K_0+K_d]$ required is upper bounded by the right-hand sides of (2.163) and (2.164). For the test phase, using the assumption $h(\tilde{Q},c,\min{(1-\mathcal{D}(\mathcal{G})\beta,\beta)},\epsilon) < I(i)$, for all $r \geq 1$, we have that (2.127) becomes

$$
\begin{aligned}
&\mathbb{E}_i\left[\max_{1\leq\ell\leq L} T_{i,\ell}^r \middle| \{\lfloor \hat{I}_\ell^\epsilon \rfloor\}_{\ell\in[L]}\right]\\
&=\left((1+o(1))\frac{|\log(c)|}{\min_{\ell\in[L]}\lfloor \hat{I}_\ell^\epsilon(i)\rfloor}\right)^r\\
&\overset{(a)}{\leq}\left((1+o(1))\frac{|\log(c)|}{I(i)-h(\tilde{Q},c,\min{(1-\mathcal{D}(\mathcal{G})\beta,\beta)},\epsilon)}\right)^r, \qquad (2.172)
\end{aligned}
$$

with probability at least

$$
(1-\exp{(-2q(\tilde{Q},c,\min{(1-\mathcal{D}(\mathcal{G})\beta,\beta)},\epsilon))}), \qquad (2.173)
$$

where (a) follows from (2.169). For the stopping phase, since $C > \log_2(L+2)+\log_2 M$, the local decisions and d_ℓ^n can be communicated at each time instance. Hence, the time N^s of the decision phase is finite from Theorem 5 and the fact that the probability the graph is strongly connected at each time instance is at least $(1-|\mathcal{E}|\epsilon) > 0$. Similar to (2.129), the result follows by combining the time for all the three phases of CCT. □

2.12.3 Proofs of Miscellaneous Results

In this section, we present results used to bound the time $\tau(N_{i,\ell})$ in Theorem 3 and 5 (see (2.107)). Let $X_1 \dots X_n$ be i.i.d. random variables and let the time

$$T = \sup\Big\{n \ : \sum_{k=1}^{n} X_k > 0\Big\}. \tag{2.174}$$

This is the last n at which

$$S_n > 0, \tag{2.175}$$

where $S_n = \sum_{k=1}^{n} X_k$, $n \geq 1$, and $S_0 = 0$.

Lemma 4. *For all $r \geq 1$, if $\mathbb{E}\big[|X_1|^{r+1}\big] < \infty$ and $\mathbb{E}\big[X_1\big] \leq -\mu_0 < 0$, then we have*

$$\mathbb{E}[T^r] \leq r\left(\frac{2}{\mu_0}\right)^r \mathbb{E}[(S^*)^r] + \sum_{k=1}^{\infty} rk^{r-1}\mathbb{P}\big(S_k + k\mu_0/2 > 0\big), \tag{2.176}$$

where $S^* = \max_{j\geq 0} S_j$.

Proof. We have

$$\begin{aligned}
\mathbb{E}[T^r] &\leq \sum_{k=1}^{\infty} rk^{r-1}\mathbb{P}(T \geq k) \\
&= \sum_{k=1}^{\infty} rk^{r-1}\mathbb{P}(\max_{j\geq k} S_j > 0) \\
&= \sum_{k=1}^{\infty} rk^{r-1}\mathbb{P}\Big(\max_{j\geq k}\big(S_j - S_k\big) + S_k > 0\Big) \\
&= \sum_{k=1}^{\infty} rk^{r-1}\mathbb{P}\Big(S^* + S_k > 0\Big),
\end{aligned} \tag{2.177}$$

where S^* is an independent copy of $\max_{j\geq 0} S_j$, therefore we loosely use the same symbol.

Now, along the same lines of proof as in [114, Theor. D], we have

$$\begin{aligned}
&\sum_{k=1}^{\infty} rk^{r-1}\mathbb{P}\Big(S^* + S_k > 0\Big) \\
&= \int_0^{\infty} \sum_{k=1}^{\lfloor 2\xi/\mu_0 \rfloor} rk^{r-1}\mathbb{P}\big(S_k > -\xi\big)\, d\mathbb{P}\big(S^* \le \xi\big) \\
&+ \int_0^{\infty} \sum_{k=\lfloor 2\xi/\mu_0 \rfloor + 1}^{\infty} rk^{r-1}\mathbb{P}\big(S_k + \mu_0 k/2 > \mu_0 k/2 - \xi\big) d\mathbb{P}\big(S^* \le \xi\big). \qquad (2.178)
\end{aligned}$$

The first integral at the right-hand side of (2.178) can be bounded as

$$\begin{aligned}
\int_0^{\infty} \sum_{k=1}^{\lfloor 2\xi/\mu_0 \rfloor} rk^{r-1}\mathbb{P}\big(S_k > -\xi\big)\, d\mathbb{P}\big(S^* \le \xi\big) &\le \int_0^{\infty} \sum_{k=1}^{\lfloor 2\xi/\mu_0 \rfloor} rk^{r-1}\, d\mathbb{P}\big(S^* \le \xi\big) \\
&\le \int_0^{\infty} r\,(2\xi/\mu_0)^r\; d\mathbb{P}\big(S^* \le \xi\big) \\
&= r\left(\frac{2}{\mu_0}\right)^r \mathbb{E}\big[(S^*)^r\big]. \qquad (2.179)
\end{aligned}$$

The second integral at the right-hand side of (2.178) can be bounded as

$$\begin{aligned}
&\int_0^{\infty} \sum_{k=\lfloor 2\xi/\mu_0 \rfloor + 1}^{\infty} rk^{r-1}\mathbb{P}\big(S_k + \mu_0 k/2 > \mu_0 k/2 - \xi\big)\, d\mathbb{P}\big(S^* \le \xi\big) \\
&\le \int_0^{\infty} \sum_{k=\lfloor 2\xi/\mu_0 \rfloor + 1}^{\infty} rk^{r-1}\mathbb{P}\big(S_k + \mu_0 k/2 > 0\big)\, d\mathbb{P}\big(S^* \le \xi\big) \\
&\le \int_0^{\infty} \sum_{k=1}^{\infty} rk^{r-1}\mathbb{P}\big(S_k + \mu_0 k/2 > 0\big)\, d\mathbb{P}\big(S^* \le \xi\big) \\
&\le \sum_{k=1}^{\infty} rk^{r-1}\mathbb{P}\big(S_k + \mu_0 k/2 > 0\big), \qquad (2.180)
\end{aligned}$$

where the first inequality follows from the fact that integration variable verifies $\xi \le \mu_0 k/2$.

The claim of the Lemma now follows by (2.179) and (2.180). □

Corollary 10.1. *Let $X_1, \ldots, X_n$ be i.i.d. random variables such that $\mathbb{E}[|X_1|^{r+1}] < \infty$ and $\mathbb{E}[X_1] \geq \mu_0 > 0$. Also, let $S_0 = 0$, $S_n = \sum_{k=1}^{n} X_k$, $n \geq 1$, and*

$$T = \sup\Big\{n \ : \sum_{k=1}^{n} X_k < 0\Big\}. \tag{2.181}$$

Then, for all $r \geq 1$, we have

$$\mathbb{E}[T^r] \leq r\Big(\frac{2}{\mu_0}\Big)^r \mathbb{E}\bigg[\Big(-\min_{j\geq 0} S_j\Big)^r\bigg] + \sum_{k=1}^{\infty} r k^{r-1} \mathbb{P}\big(S_k - k\mu_0/2 < 0\big). \tag{2.182}$$

Proof. The proof follows from replacing X_k by $-X_k$ in Lemma 4. □

Lemma 5. *Let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed random variables with zero mean and finite $(r+1)^{th}$ absolute moment, namely for all $r \geq 1$, we have $\mathbb{E}[|X|^{r+1}] < \infty$. Then, for all $r \geq 1$, we have*

$$\sum_{n=1}^{\infty} n^{r-1} \mathbb{P}\Big(\Big|\sum_{k=1}^{n} X_k\Big| > n\Big) < \infty. \tag{2.183}$$

Proof. The proof technique is borrowed from [60]. Event $A = \{|\sum_{k=1}^{n} X_k| > n\}$ is written as a subset of the union of three events i.e. $A \subset A_n^{(1)} \cup A_n^{(2)} \cup A_n^{(3)}$. We bound the probability of these three events, and show that for all $i \in [3]$, we have

$$\sum_{n=1}^{\infty} n^{r-1} \mathbb{P}(A_n^{(i)}) < \infty. \tag{2.184}$$

Thus, (2.183) follows from (2.184).

Let $2^i \leq n < 2^{i+1}$, with $i \geq 0$. The events $A_n^{(1)}$, $A_n^{(2)}$ and $A_n^{(3)}$ are defined as follows:

$$\begin{aligned}
A_n^{(1)} &= \{\text{There exists } k \leq n \text{ such that } |X_k| > 2^{i-2}\},\\
A_n^{(2)} &= \{\text{There exists at least two integers } k_1, k_2 \leq n \text{ such that}\\
&\qquad |X_{k_1}| > n^{4/5} \text{and } |X_{k_2}| > n^{4/5}\},\\
A_n^{(3)} &= \Big\{\Big|\sum_{k\in N'} X_k\Big| > 2^{i-2}\Big\},
\end{aligned}$$

where $N' = [n]\backslash\{k \leq n : |X_k| > n^{4/5}\}$. If the event $A_n^{(1)} \cup A_n^{(2)} \cup A_n^{(3)}$ does not occur, then we have

$$\Big|\sum_{k=1}^{n} X_k\Big| \leq 2^{i-2} + 2^{i-2} < n. \tag{2.185}$$

Hence, $A \subset A_n^{(1)} \cup A_n^{(2)} \cup A_n^{(3)}$, and $\mathbb{P}(A) \leq \mathbb{P}(A_n^{(1)}) + \mathbb{P}(A_n^{(2)}) + \mathbb{P}(A_n^{(3)})$. Therefore,

$$\sum_{n=1}^{\infty} n^{r-1}\mathbb{P}(A) \leq \sum_{n=1}^{\infty} n^{r-1}\mathbb{P}(A_n^{(1)}) + \sum_{n=1}^{\infty} n^{r-1}\mathbb{P}(A_n^{(2)}) + \sum_{n=1}^{\infty} n^{r-1}\mathbb{P}(A_n^{(3)}). \tag{2.186}$$

Now, we bound the probability of all three events at the right-hand side of the above equation.

Let $a_i = \mathbb{P}(|X_k| > 2^i)$. We have

$$\begin{aligned}
\sum_{i=0}^{\infty} 2^{i(r+1)-1} a_i &\overset{(a)}{\leq} \sum_{i=0}^{\infty} 2^{i(r+1)}(a_i - a_{i+1})\\
&\overset{(b)}{\leq} \mathbb{E}[|X_k|^{r+1}]\\
&\overset{(c)}{<} \infty,
\end{aligned} \tag{2.187}$$

where (a) follows from the fact that

$$\frac{1}{2}\sum_{i=0}^{\infty} 2^{i(r+1)} a_i \geq \frac{1}{2^{r+1}} \sum_{i=1}^{\infty} 2^{i(r+1)} a_i = \sum_{i=0}^{\infty} 2^{i(r+1)} a_{i+1}, \tag{2.188}$$

which implies

$$\sum_{i=0}^{\infty} 2^{i(r+1)} a_i - \frac{1}{2} \sum_{i=0}^{\infty} 2^{i(r+1)} a_i \geq \sum_{i=0}^{\infty} 2^{i(r+1)} a_{i+1}, \tag{2.189}$$

(b) follows from the definitions of a_i and $\mathbb{E}[|X_k|^{r+1}]$, after exploiting $\int_{y_1}^{y_2} y dy \geq \int_{y_1}^{y_2} y_1 dy$, and (c) follows from the assumption of the lemma. Thus, using (2.187), we have

$$\sum_{i=0}^{\infty} 2^{i(r+1)} a_i < \infty. \tag{2.190}$$

Now, we bound the probability of the event at the right-hand side of (2.186) that involves $A_n^{(1)}$

$$\begin{aligned}
\sum_{n=1}^{\infty} n^{r-1} \mathbb{P}(A_n^{(1)}) &= \sum_{n=1}^{\infty} n^{r-1} \mathbb{P}(\exists k \leq n : |X_k| > 2^{i-2} \text{ where } i \text{ verifies } 2^i \leq n < 2^{i+1}) \\
&\overset{(a)}{\leq} \sum_{n=1}^{\infty} n^r \mathbb{P}(|X_k| > 2^{i-2} \text{ where } i \text{ verifies } 2^i \leq n < 2^{i+1}) \\
&= \sum_{i=0}^{\infty} \sum_{2^i \leq n < 2^{i+1}} n^r a_{i-2} \\
&\leq \sum_{i=0}^{\infty} \sum_{2^i \leq n < 2^{i+1}} 2^{(i+1)r} a_{i-2} \\
&= \sum_{i=0}^{\infty} 2^{i(r+1)+r} a_{i-2} \\
&< \infty,
\end{aligned} \tag{2.191}$$

where (a) follows from the union bound and the fact that X_k are i.i.d, and the last inequality follows from (2.190).

Since the $(r+1)^{st}$ absolute moment is finite, for all $k \in \mathbb{N}$, there exists a finite constant $K > 0$ such that

$$\mathbb{P}(|X_k| \geq u) \leq K/u^{r+1}. \tag{2.192}$$

Now, we bound the probability of event $A_n^{(2)}$

$$\begin{aligned}\mathbb{P}(A_n^{(2)}) &\overset{(a)}{\leq} \sum_{1\leq k_1<k_2\leq n} \mathbb{P}(|X_{k_1}| > n^{4/5} \text{and } |X_{k_2}| > n^{4/5}) \\ &\overset{(b)}{\leq} n^2 \mathbb{P}(|X_1| > n^{4/5})\, \mathbb{P}(|X_2| > n^{4/5}) \\ &\overset{(c)}{\leq} K^2 n^2 n^{-4(r+1)/5} n^{-4(r+1)/5}, \end{aligned} \tag{2.193}$$

where (a) follows from the definition of the event and the union bound, (b) follows from the independence of the random variables and a bound on the number of possible combinations of k_1 and k_2, and (c) follows from (2.192). Therefore, we have

$$\begin{aligned}\sum_{n=1}^{\infty} n^{r-1} \mathbb{P}(A_n^{(2)}) &\overset{(a)}{\leq} \sum_{n=1}^{\infty} K^2 n^{r-1} n^2 n^{-8(r+1)/5} \\ &= \sum_{n=1}^{\infty} K^2 n^{-3r/5-3/5} \\ &\overset{(b)}{<} \infty, \end{aligned} \tag{2.194}$$

where (a) follows from (2.193), and (b) follows as $r \geq 1$.

Now, we bound the probability of event $A_n^{(3)}$. Let

$$X_k^+ = \begin{cases} X_k & |X_k| < n^{4/5}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.195}$$

There exist finite positive constants $K^{(1)}, K^{(2)}$, such that

$$\begin{aligned}
\mathbb{E}\Big[\Big|\sum_{k=1}^{n} X_k^+\Big|^{2r}\Big] &\overset{(a)}{\leq} \mathbb{E}\Big[\sum_{k=1}^{n} |X_k^+|^{2r}\Big] + \sum_{1\leq k_1,k_2\leq n} \mathbb{E}\big[|X_{k_1}^+|^{2r-1}\big]\mathbb{E}\big[|X_{k_2}^+|\big] + \dots \\
&\overset{(b)}{\leq} \mathbb{E}\Big[\sum_{k=1}^{n} n^{4(r-1)/5}|X_k^+|^{r+1}\Big] \\
&\qquad + \sum_{1\leq k_1,k_2\leq n} \mathbb{E}\big[n^{4(r-2)/5}|X_{k_1}^+|^{r+1}\big]\mathbb{E}\big[|X_{k_2}^+|\big] + \dots \\
&\overset{(c)}{\leq} K^{(1)} n^{4r/5} r^{2r}\Big(n^{-4/5} + n^{-8/5} + \dots\Big) \\
&\overset{(d)}{\leq} K^{(1)} n^{4r/5} r^{2r} \frac{n^{-4/5}}{1-n^{-4/5}} \\
&\leq K^{(2)} n^{4(r-1)/5}, \qquad (2.196)
\end{aligned}$$

where (a) follows from the multinomial expansion of $(\sum_{k=1}^{n}|X_k^+|)^{2r}$, and the independence of the random variables, (b) follows from (2.195), (c) follows from the following facts that

- $(r+1)^{st}$ absolute moment of X_k^+ is finite;
- the coefficient of multinomial expansion is of the form $2r!/(k_1!\dots k_n!)$ such that $k_1 + \dots + k_n = 2r$ and can be bounded as $\mathcal{O}(2r^{2r})$ independent of n;
- the largest coefficient of n in the expansion is $n^{4r/5}$ present in the first term in (b);
- the remaining coefficient of n will form a finite geometric progression, more specifically $n^{-4/5}, n^{-8/5}, \dots$;

and (d) follows from the fact that sum of the geometric progression $n^{-4/5}, n^{-8/5}, \dots$ can be bounded by $n^{-4/5}/(1-n^{-4/5})$.

Thus, using (2.196), there exists $K^{(3)} > 0$ such that

$$\mathbb{P}\Big(\Big|\sum_{k=1}^{n} X_k^+\Big| > n/8\Big) \leq \frac{K^{(3)} n^{4(r-1)/5}}{n^{2r}}, \qquad (2.197)$$

and

$$\begin{aligned}\mathbb{P}(A_n^{(3)}) &= \mathbb{P}\Big(\Big|\sum_{k=1}^{n} X_k^+\Big| > 2^{i-2}\Big)\\ &\overset{(a)}{\leq} \mathbb{P}\Big(\Big|\sum_{k=1}^{n} X_k^+\Big| > n/8\Big)\\ &\overset{(b)}{\leq} K^{(3)} n^{4(r-1)/5} n^{-2r}, \end{aligned} \tag{2.198}$$

where (a) follows by $n/8 < 2^{i-2}$, and (b) follows from (2.197). Thus, we have

$$\begin{aligned}\sum_{n=1}^{\infty} n^{r-1}\mathbb{P}(A_n^{(3)}) &\overset{(a)}{\leq} \sum_{n=1}^{\infty} n^{r-1} K^{(3)} n^{4(r-1)/5} n^{-2r}\\ &= \sum_{n=1}^{\infty} K^{(3)} n^{-r/5} n^{-9/5} \overset{(b)}{<} \infty, \end{aligned} \tag{2.199}$$

where (a) follows from (2.198), and (b) from the convergence of summation for $r \geq 1$. Finally, using (2.191), (2.194) and (2.199), we have that (2.183) follows. □

Now, we combine the results in Corollary 10.1 and Lemma 5. Let $X_1, \ldots, X_n$ be i.i.d. random variables with $\mathbb{E}[X_1] = \mu_x > 0$, and $\mathbb{E}[|X_1|^{r+1}] < \infty$ for all $r \geq 1$. Let $S_0 = 0$, $S_n = \sum_{k=1}^{n} X_k$ for $n \geq 1$, and $T = \sup\Big\{n \,:\, S_n < 0\Big\}$. Using (2.182), we have

$$E[T^r] \leq r\Big(\frac{2}{\mu_x}\Big)^r \mathbb{E}\Big[\Big(-\min_{j\geq 0} S_j\Big)^r\Big] + \sum_{k=1}^{\infty} r k^{r-1}\mathbb{P}\big(S_k - k\mu_x/2 < 0\big). \tag{2.200}$$

The first term at the right-hand side of (2.200) is finite because of the assumptions $\mu_x > 0$ and $\mathbb{E}[|X_1|^{r+1}] < \infty$ [114]. The second term can be bounded as follows

$$\begin{aligned}\sum_{k=1}^{\infty} r k^{r-1}\mathbb{P}\big(S_k - k\mu_x/2 < 0\big) &= \sum_{k=1}^{\infty} r k^{r-1}\mathbb{P}\big(2k - 2S_k/\mu_x > k\big)\\ &\leq \sum_{k=1}^{\infty} r k^{r-1}\mathbb{P}\Big(\Big|2k - 2S_k/\mu_x\Big| > k\Big) < \infty, \end{aligned} \tag{2.201}$$

where the last inequality follows by Lemma 5 applied to the zero-mean i.i.d. variables $\{2 - 2X_i/\mu_x\}_{i=1}^{\infty}$. Thus, we arrive at

$$\mathbb{E}[T^r] < \infty. \tag{2.202}$$

Chapter 3

Bounded Knapsack Bandits in crowd-sourcing systems

3.1 Introduction

Crowdsourcing systems (CS) have emerged as a valuable tool for several organizations to outsource a variety of tasks to a population of diverse workers at low cost. Some of the key players in the crowdsourcing market include for example *Amazon Mechanical Turk, Upwork, Freelancer* and *uTest*. In these systems, guaranteeing the quality of the work remains a key challenge, due to the limited a priori information about the ability of the workers. Thus, there is interest in developing automated methods for the collection and aggregation of information from the workers, incentive schemes to hire expert workers, and schemes for determining the quality of the tasks being done.

CS research has mostly focused on distributed methods where there is very limited interaction between workers and task master. The interaction is typically limited to the assignment of a gold-set of tasks to evaluate workers' performance prior to the assignment of the actual set of tasks, and does not provide a way to continuously monitor the quality of the work in real time. Dishonest workers can perform well on a gold-set of tasks and, not being evaluated on-line on a competitive basis, underperform during the actual working phase. Alternatively, the gold set can be mixed with all the assigned tasks in a way that the workers cannot distinguish between them. This is helpful to detect underperforming workers, but it wastes resources, and does not ensure

continuous monitoring of the quality of the work.

In this paper, we develop a notion of *Limited-information Crowdsourcing Systems* (LCS) that is desirable from both the task master and workers perspective. In LCS, workers express their interest in doing the tasks, quote their charges per task, and provide an upper limit on the number of tasks they are willing to perform. The tasks can then be assigned in burst or one-by-one to the workers, as long as the workers' constraints are satisfied. Given these constraints, unlike traditional CS, the workers do not need to be assigned all of their tasks at the same time. The workers' selection policy is not limited to be of the form "take-it" or "leave-it," but it can include workers who are still available after having completed a certain number of tasks, and that may be assigned additional tasks at a later time. This eliminates the requirement of having gold-set of tasks, and allows the task master to continuously monitor the quality of the work and assign tasks based on the estimated workers' ability, thus creating a competitive environment. Additionally, the workers are incentivized to perform tasks satisfactorily in order to maximize their earnings, while satisfying their load constraint.

This new formulation also poses new challenges. In our setting, the workers' selection algorithm needs to balance an exploration-exploitation trade-off, since the workers' ability is initially unknown to the task master and is learned on-line. This trade-off is not considered in traditional CS due to the limited interaction between the workers and the task master, but it is a classic one in the field of Multi-Armed Bandits (MAB) [124]. This is a class of problems dealing with decision making under uncertainty, where the actions have rewards that have to be learned through observations. Thus, the main challenge in LCS is to determine an efficient workers' selection scheme and to estimate of the abilities of the workers. To exploit the similarity of LCS with MAB, we reformulate the LCS problem in terms of a Bounded Knapsack Problem (BKP) that is equivalent to an arm-limited, budget-limited MAB. Given a strategy to estimate the workers' ability in real time, we use the B-KUBE algorithm developed in [213] for workers' selection. This algorithm has previously only been evaluated experimentally, and we provide provable performance guarantees, showing that its expected regret is $O(\log B)$, where B is the

maximum available budget. Since it has been shown in [16] that the expected regret for any algorithm is at least $\Omega(\log B)$, our results imply that B-KUBE is order optimal. Thus, we close the gap in the literature of arm-limited, budget-limited MAB by providing the first order optimal bounds of an algorithm in the current MAB setup. We then formalize the notion of workers' ability and propose an online strategy to estimate it. We also experimentally evaluate B-KUBE in conjunction with our strategy for estimating the workers' ability, showing that it outperforms other state-of-the-art MAB algorithms applied in the same setting. Thus, the contributions of the work are two fold: providing an optimal scheme for a MAB setup and using it in a crowdsourcing setting conjunction with an estimation scheme.

The organization of the paper is as follows: Section 2 describes the problem formulation; Section 3 discusses related work; Section 4 describes usage of B-KUBE for workers' selection and gives its performance guarantees; section 5 describes a strategy for estimating the workers' performance in real time; section 6 provides experimental evaluation of of B-KUBE in conjunction with this strategy; section 7 concludes the work.

3.2 Problem Formulation

We consider a labeling task in LCS, but this formulation can be easily modified to accommodate a different type of work. We assume the task master has a budget B and needs to label data with one of L labels. There are K workers interested in performing the labeling tasks. For every $k \in [K]$, the number of evaluations a worker can perform is limited by M_k and the cost of each evaluation is c_k. The objective of the task master is to minimize the average classification error

$$\epsilon = \frac{1}{T} \sum_i \mathbb{P}(\hat{l}_i \neq l_i^*) \tag{3.1}$$

where $\hat{l}_i$ and l_i^* are the predicted label and true label of the task i respectively, and T is the total number of labeling tasks. This is a common measure of performance considered in crowdsourcing systems works [103, 102, 109]. Thus, letting x_k be the number of evaluations assigned to each

worker, we define the LCS problem as follows

$$\begin{aligned} &\min \epsilon \text{ s.t.} \\ &\sum_k x_k c_k \leq B, \\ &\forall k \in [K] : 0 \leq x_k \leq M_k, \\ &\text{and } x_k \text{ is an integer.} \end{aligned} \tag{3.2}$$

We now reformulate the problem in (3.2) as a Bounded Knapsack problem (BKP). Assume that the measure of a worker's performance is given by a value contribution v_k. This value contribution is a measure of information contributed by a worker to the system after each evaluation. Minimizing ϵ in the LCS problem is then analogous to maximizing the aggregated value contributions in the following BKP

$$\begin{aligned} &\max_{\{x_k\}} \sum_k x_k v_k \text{ s.t.} \\ &\sum_k x_k c_k \leq B, \\ &\forall k \in [K] : 0 \leq x_k \leq M_k, \\ &\text{and } x_k \text{ is an integer.} \end{aligned} \tag{3.3}$$

The key benefit of the reformulation to BKP is that it provides an insight on the optimal aggregation of two different attributes, cost and value contribution, of the workers. Despite this equivalent formulation, the original LCS problem cannot be solved using standard BKP techniques. The value contributions typically are assumed to be known in BKP [107], while they need to be estimated in our setting. Nevertheless, the problem in (3.3) is also equivalent to an arm-limited, budget-limited stochastic MAB problem, whose expected rewards correspond to the unknown value contributions.

In a stochastic MAB problem, there are K arms of a single "bandit" machine. Pulling of

each arm delivers a reward that is independently drawn from an unknown distribution. An agent chooses to pull arms with the goal of maximizing the expected sum of the rewards received over a sequence of pulls.

We consider a popular stochastic model, from the literature of CS, for modeling the workers' responses. In this model, a worker k can be assigned a task multiple times and the correct label is predicted each time with probability p_k independent of the past responses of the worker about the task [103, 102, 83, 1, 246, 109, 213]. Given a task i, for all workers $k \in [K]$, we assume that the probability of predicting any incorrect label is the same for all labels independent of the task i and true label l_i^*, namely for all $\hat{l}_{i,k} \in [L]$ we have $\mathbb{P}(\hat{l}_{i,k} \neq l_i^*) = (1 - p_k)/(L - 1)$, where $\hat{l}_{i,k}$ is the predicted label of task i by the worker k. We also assume that the value contribution of a worker remains the same irrespective of the true label, namely for all $i \in [T]$ and $l_i^* \in [L]$ we have $v_k(l_i^*) = v_k$. These assumptions are only made for ease of presentation of our estimation strategy for value contributions and all the theoretical results provided in the paper do not rely on them.

The workers in LCS are equivalent to arms in MAB, and the task master plays the same role as the agent in MAB. The value contributions of the workers are analogous to the rewards of the arms. However, while the reward realization is immediately known after each pull, value contributions need to be estimated as the worker's ability in a real LCS scenario. Since M_k in LCS corresponds to a limit on the number of times an arm can be pulled, and c_k corresponds to the cost of pulling each arm in MAB, it follows that our problem corresponds to an arm-limited, budget-limited MAB where the realizations of the rewards depend on the workers' ability.

The regret of an algorithm A for a given budget B is defined as:

$$R^A(B) = v^*(B) - v^A(B)$$

where $v^*(B)$ is the optimal solution of the BKP in (3.3) and $v^A(B)$ is the aggregated value contributions using algorithm A.

3.3 Related Work

Several heuristic algorithms have been proposed for labeling tasks in CS, however, the performance of these inference algorithms is typically intractable [96, 229, 248]. In [103], an algorithm was proposed for the evaluation of homogeneous labeling tasks, i.e., all the tasks are equally difficult to label. It was proved that the algorithm is order optimal in the number of evaluations required per task required to obtain a desired classification error, when the number of tasks and workers tends to infinity. Thus, the work concluded that using an adaptive algorithm for task assignment has no gain in traditional CS. The model studied in [103] was generalized to heterogeneous labeling tasks in [83]. In this case, the authors showed that adaptive assignment of tasks leads to significant gains both in theory and practice. Unlike our work, the solution in this work is limited to weighted MV and binary labeling of the tasks. In addition, [109] presented tight achievable lower bounds for heterogeneous labeling tasks and proposed an order optimal scheme. For a similar model, [239] exploits the notions of iterative improvement and redundancy for translation tasks outsourced to CS. The work in [84] proposed an online task assignment scheme based on exploration and exploitation for heterogenous tasks. Their system model is budget constrained by assigning a limit on the number of evaluations for each task.

All of the above works consider equal incentives for all the workers and minimize the number of evaluations required per task. However, in a real life scenario a more efficient worker would expect higher incentives for his or her work. Our model allows for different costs per worker, and plans the assignment of the tasks accordingly. Additionally, the model also accounts for a maximum number of tasks that can be performed by a worker.

The Multi-Armed Bandits (MAB) problem is closely related to our crowdsourcing problem. A variety of budget constrained models have been studied in the MAB setup [32, 77, 10]. These works consider a budget-limited exploration in the initial phase followed by a cost-free exploitation phase. However, in a real world setting such as the one considered in LCS, the exploitation phase is not free of cost. This limitation is addressed in the budget-limited MAB

problem, where both the exploration and exploitation phase are limited by a single budget. This model also considers different costs for arm selection. Two different policies were proposed in this setting, called ϵ-first policy and KUBE [210, 211]. However, they did not consider a limit on the number of times an arm can be pulled, which is analogous to limiting the number of tasks a worker can do in LCS.

Later, the ϵ-first policy in [210] was extended to an arm-limited, budget-limited MAB and the regret of this new policy is $O(B^{2/3})$, where B is the budget [213]. However, the lower bound on the regret for any algorithm is of the order $\Omega(\log B)$ [16]. It follows that the extended ϵ-first policy is not optimal. Additionally, the KUBE algorithm was also extended to arm-limited, budget-limited MAB [213]. However, the work does not provide any theoretical performance analysis for this new algorithm, called B-KUBE. In this paper, we show that B-KUBE is indeed order optimal, achieving the lower bound presented in [16].

Other extensions of the MAB setup to CS have been considered in the literature. The workers' selection criteria for different cost of workers is studied in [1]. Two schemes were proposed for learning the ability of the labelers for equal and unequal incentives, respectively [55, 246]. Unlike our model, their system is not task limited by the workers and budget limited by the task master. BTASC is a workers' selection scheme proposed for spatial CS, however, it does not have any theoretical guarantees[150]. It is sub-optimal compared to BKUBE as it does not account for different costs paid to the workers. Also,the computation complexity of the scheme is $O(BK^2)$, whereas, the computation complexity for B-KUBE is $O(BK\log(K))$.

Also, there has been a large amount of work on bandits with knapsack [20, 52, 83, 25, 5]. In [20, 52, 83, 25], the work focuses on unbounded multidimensional knapsack problem in MAB, whereas, our work studies the bounded knapsack problem (BKP). In other words, the setup of these works do not consider a limit on the number of tasks that can be performed by each worker. In [20] and [83], workers arrive sequentially, and the workers' selection policy has to be of the form "take-it" or "leave-it". Therefore, unlike LCS, no worker is accessible later for task assignment once left. In [5], the work assumes that the constraints of the knapsack problem form

Algorithm 4. Bounded KUBE algorithm

Initialization: $n = 1$; $B_n = B$; For all k, $m_k = M_k$
while selecting a worker is feasible **do**
  if $n \leq K$ **then**
    Initialization Phase: assign $i(n) = n$
  else
    $\{m^*_{k,n}\}$ =greedyAlgoForBKP($\hat{v}_k, m_k, n, B_n$)
    Choose $i(n)$ with $\mathbb{P}(i(n) = j) = \frac{m^*_{j,n}}{\sum_k m^*_{k,n}}$
  end if
  Assign the task to $i(n)$
  Update the value contribution $\hat{v}_{i(n)}$ of $i(n)$
  $B_{n+1} = B_n - c_{i(n)}$
  $m_{i(n)} = m_{i(n)} - 1$
  $n = n + 1$
end while

a simplex. Therefore, the focus is on a perfectly convex knapsack problem. Unlike our work, this problem setup does not capture the limit on the number of tasks that can be performed by each worker which is an integer programming problem. Additionally, upper confidence bounds proposed in [5, 52] are different than the one used in B-KUBE. Extension of the policy proposed in [25] to BKP setting is of the form of Bounded ϵ-F policy which is suboptimal with respect to BKUBE, and its performance bounds cannot be improved [213].

3.4 Workers' Selection

We perform workers' selection using B-KUBE, which is described in Algorithm 4, where n denotes the iteration for worker's selection, B_n is the remaining budget before the n^{th} iteration, m_k is the remaining number of tasks a worker can perform, and $i(n)$ is the worker selected in the n^{th} iteration.

In each iteration, the task master checks the feasibility of worker's selection, i.e., whether there exists a $k \in [K]$ such that $c_k \leq B_n$ and $m_k > 0$. The first K iterations of B-KUBE constitute the initialization phase, where all the workers are selected once. For the remaining iterations, B-KUBE selects a worker j with probability $m^*_{j,n}/\sum_k m^*_{k,n}$, where $m^*_{j,n}$ is the number of selections of worker j proposed by the density-ordered greedy algorithm (DGA) for BKP at

Algorithm 5. Density Ordered Greedy Algorithm for BKP

Function name: greedyAlgoForBKP
Input: $\hat{v}_k, m_k, n, B_n$
Output:$\{m^*_{k,n}\}$
Initialization: $\hat{w}_k = \hat{v}_k + \sqrt{\frac{2\log(n)}{M_k - m_k}}$, $m^*_{k,n} = 0\ \forall k$
$\hat{e} = \{e_1, \ldots, e_k\}$ is the list of $(\hat{w}_k, c_k, m^*_{k,n}, m_k)$ sorted in decreasing order with respect to $\hat{w}_k/c_k$
$c = 0$
for $j = 1$ to K **do**
  if $c + \hat{e}(c_j) \leq B_n$ **then**
    assign task to j^{th} worker in $\hat{e}$
    $\hat{e}(m^*_{j,n}) = \min\left(\hat{e}(m_j), \lfloor \frac{B(n)-c}{\hat{e}(c_j)} \rfloor\right)$
    $c = c + \hat{e}(m^*_{j,n})\hat{e}(c_j)$
  else
    $\hat{e}(m^*_{j,n}) = 0$
  end if
end for

the n^{th} iteration.

DGA for BKP is described in Algorithm 5. It gives the number of selections of the workers for the remaining budget B_n. The algorithm computes the upper confidence bound value contribution $\hat{w}_k$, using the estimated value contribution $\hat{v}_k$, as

$$\hat{w}_k = \hat{v}_k + \sqrt{\frac{2\log(n)}{M_k - m_k}}, \tag{3.4}$$

and utilizes the entire B_n to select the workers as many times as possible, taking into account their individual limit m_k at the n^{th} iteration. The workers are selected in decreasing order of their estimated efficiencies $\hat{e}_k = \hat{w}_k/c_k$.

To analyze the performance of B-KUBE, we assume that the budget $\sum_k c_k < B \leq \sum_k c_k M_k$, the value contribution v_k has support in $[0, 1]$, and the cost $c_k \geq 1, \forall k \in [K]$. All results can easily be generalized using an appropriate scaling factor.

We start by recalling some results from the literature of BKP that are useful in our setting.

The BKP formulation in (3.3) can be relaxed to the linear problem LP-BKP

$$\begin{aligned} &\max_{\{x_k\}} \sum_k x_k v_k \\ &\text{such that } \sum_k x_k c_k \leq B, \\ &\forall k \in [K] : 0 \leq x_k \leq M_k. \end{aligned} \tag{3.5}$$

The following lemma provides the optimal workers' selection strategy for LP-BKP.

Lemma 6. [107]. *If the workers are sorted in decreasing order of their efficiencies* $e_k = v_k/c_k$*, where* $e_1 \geq e_2 \geq \ldots \geq e_K$*, then the optimal workers' selection strategy for LP-BKP is*

$$x_k^* = \begin{cases} M_k & \forall k = 1, 2, \ldots, s-1, \\ (B - \sum_{k=1}^{s-1} c_k M_k)/c_s & k = s, \\ 0 & \forall k = s+1, \ldots, K, \end{cases} \tag{3.6}$$

where the splitting worker s *is such that* $\sum_{k=1}^{s-1} c_k M_k \leq B$ *and* $\sum_{k=1}^{s} c_k M_k > B$*. The maximum aggregated value contribution is*

$$v_{LP-BKP}^* = \sum_{k=1}^{s} x_k^* v_k. \tag{3.7}$$

Letting v_{BKP}^* be the maximum aggregated value contributions that can be obtained from BKP and v' be the aggregated value contribution corresponding to the selection strategy $\lfloor x^* \rfloor = (x_1^*, x_2^*, \ldots, \lfloor x_s^* \rfloor, 0, 0 \ldots)$, by Lemma 6 we have

$$v' \leq v_{BKP}^* \leq v_{LP-BKP}^* \leq v' + v_s. \tag{3.8}$$

The key idea for obtaining a regret bound for B-KUBE is now to determine the number

of times a worker k is selected more than the number of selections of worker k as proposed by $\lfloor x^* \rfloor$. This will provide a bound on the regret of B-KUBE assuming $\lfloor x^* \rfloor$ is the optimal workers' selection strategy. This bound can then be combined with (3.8) to obtain the regret bound for B-KUBE.

It is worth pointing out the main challenges for the theoretical evaluation of B-KUBE compared to that of KUBE. In the KUBE setup, the computation of the regret bound simply corresponds to determining the expected number of times the most efficient worker is not selected. In the B-KUBE setup, the optimal selection of workers is not limited to a single most efficient worker, and a simplification like the one for KUBE is not possible. We overcome this difficulty by assuming that a feasible solution of BKP is the optimal selection strategy, and bounding the sub-optimal workers' selection based on this assumption. The other challenge is that the selection of the splitting worker s in $\lfloor x^* \rfloor$ is not always optimal. We solve this challenge by giving a bound on the expected number of times a worker k is selected more than the number of selections of worker k as proposed by $\lfloor x^* \rfloor$, as follows

Theorem 11. *For a given budget B, let B-KUBE perform N iterations. Assume that $\lfloor x^* \rfloor$ is the optimal selection strategy for the workers. Then, the expected number of times a worker k is selected more than the number of selections proposed by $\lfloor x^* \rfloor$ is*

$$\mathbb{E}\big[N_k(N)|N\big] \leq \left(\frac{8}{\min\{Q_{\min}^2, d_s^2\}} + \left(\frac{C_{max}}{C_{min}}\right)^2 \right) \log N + \frac{\pi^2}{3} + 1, \tag{3.9}$$

where

$$\begin{aligned} Q_{min} &= \min_{k \notin I^* \cup \{s\}} \big|e_k - e_s\big| \\ &= \min_{k \notin I^* \cup \{s\}} \big|v_k/c_k - v_s/c_s\big|, \end{aligned} \tag{3.10}$$

I^* is the set of the top $s-1$ workers, arranged in decreasing order of their efficiencies e_k, s is the splitting worker, $d_s = |v_{s-1}/c_{s-1} - v_s/c_s|$, $C_{max} = \max_{k\in[K]} c_k$ and $C_{min} = \min_{k\in[K]} c_k$.

From Theorem 11, it follows that assuming $\lfloor x^* \rfloor$ is the optimal selection strategy, using B-KUBE the selection of sub-optimal workers grows only logarithmically with N and we can conclude that B-KUBE favors the selection of workers as proposed by $\lfloor x^* \rfloor$. Additionally, Q_{min} and d_s measure the minimum separation between the optimal and sub-optimal selections, hence, they are the leading constants of $\log(N)$ in Theorem 11. Intuitively, it is more difficult to identify the optimal selection strategy $\lfloor x^* \rfloor$ if the abilities of the workers at the boundary of the optimal and sub-optimal selections are close. Theorem 11 recovers the result of the stochastic bandits, which are neither budget limited nor arm limited, with an additional constant factor of one in the leading term $\log(N)$ [16]. The minimum separation between the optimal and sub-optimal selections reduces to the same measure as proposed in [16].

Finally, the following theorem provides the regret bound for B-KUBE.

Theorem 12. *The expected regret for B-KUBE is* $O(\log(B))$.

The lower bound on the regret is $\Omega(\log N)$, where N is the total number of iterations [16]. In a budget-limited scenario, the number of iterations N is $\Theta(B)$, since $N \in [B/C_{max}, B/C_{min}]$. It follows that the lower bound on the regret in a budget-limited scenario is $\Omega(\log B)$ and B-KUBE is order optimal for arm-limited, budget-limited MAB.

3.5 Value Contributions of Workers

At each step n, workers' selection policy discussed in the previous section is dependent on the realization of $i(n)^{th}$ worker's value contribution for the update of its empirical estimate $\hat{v}_{i(n)}$. Therefore, we now focus on the determination of the ability of the workers in terms of value contributions, and propose a strategy for estimating the value contribution in real time.

Let the inference function $f_k(l, \hat{l})$ denote the contribution of the k^{th} worker to the label l when $\hat{l}$ is the label predicted by the k^{th} worker. Then, for all $l \in [L]$, the accumulated contribution

to the label l after M evaluations of task i is

$$s_{i,l} = \sum_{n=1}^{M} \sum_{k=1}^{K} f_k(l, \hat{l}_i^{(n)}) y_{k,n}, \tag{3.11}$$

where $y_{k,n}$ is an indicator function which is unity if the n^{th} evaluation of the task is performed by the k^{th} worker, and $\hat{l}_i^{(n)}$ is the predicted label of task i at n^{th} evaluation. The decision rule is

$$\hat{l}_i = \arg\max_{l \in [L]} s_{i,l}. \tag{3.12}$$

The inference function $f_k(.,.)$ is assumed to be non-negative, and bounded for all $k \in [K]$. Any generalized inference rule for labeling task is captured by (3.11) and (3.12). Special cases include majority voting, weighted majority voting and Maximum A Posteriori (MAP) decision rule.

Two key properties play an important role in the design of the inference function. First, the function should account for the characteristics of an individual worker. For example, if a worker is expected to confuse between the two labels, then the contribution of the inference function to them should be similar when one of these labels is predicted. This knowledge can be acquired from the prior knowledge about the workers' ability, if available. Second, the inference function can be designed by the task master based on the knowledge of the labeling tasks. If two labels are similar to each other, then the contributions to them should be similar, for all the workers, when one of these labels is predicted. Other properties that the task master can consider while designing the inference function are the difficulty level of the tasks and the prior distribution on the labels. Clearly, while all of the above properties can be used to design an appropriate inference function, it is not mandatory to use any these properties. For example, a popular inference rule that does not account for these properties is majority voting (MV), while weighted majority voting takes into account the efficiency of the workers.

The following theorem provides the value contribution of each worker and the relation between the accumulated value contribution and the classification error for each task.

Theorem 13. *Given a task i, for the inference rule in (3.11) and (3.12), the value contribution v_k for the k^{th} worker is*

$$v_k(l_i^*) = \min_{l \neq l_i^*} \mathbb{E}_{l_i^*}\Big[f_k(l_i^*, Y) - f_k(l, Y)\Big]. \tag{3.13}$$

Additionally, the classification error $\epsilon_i = \mathbb{P}(\hat{l}_i \neq l_i^*)$ and the accumulated value contribution after M evaluations of a task are related as

$$\sum_{n=1}^{N}\sum_{k=1}^{K} v_k(l_i^*).y_{k,n} \geq \sqrt{MQ^2 \log \frac{L-1}{\epsilon_i}}, \tag{3.14}$$

where $Q = \max_{k\in[K]} \max_{l^*\in[L]} \max_{\hat{l}\in[L]} f_k(l^*, \hat{l})$.

In LCS, the value contributions of the workers are unknown and need to be estimated online. The workers' responses are modeled by a stochastic model where a worker k can be assigned a task multiple times and the correct label is predicted each time with probability p_k independent of the past responses of the worker about the task. Therefore, using (3.13), the estimation of the value contribution in LCS is based on the knowledge of true label of task i l_i^* and the estimate of p_k of the worker k. In practise, the true label l_i^* for a task i is unknown. To circumvent this issue in practical crowdsourcing systems, the ground truth l_i^* is estimated by $\hat{l}_i$ after m^{th} evaluation (3.12). Following the estimate of l_i^*, we estimate p_k for each worker based on its empirical mean. Let m^{th} evaluation of a task i is assigned to a worker k. The worker k is said to have labeled the task correctly if the predicted label at the m^{th} evaluation $\hat{l}_i^{(m)}$ is the same as $\hat{l}_i$, which is an estimate of l_i^* after m evaluations. Since the probability of predicting the correct label is independent of the true label, the empirical estimate of p_k is then updated as the ratio of correctly predicted labels to the total number of evaluations performed by the worker, namely

$$\hat{p}_k = \frac{\hat{p}_k \sum_{n=1}^{m-1} y_{k,n} + \mathbb{1}_{\{\hat{l}_i^{(m)} = \hat{l}_i\}} y_{k,m}}{\sum_{n=1}^{m} y_{k,n}}. \tag{3.15}$$

Using the estimate of p_k, the value contribution v_k is estimated according to (3.13), where the expectation is computed using the empirical estimate of p_k. Under the assumption that the value contribution is independent of the true label i.e. for all $l_i^* \in [L]$ $v_k(l_i^*) = v_k$, the current estimate of the value contribution can be used for the workers' selection in the next iteration.

Now, we briefly re-visit the reformulation of LCS problem in (3.2) to BKP in (3.3). The reformulation of LCS problem to BKP is dependent on the inference rule. The average classification error ϵ (3.1) is the average of ϵ_i. Using Theorem 13, for a generalized inference rule in (3.12), the upper bound on the classification error ϵ_i decays exponentially with the increase in aggregated value contributions from the workers for a task i. Thus, minimizing the ϵ_i can be reformulated as maximizing the aggregated value contributions from the workers for task i. Hence, BKP in (3.3) follows from LCS problem in (3.2). The key benefit of the reformulation to BKP is that it provides an insight on the optimal aggregation of two different attributes, cost and value contribution, of the worker, and facilitate their comparison on a single scale defined as efficiency in Lemma 6. A similar transformation of the problem for labeling tasks, with different constraints, has been considered earlier for special cases such as weighted majority voting and majority voting [83, 1]. However, we formalize the notion of the value contribution for a generalized form of inference rule which recovers the transformation derived for weighted majority voting and majority voting in the literature as a special case.

3.6 Performance Evaluation

We now compare the performance of B-KUBE in conjunction with our value contribution estimation strategy, with three benchmark MAB algorithms for workers' selection using the same value contribution estimation strategy in LCS setup. The benchmark algorithms are Bounded ϵ-First (Bounded ϵ-F), Trailsourcing, and Budget-Limited ϵ-First (ϵ-F). Bounded ϵ-F and ϵ-F are described in [213], whereas, trailsourcing is a special case of Bounded ϵ-F. Bounded ϵ-F consists of separate exploration and exploitation phases. It allocates an ϵ fraction of the total budget for

exploration to estimate the value contributions of the workers. The exploitation phase in Bounded ϵ-F is a single step assignment phase where the labeling tasks are assigned to the workers based on their estimated value contributions. Trailsourcing is a simpler version of Bounded ϵ-F with only one round of exploration phase i.e. each worker is selected exactly once in the exploration phase. Budget-Limited ϵ-First has the same exploration phase as Bounded ϵ-F but in the exploitation phase it assigns all the tasks to a single worker with maximum estimated efficiency.

Like Bounded ϵ-F, the task assignment schemes studied in the literature of traditional CS are based on learning the quality parameters of the workers in the first stage followed by a single step assignment of the tasks to the workers [102, 84, 83, 1, 150]. These schemes are sub-optimal with respect to Bounded ϵ-F as they do not consider the unequal incentives for the workers. Additionally, in [213], the authors also argue that the theoretical regret bounds of Bounded ϵ-F cannot be improved for any estimation scheme for quality parameters of the workers. Thus, we limit ourselves to the above mentioned three schemes for the performance comparison. We compare BKUBE directly with Bounded ϵ-F, and show that BKUBE outperforms it both experimentally and theoretically.

LCS is a novel system proposed in this work, therefore, an appropriate real data set is not available for labeling tasks in this setup. Thus,the algorithms are compared in an experimental setup. Additionally, the evaluations in a simulated setup are common for CS as the other schemes proposed in the literature are mostly evaluated in a simulated environment[102, 1, 103, 109]. We perform the comparison in a setup where twenty workers express their interest to perform binary labeling tasks i.e. $K = 20$ and $L = 2$. In this setup, the labels are considered to be equally likely and the tasks are assumed to be equally difficult. The experiments are performed for two different set of workers. In set A, every worker predicts the true label with probability $p_k > 1/2$. The set B contains 15 workers from set A and 5 spammers, i.e, $p_k = 0.5$. MV is used as the inference rule for labeling the tasks. Since MV does not account for any prior information about the labels and the workers, it provides a neutral environment to capture the performance of the algorithms for workers' selection in LCS. By Theorem 13, the value contribution v_k of a worker

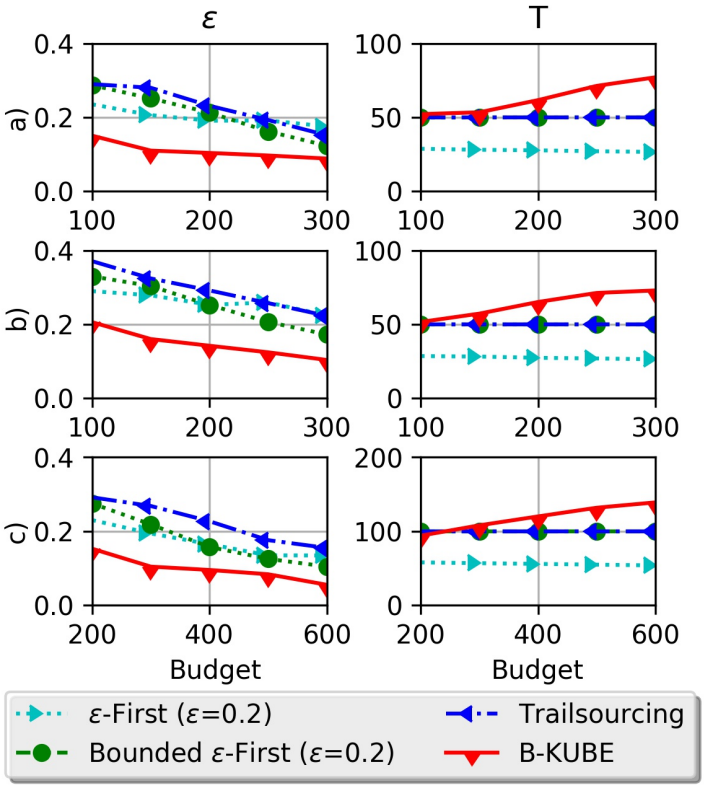


Figure 3.1. The first and second column of plots are corresponding to the classification error ϵ and number of tasks T performed by the workers respectively. a) T=50 and Set A workers b) T=50 and Set B workers c) T=100 and Set A workers

k is $v_k = 2p_k - 1$. In this setup, p_k is randomly chosen from the uniform distribution over the interval $[0.5, 1]$. The value contribution v_k can be computed from p_k. Given v_k, c_k is randomly chosen from the uniform distribution over the interval $[v_k, 1 + v_k]$ as a worker with higher value contribution will expect more incentives.

Assignment of the labeling tasks to the workers is a single step process in all the three benchmark algorithms. Therefore, we evaluate the performance of these algorithms for two different set of tasks with number of tasks $T = 50$ and $T = 100$ in each set and the limit M_k on the number of tasks a workers can perform is $0.6T$ for all the workers. Unlike the benchmark algorithms, B-KUBE evaluates one task at a time and moves to a different task whenever the algorithm is confident that the estimated label of the current task is correct. For the evaluations

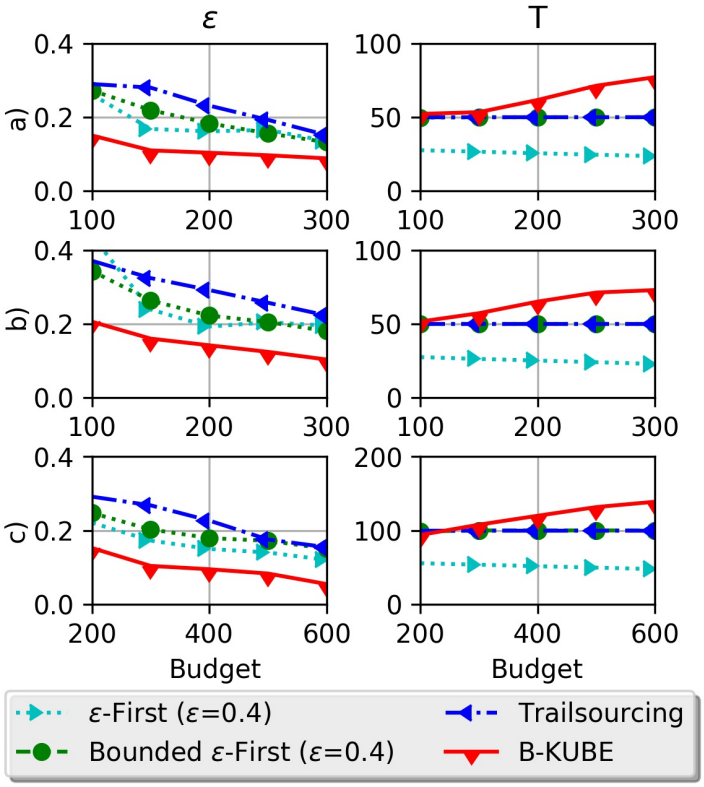


Figure 3.2. The first and second column of plots are corresponding to the classification error ϵ and number of tasks T performed by the workers respectively. a) T=50 and Set A workers b) T=50 and Set B workers c) T=100 and Set A workers

of B-KUBE, we use the criteria proposed in [1] to move on to the next task.

For a given budget, the two key performance measures of the algorithms are: classification error and the number of tasks being performed in LCS. The classification error can be reduced by assigning a task to a large number of workers and aggregating the contributions from the workers to predict the final label of the task. However, this will reduce the number of tasks that can be performed in a limited budget. Thus, there is a trade-off between these two performance measures. The evaluations show that B-KUBE outperforms all the three benchmarks for both the performance measures simultaneously, see Figure 3.1 and 3.2.

As the budget B increases, the classification error decreases for all the algorithms. This is expected, since a larger number of evaluations of the labeling tasks can be performed if more

budget is available. The key observation is that B-KUBE has the smallest classification error whereas the three benchmark algorithms have a higher classification error even after utilizing the available budget to perform less number of tasks in comparison to B-KUBE. Additionally, the classification error of ϵ-F is close to that of Bounded ϵ-F, however, the number of tasks performed by ϵ-F are less than the number of tasks performed by Bounded ϵ-F. This is because the tasks are only assigned to the most efficient worker estimated during the exploration phase. As a consequence, this limits the number of tasks T performed by ϵ-F (Fig. 3.1 and 3.2). Another important observation is that the gap between the classification error of the three benchmark algorithms and B-KUBE reduces as the budget increases. This is because the optimal solution of the BKP includes more and more less efficient workers as the budget increases and the absolute gains from the correct identification of the optimal workers decreases for a large budget. In other words, the losses due to selection of a worker from the sub-optimal set, according to BKP, reduces for large budget.

Figure 3.1(b) and 3.2(b) shows the performance of the algorithms in presence of the spammers for the same setting as in Fig 3.1(a) and 3.2(a) respectively. An important remark here is that the optimal solution for BKP doesn't include any spammer for the values of B considered in the setup. B-KUBE performs better than the three benchmark algorithms in the presence of spammers as well. However, there is a significant increase in the classification error of the B-KUBE for small budget i.e. $B = 100$. The key reason is the absence of a pure exploration phase in B-KUBE which limits the opportunity to identify the spammers. For large budget $B = 300$, the classification error of B-KUBE does not increase significantly as the algorithm is able to utilize the budget efficiently for the identification of spammers. On contrary, this is not true for the three benchmark algorithms.

In conclusion, B-KUBE has a smaller classification error and performs a larger number of tasks in comparison to the three benchmark algorithms. Note that B-KUBE and Bounded ϵ-F are the DGA based extension of KUBE and ϵ-First policies from a budget-limited MAB setup to an arm-limited, budget-limited MAB setup . Finally, the performance trends of B-KUBE

and Bounded ϵ-F in the current setup are similar to the ones of KUBE and ϵ-First policy in a budget-limited MAB setup reported in [211].

3.7 Conclusion

We proposed a notion of Limited-information Crowdsourcing Systems. Unlike traditional CS, LCS monitors the labeling of every single task by a worker in real time, and controls the further assignment of the tasks to the workers based on the estimated value contribution. Due to this form of continuous monitoring, the task master can choose not to assign a task to a worker, and return later to the same worker after exploring other workers, thus, eliminating the requirement of gold-set of tasks. The key challenges in this new setup are determining an efficient workers' selection policy and estimating the value contributions of the workers in real time.

We used B-KUBE to resolve the first challenge and provided its performance analysis, showing that it is an order optimal policy for workers' selection in a budget limited arm limited MAB setup. This work closes the gap in the literature of current MAB setup, showing that B-KUBE is order optimal. To resolve the second challenge, we first introduced the value contributions of the workers for any inference rule and then provided the explicit relation between the accumulated value contribution from the workers and the classification error. We also proposed a strategy to estimate the value contributions of the workers.

We compared the performance of B-KUBE in conjunction with our value contribution estimation strategy, with three benchmark MAB algorithms using the same value contribution estimation strategy in LCS setup. Our experimental evaluations show that B-KUBE outperforms all the three benchmark algorithms for both the performance measures simultaneously. However, it is worth noticing that B-KUBE has a higher computational complexity than the benchmarks evaluated here.

The MAB setup considered in this paper is important as it has extension to various applica-

tions like recommendation systems and learning optimal causal intervention. In recommendation systems, the selection of items is analogous to the workers' selection and value of the items need to estimated online from the user's prospective like value contributions in LCS. Likewise, the current MAB setup can be used to learn an optimal causal intervention in Directed Acyclic Graphs. In this application, the intervention selection is analogous to workers' selection and the reward corresponding to the intervention is analogous to workers' value contribution. The budget constraint is applicable to these applications in a similar way as to the current LCS setup. Hence, there exists many applications where the current MAB setup can be used along with an online estimation scheme, depending on the application, to design an efficient multi-agent system. Likewise, the work can be applied to various Multi-agent systems as the budget limited arm limited MAB setup is a popular model for constraining the systems.

Additionally, the work introduces a notion of LCS which triggers another research direction for crowdsourcing systems. The value contributions of the workers can be formulated for more complicated tasks, for example translation and testing, that require variety of skills to complete. If a task requires z skills to be completed then the value contribution of a worker can be modeled as a z dimensional vector where each dimension of the vector corresponds to a particular skill required for completing the task. Designing the workers' selection policy and online strategy to estimate the value contributions of the workers for such tasks is challenging and is left as future work.

3.8 Acknowledgment

Chapter 3, in full, is a reprint of the material as it appears in Anshuka Rangi and Massimo Franceschetti, "Multi-armed bandit algorithms for crowdsourcing systems with online estimation of workers' ability", *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, July 2018. The dissertation author was the primary investigator and author of this paper.

3.9 Appendix

3.9.1 Proof of Theorem 11

Let $i(n)$ be the worker selected at the n^{th} iteration of B-KUBE; B_n is the residual budget before the n^{th} iteration of B-KUBE; $m_{(k,n)}$ is the remaining number of tasks a worker k can perform at the n^{th} iteration of B-KUBE; $\hat{x}(n)$ is an estimate of $\lfloor x^*_{B_n} \rfloor$ by DGA using the estimated efficiencies $\hat{e}_k = \hat{w}_k/c_k$, where $x^*_{B_n}$ is the solution proposed by Lemma 6 for a given budget B_n and the set $\{m_{(k,n)}\}$ of the remaining tasks that can be performed by the workers; $j \notin \hat{x}(n)$ implies that an additional selection of worker j is not proposed by selection strategy $\hat{x}(n)$; $\hat{s}(n)$ is the estimated splitting worker by DGA at the n^{th} iteration; $N_k(N)$ denotes the number of times the worker k is selected more than the number of selections proposed by $\lfloor x^* \rfloor$ when B-KUBE stops after N iterations.

The following two lemmas are the key components of the proof of Theorem 11.

Lemma 7. *Let B-KUBE perform N iterations. For all $1 \leq n \leq N$, if a worker j is selected, then*

$$\mathbb{P}\big(i(n) = j|N\big) \leq \mathbb{P}\big(j \in \hat{x}(n)|N\big) + \Big(\frac{C_{\max}}{C_{\min}}\Big)^2 \frac{1}{N-n+1}. \tag{3.16}$$

Proof. We consider the n^{th} iteration and assume that the estimated efficiencies of the workers $\hat{e}_k = \hat{w}_k/c_k$ are such that $\hat{e}_1 \geq \hat{e}_2 \geq \ldots \geq \hat{e}_K$. For convenience, we drop the conditioning on N in the notation. Let $M^*(B_n, \{m_{(k,n)}\}) = \{m^*_{k,n}\}$, where $m^*_{k,n}$ is the number of selections of worker k proposed by DGA at the n^{th} iteration. For a given B_n and $\{m_{(k,n)}\}$, using Bayes rule, and the fact that $i(n)$ is independent of B_n and $\{m_{(k,n)}\}$, given $M^*(B_n, \{m_{(k,n)}\})$, we have

$$\begin{aligned}\mathbb{P}\big(i(n) = j|B_n, \{m_{(k,n)}\}\big) = \sum_{\{m^*_{(k,n)}\}} & \mathbb{P}\big(i(n) = j|M^*(B_n, \{m_{(k,n)}\})\big) \\ & \cdot \mathbb{P}\big(M^*(B_n, \{m_{(k,n)}\})\big)|B_n, \{m_{(k,n)}\}).\end{aligned} \tag{3.17}$$

DGA proposes the selection of the first $\hat{s}(n)-1$ workers up to their maximum remaining capacity $m_{(k,n)}$ and selects the worker $\hat{s}(n)$ as many times as feasible. These selections are same as the ones suggested by $\hat{x}(n)$. Since the selection strategies $\hat{x}(n)$ and $M^*(B_n, \{m_{(k,n)}\})$ are same for the first $\hat{s}(n)$ workers, the remaining budget after the selections of the first $\hat{s}(n)$ workers is at most $c_{\hat{s}(n)}$, otherwise, the worker $\hat{s}(n)$ can be selected one more time. Thus, the number of workers' selections suggested by $M^*(B_n, \{m_{(k,n)}\})$ in addition to $\hat{x}(n)$ can be bounded as:

$$\sum_{i \notin \hat{x}(n)} m^*_{i,n} \leq \frac{c_{\hat{s}(n)}}{C_{\text{min}}}. \tag{3.18}$$

Additionally, the total number of selections as proposed by DGA can be bounded as:

$$\sum_{k=1}^{K} m^*_{k,n} \geq \frac{B_n}{C_{\text{max}}}. \tag{3.19}$$

The following inequalities can be obtained by combining eq. (3.18) and (3.19), and using the fact that $c_{\hat{s}(n)} \leq C_{\text{max}}$,

$$\frac{\sum_{i \notin \hat{x}(n)} m^*_{i,n}}{\sum_{k=1}^{K} m^*_{k,n}} \leq \frac{c_{\hat{s}(n)}}{C_{\text{min}}} . \frac{C_{\text{max}}}{B_n} \leq \left(\frac{C_{\text{max}}}{C_{\text{min}}}\right)^2 \frac{C_{\text{min}}}{B_n}. \tag{3.20}$$

Additionally, before each iteration n of B-KUBE, the remaining number of iterations are $N-n+1$. Therefore, the residual budget B_n is at least $C_{\text{min}}(N-n+1)$. Thus, we have

$$\frac{C_{\text{min}}}{B_n} \leq \frac{1}{N-n+1}. \tag{3.21}$$

Now, the probability on the right-hand side of eq. (3.17) can be written as

$$
\begin{aligned}
&\mathbb{P}\Big(i(n)=j|M^*(B_n,\{m_{(k,n)}\})=\{m^*_{(k,n)}\}\Big)\\
&\overset{(a)}{=}\mathbb{P}\Big(i(n)=j,j\in\hat{x}(n)|M^*(B_n,\{m_{(k,n)}\})=\{m^*_{k,n}\}\Big)\\
&\qquad+\mathbb{P}\Big(i(n)=j,j\notin\hat{x}(n)|M^*(B_n,\{m_{(k,n)}\})=\{m^*_{(k,n)}\}\Big)\\
&\overset{(b)}{\leq}\mathbb{P}\Big(j\in\hat{x}(n)|M^*(B_n,\{m_{(k,n)}\})\Big)\\
&\qquad\cdot\mathbb{P}\Big(i(n)=j|j\in\hat{x}(n),M^*(B_n,\{m_{(k,n)}\})\Big)\\
&\qquad+\frac{\sum_{i\notin\hat{x}(n)}m^*_{(i,n)}}{\sum_{k=1}^{K}m^*_{(k,n)}}\mathbb{P}\Big(j\notin\hat{x}(n)|M^*(B_n,\{m_{(k,n)}\})\Big)\\
&\overset{(c)}{\leq}\mathbb{P}\Big(j\in\hat{x}(n)|M^*(B_n,\{m_{(k,n)}\})\Big)+\frac{\sum_{i\notin\hat{x}(n)}m^*_{(i,n)}}{\sum_{k=1}^{K}m^*_{(k,n)}}\\
&\overset{(d)}{\leq}\mathbb{P}\Big(j\in\hat{x}(n)|M^*(B_n,\{m_{(k,n)}\})\Big)+\Big(\frac{C_{\max}}{C_{\min}}\Big)^2\frac{1}{N-n+1},
\end{aligned}
\tag{3.22}
$$

where (a) follows from the fact that two events are mutually exclusive; (b) follows because B-KUBE chooses worker j with probability $m^*_{j,n}/\sum_{k=1}^{K}m_{(k,n)}$ and $j\notin\hat{x}(n)$; (c) follows because the probability is bounded by 1; (d) follows by combining eq. (3.20) and (3.21). The lemma follows by combining eq. (3.17) and (3.22), and using Bayes rule. □

Lemma 8. *Let $\lfloor x^*\rfloor$ be the optimal workers' selection strategy. If $j\notin\lfloor x^*\rfloor$ and $j\in\hat{x}(n)$, then there is at least one worker $k^{'}\in\lfloor x^*\rfloor$ whose estimated efficiency is less than the estimated efficiency of the worker j i.e. $\hat{e}_{k'}\leq\hat{e}_j$ and the worker $k^{'}$ can perform additional tasks.*

Proof. If $\lfloor x^*\rfloor$ is the optimal workers' selection strategy at budget B, then for any budget $B^{'}<B$ the optimal selection strategy $\lfloor x^*_{B'}\rfloor$ is a subset of the selections proposed by $\lfloor x^*\rfloor$. This can be seen from Lemma 6.

We can say that if a worker $j\notin\lfloor x^*\rfloor$, then $j\notin\lfloor x^*_{B_n}\rfloor$ for the residual budget B_n as $B_n\leq B$. Now $\hat{x}(n)$ is an estimate of $\lfloor x^*_{B_n}\rfloor$ by DGA, and $j\in\hat{x}(n)$ according to the hypothesis in this lemma. Thus, there is at least one worker $k^{'}\in\lfloor x^*_{B_n}\rfloor$ whose estimated efficiency is less than the estimated efficiency of worker j by DGA i.e. $\hat{e}_{k'}\leq\hat{e}_j$. Also, $k^{'}\in\lfloor x^*_{B_n}\rfloor$ implies that the worker $k^{'}$ can still perform tasks. As $\lfloor x^*_{B_n}\rfloor$ is a subset of $\lfloor x^*\rfloor$ and $k^{'}\in\lfloor x^*_{B_n}\rfloor$, therefore

$k^{'} \in \lfloor x^{*} \rfloor$ and the worker $k^{'}$ can perform more tasks. $\square$

Using the above lemmas, we continue with the proof of Theorem 11. Without loss of generality, let us assume that the efficiencies e_k of the workers are such that $e_1 \geq e_2 \geq \ldots \geq e_K$. The notation of conditioning N is dropped for convenience.

If $\lfloor x^{*} \rfloor$ is the optimal selection strategy, then by Lemma 6 the selection of the first $s-1$ workers is always optimal. Thus, if a worker $j \notin \lfloor x^{*} \rfloor$,then the worker $j \geq s$. According to $\lfloor x^{*} \rfloor$, the selection of these workers is always sub-optimal with the exception of the s^{th} worker. Therefore, $j = s$ will be handled separately in the proof. Thus, $N_j(N)$ for $j \notin \lfloor x^{*} \rfloor$ can be written as

$$\begin{aligned} N_j(N) &\leq 1 + \min\Big(\sum_{n=K+1}^{N} \{i(n) = j\}, M_j \Big) \\ &\leq 1 + \sum_{n=K+1}^{N} \{i(n) = j\}. \end{aligned} \tag{3.23}$$

Taking expectations on both sides, for $1 \leq l \leq M_j$, we have

$$\begin{aligned} \mathbb{E}\big[N_j(N)\big] &\leq 1 + \sum_{n=K+1}^{N} \mathbb{P}\Big(i(n) = j\Big) \\ &\overset{(a)}{\leq} 1 + \sum_{n=K+1}^{N} \mathbb{P}\Big(j \in \hat{x}(n)\Big) + \sum_{n=K+1}^{N} \Big(\frac{C_{max}}{C_{min}}\Big)^2 \frac{1}{N-n+1} \\ &\overset{(b)}{\leq} l + \sum_{n=K+1}^{N} \mathbb{P}\Big(j \in \hat{x}(n), M_j(n) \geq l\Big) \\ &\quad + \sum_{n=K+1}^{N} \Big(\frac{C_{max}}{C_{min}}\Big)^2 \frac{1}{N-n+1}, \end{aligned} \tag{3.24}$$

where (a) follows from Lemma 7; (b) follows from the intersection of events $j \in \hat{x}(n)$ and $M_j(n) \geq l$ where $M_j(n)$ is the number of times worker j is selected before n^{th} iteration i.e. $M_j(n) = M_j - m_{(j,n)}$.

Let $b_{N,m_{(j,n)}} = \sqrt{2 \log N / M_j - m_{(j,n)}}$ and $b_{N,m_j} = \sqrt{2 \log N / M_j - m_j}$. Now, con-

sider the event $A(n,j) = \{j \in \hat{x}(n), M_j(n) \geq l\}$ on the right-hand side of (3.24). By hypothesis, we have $j \notin \lfloor x^* \rfloor$, however, $j \in \hat{x}(n)$, thus by Lemma 8 $\exists k' \in \lfloor x^* \rfloor$ such that $\hat{e}_{k'} \leq \hat{e}_j$. Note that Lemma 8 also accounts for the sub-optimal selections of the splitting worker s. It follows that the probability of the event $A(n,j) = \{j \in \hat{x}(n), N_j(n) \geq l\}$ can be simplified as:

$$\begin{aligned}
\sum_{n=K+1}^{N} \mathbb{P}\Big(A(n,j)\Big) &\leq \sum_{n=K+1}^{N} \mathbb{P}\Big(\hat{e}_j \geq \hat{e}_{k'}; M_j(n) \geq l\Big) \\
&= \sum_{n=K+1}^{N} \mathbb{P}\Big(\frac{\hat{v}_{j,m_{(j,n)}}}{c_j} + \frac{b_{n,m_{(j,n)}}}{c_j} \\
&\qquad \geq \frac{\hat{v}_{k',m_{(k',n)}}}{c_{k'}} + \frac{b_{n,m_{(k',n)}}}{c_{k'}}; M_j(n) \geq l\Big) \\
&\leq \sum_{n=K+1}^{N} \mathbb{P}\Big(\max_{l \leq m_j \leq \min(n, M_j)} \frac{\hat{v}_{j,m_j}}{c_j} + \frac{b_{n,m_j}}{c_j} \\
&\qquad \geq \min_{1 \leq i \leq n} \Big\{\frac{\hat{v}_{k',m_{(k',i)}}}{c_{k'}} + \frac{b_{n,m_{(k',i)}}}{c_{k'}}\Big\}\Big) \\
&\leq \sum_{n=1}^{N} \sum_{i=1}^{n} \sum_{m_j=1}^{n} \mathbb{P}\big(F\big), \qquad (3.25)
\end{aligned}$$

where the event F is defined as follows:

$$\frac{\hat{v}_{j,m_j}}{c_j} + \frac{b_{n,m_j}}{c_j} \geq \frac{\hat{v}_{k',m_{(k',i)}}}{c_{k'}} + \frac{b_{n,m_{(k',i)}}}{c_{k'}}. \qquad (3.26)$$

The event F occurs only if at least one of the events among C, D and E occurs where

$$\text{C: } \frac{\hat{v}_{k',m_{(k',i)}}}{c_{k'}} + \frac{b_{n,m_{(k',i)}}}{c_{k'}} \leq \frac{v_{k'}}{c_{k'}}, \qquad (3.27)$$

$$\text{D: } \qquad \frac{v_j}{c_j} \leq \frac{\hat{v}_{j,m_j}}{c_j} + \frac{b_{n,m_j}}{c_j}, \qquad (3.28)$$

$$\text{E:} \qquad \frac{v_{k'}}{c_{k'}} \leq \frac{v_j}{c_j} + 2\frac{b_{n,m_j}}{c_j}. \tag{3.29}$$

This claim can be proved by contradiction. Thus, the probability of the event F can be bounded as

$$\mathbb{P}\big(F\big) \leq \mathbb{P}\big(C\big) + \mathbb{P}\big(D\big) + \mathbb{P}\big(E\big). \tag{3.30}$$

Using Chernoff-Hoeffding inequalities, the probability of the events C and D can be bounded as

$$\mathbb{P}(C) \leq exp\big(-2b^2_{(n,m_{(k',i)})}(M_{k'} - m_{(k',i)})\big) = n^{-4}, \tag{3.31}$$

$$\mathbb{P}(D) \leq exp\big(-2b^2_{(n,m_j)}(M_j - m_j)\big) = n^{-4}. \tag{3.32}$$

Next, we show that for $l \geq 8\log N/\min\{Q^2_{\text{min}}, d_s^2\}$, we have $\mathbb{P}(E) = 0$. The analysis is split into two cases: $j > s$ and $j = s$.

Case 1: For $j > s$ and $l \geq 8\log N/Q^2_{min}$, we have

$$\begin{aligned}
\frac{v_{k'}}{c_{k'}} - \frac{v_j}{c_j} - \frac{2b_{n,m_j}}{c_j} &\overset{(a)}{\geq} \frac{v_{k'}}{c_{k'}} - \frac{v_j}{c_j} - 2b_{n,m_j} \\
&\overset{(b)}{\geq} \frac{v_{k'}}{c_{k'}} - \frac{v_j}{c_j} - 2\sqrt{\frac{2\log n}{l}} \\
&\overset{(c)}{\geq} \frac{v_{k'}}{c_{k'}} - \frac{v_j}{c_j} - 2\sqrt{\frac{2Q^2_{\text{min}}\log n}{8\log N}} \\
&\overset{(d)}{\geq} \frac{v_{k'}}{c_{k'}} - \frac{v_j}{c_j} - Q_{\text{min}} \\
&\overset{(e)}{\geq} \frac{v_{k'}}{c_{k'}} - \frac{v_j}{c_j} - Q_j \\
&\overset{(f)}{\geq} 0,
\end{aligned} \tag{3.33}$$

where (a) follows from the fact that $\forall j \in [K]$, $c_j \geq 1$; (b) and (c) use the fact that $m_j \geq l \geq 8 \log N/Q_{\text{min}}^2$; (d) follows from the fact that $n \leq N$; (e) and (f) follow from the fact that $Q_{\text{min}} = \min_{j\in[K]} Q_j$ where $Q_j = |v_s/c_s - v_j/c_j|$ and s is the splitting worker.

Case 2: For $j = s$, we have by Lemma 8 that $k^{'} < s$. By following the same steps as in case 1, it can be shown that $P(E) = 0$ for $l \geq 8 \log N/d_s^2$ where $d_s = |v_{s-1}/c_{s-1} - v_s/c_s|$.

Thus, for $l \geq 8 \log N/ \min\{Q_{\text{min}}^2, d_s^2\}$, we have $P(E) = 0$. Now combining this fact with eq. (3.25), (3.31) and (3.32), we have

$$\sum_{n=K+1}^{N} \mathbb{P}\Big(j \in \hat{x}(n), N_j(n) \geq l\Big) \leq \sum_{n=1}^{N}\sum_{i=1}^{n}\sum_{m_j=1}^{n} 2n^{-4} \leq \frac{\pi^2}{3}. \tag{3.34}$$

The third term in eq. (3.24) can be bounded as

$$\sum_{n=K+1}^{N} \left(\frac{C_{\text{max}}}{C_{\text{min}}}\right)^2 \frac{1}{N-n+1} \leq \left(\frac{C_{\text{max}}}{C_{\text{min}}}\right)^2 \log(N). \tag{3.35}$$

Thus, for $l = 8 \log N/\min\{Q_{\text{min}}^2, d_s^2\} + 1$, and combining eq. (3.24), (3.34) and (3.35) the result follows

$$\mathbb{E}\big[N_j(N)|N\big] \leq \frac{8 \log N}{\min\{Q_{\text{min}}^2, d_s^2\}} + \left(\frac{C_{\text{max}}}{C_{\text{min}}}\right)^2 \log(N) + \frac{\pi^2}{3} + 1. \tag{3.36}$$

3.9.2 Proof of Theorem 12

Proof. The key idea is to use the result from Theorem 11 to bound the regret of B-KUBE assuming $\lfloor x^* \rfloor$ is the optimal workers' selection strategy. This bound can then be combined with eq. (3.8) to obtain the regret bound for B-KUBE.

Let $\Delta_j = v_{\text{max}} - v_j$ where $v_{\text{max}} = \max_{j\in[K]} v_j$, I^* is the set of top $s-1$ workers when arranged in decreasing order of their efficiencies and s is the splitting worker. The regret of

B-KUBE is

$$
\begin{aligned}
R^{B-KUBE}(B) &= v^{*}_{BKP} - v^{B-KUBE}(B) \\
&= v^{*}_{BKP} - v^{'} + v^{'} - v^{B-KUBE}(B) \\
&\overset{(a)}{\leq} v_s + \mathbb{E}_N\Bigg[\sum_{j\notin I^*}\Delta_j \mathbb{E}\big[N_j(N)|N\big]\Bigg], \\
&\overset{(b)}{\leq} \mathbb{E}_N\Bigg[\sum_{j\notin I^*}\Delta_j\Bigg(\Big(\frac{8}{\min\{Q^2_{\min}, d^2_s\}} + \Big(\frac{C_{\max}}{C_{\min}}\Big)^2\Big) \\
&\qquad\qquad \cdot \log\left(\frac{B}{C_{\min}}\right) + \frac{\pi^2}{3} + 1\Bigg)\Bigg] + 1 \\
&= \sum_{j\notin I^*}\Delta_j\Bigg(\Big(\frac{8}{\min\{Q^2_{\min}, d^2_s\}} + \Big(\frac{C_{\max}}{C_{\min}}\Big)^2\Big) \\
&\qquad\qquad \cdot \log\left(\frac{B}{C_{\min}}\right) + \frac{\pi^2}{3} + 1\Bigg) + 1,
\end{aligned} \tag{3.37}
$$

where (a) follows from eq. (3.8) and Theorem 11; (b) follows from $N \leq B/C_{\min}$ and $v_s \leq 1$. Hence, the regret bound of B-KUBE follows. □

3.9.3 Proof of Theorem 13

Proof. The classification error can be written as follows

$$
\begin{aligned}
\mathbb{P}(\hat{l} \neq l^*) &= \mathbb{P}\big(\cup_{l\neq l^*}(s_l > s_{l^*})\big) \\
&\overset{(a)}{\leq} \sum_{l\neq l^*}\mathbb{P}\Big(s_l - s_{l^*} - \mathbb{E}(s_l - s_{l^*}) \geq \mathbb{E}(s_{l^*} - s_l)\Big) \\
&\overset{(b)}{\leq} \sum_{l\neq l^*}\exp\left(\frac{-\big(\mathbb{E}(s_{l^*} - s_l)\big)^2}{2Q^2M}\right) \\
&\overset{(c)}{\leq} (L-1)\exp\Big(\frac{-\min_{l\neq l^*}\big(\mathbb{E}(s_{l^*} - s_l)\big)^2}{2Q^2M}\Big) \\
&\overset{(d)}{=} (L-1)\exp\Big(\frac{-\big(\min_{l\neq l^*}\mathbb{E}(s_{l^*} - s_l)\big)^2}{2Q^2M}\Big) \\
&\overset{(e)}{\leq} (L-1)\exp\Big(\frac{-(\sum_{k=1}^{K} v_k(l^*)y_{k,n})^2}{2Q^2M}\Big),
\end{aligned} \tag{3.38}
$$

where (a) follows from the union bound; (b) follows from the Azuma-Hoeffding inequality because $\mathbb{E}(s_{l^*} - s_l) > 0$ as inference functions are designed to favor the true hypothesis; (c) follows trivially from the fact that $\forall l \neq l^*$, we have $\min_{l \neq l^*} \left(\mathbb{E}(s_{l^*} - s_l)\right) < \mathbb{E}(s_{l^*} - s_l)$; (d) follows from $\mathbb{E}(s_{l^*} - s_l) > 0$; (e) follows from the definition of $v_k(l^*)$.

Now, if the classification error is at most ϵ, then

$$\sum_{n=1}^{N} \sum_{k=1}^{K} v_k(l^*) y_{k,n} \geq \sqrt{2MQ^2 \log \frac{L-1}{\epsilon}}. \tag{3.39}$$

□

Chapter 4

Unifying the Stochastic and the Adversarial Knapsack Bandits

4.1 Introduction

Multi-Armed Bandit (MAB) is a sequential decision making problem under uncertainty, that is based on balancing the trade-off between exploration and exploitation, i.e. "the conflict between taking actions which yield immediate rewards and taking actions whose benefits will be seen later." A common feature in various applications of MAB is that the resources consumed during the decision making process are limited. For instance, scientists experimenting with alternative medical treatments may be limited by the number of patients participating in the study as well as by the cost of the material used in the treatments. Similarly, in web advertisements, a website experimenting with displaying advertisements is constrained by the number of users who visit the site as well as by the advertisers' budgets. A retailer engaging in price experimentation faces inventory limits along with a limited number of consumers. A model which incorporates a budget constraint on these supply limits is Bandits with Knapsack (BwK). This can be seen as a game between a player and an adversary (or environment) that evolves for T rounds. The player is constrained by a budget B on the resources consumed during the decision making process. The game terminates when the player runs out of budget, therefore T is dependent on B. At each round t, the player performs an action i from a set of K actions, pays a cost for the selected action i from the budget B and receives a reward in $[0, 1]$ for the selected action i. The reward and the

cost can vary from application to application. For example, in web advertisement, the reward is the click through rate and the cost is the space occupied by the advertisement on the web page. In medical trials, the reward is the success rate of the medicine and the cost corresponds to the cost of the material used.

The Bandits with Knapsack problem can be classified into two categories: stochastic BwK and adversarial BwK. In stochastic BwK, the reward and the cost of each action is an i.i.d sequence over T rounds drawn from a fixed unknown distribution. In adversarial BwK, the sequence of the rewards and the costs associated with each action over T rounds is assigned by an oblivious adversary before the game starts. The objective of the player is to minimize the expected regret, which is the difference between the expectation of the rewards received from the best fixed action in the hindsight and the sum of rewards received by the player's action selection strategy.

The stochastic BwK setting has been extensively studied in the literature [210, 211, 52, 20, 5, 213, 4, 232, 190, 174]. The results in these works can be broadly classified into two categories depending on the regret analysis. The problem dependent bound on the expected regret is $O(\log(B))$ [211, 52, 232, 247, 174], while the problem independent bound on the expected regret is $O(\sqrt{KB})$ [4, 5, 20].

Limited attention has been received by the adversarial BwK setting [247]. In this setting, it has been assumed that the reward at round $t \leq T$ is greater than the cost at round $t \leq T$ for every action over the duration of the game [247]. Under this assumption, EXP3.M.B has been proposed and proven to be order optimal [247]. We observe here that the assumption on the reward being greater than the cost is uncommon in the literature of the BwK problem, and does not have any physical meaning in many applications. For example, in web advertisement, the click through rate (i.e., reward) and the space occupied by the advertisement on the web page (i.e., cost) cannot be compared with each other. Likewise, in a medical trial, the reward is the success rate of the medicine and the cost corresponds to the cost of the material used, and the comparison of these values has no meaning. Thus, a key question is how to design an algorithm

for the adversarial BwK in a general reward setting that achieves order optimal regret guarantees.

Another key challenge is to provide a solution that is satisfactory for both stochastic and adversarial settings. In many real-world situations, there is no information about whether the bandit model is used in a stochastic or adversarial manner. Thus, the deployed algorithm has to be able to perform well in both cases. Current algorithms in the adversarial BwK (e.g., EXP3.M.B), do not provide optimal regret guarantees in the stochastic setting, i.e. $O(\log(B))$, and algorithms in the stochastic BwK (e.g., KUBE), do not provide optimal regret guarantees in the adversarial setting, i.e. $O(\sqrt{KB})$. Currently, there is no work proposing a practical algorithm for both settings. Finally, the literature of the BwK problem typically assumes that the costs are bounded by a constant (i.e., they are independent of the budget B) and it is unknown whether state-of-the-art regret bounds hold for the case of large costs (i.e., when costs are comparable to the budget B).

4.2 Contribution

In this framework, the contribution of our work is three fold. First, we extend EXP3, a classical algorithm, proposed for the adversarial MAB setup [17], and propose EXP3.BwK, an algorithm for the adversarial BwK setup. We remove the assumption on the rewards and the costs previously used in [247] to obtain regret bounds and we show that the expected regret of EXP3.BwK is $O(\sqrt{BK \log K})$. We also show the lower bound $\Omega(\sqrt{KB})$ in the adversarial BwK setting. It follows that EXP3.BwK is order optimal. Second, we unify the stochastic and the adversarial settings by proposing EXP3++.BwK, a novel and practical algorithm which works well in both of these settings. This algorithm incurs an expected regret of $O(\sqrt{BK \log K})$ and $O(\log^2(B))$ in the adversarial and the stochastic BwK settings respectively. Note that the regret bound of EXP3++.BwK for the stochastic setting has an additional factor of $\log(B)$ in comparison to the optimal expected regret i.e. $O(\log(B))$. Thus, EXP3++.BwK exhibits an almost optimal behavior in both the stochastic and the adversarial settings. Table 4.1 summarizes

Table 4.1. Contributions to the literature of BwK.

Algorithm	Upper bound	Lower bound
KUBE for BwK [211]	$O(K\log(B)/\min_{i\in[K]}\Delta(i))$	$\Omega(\log(B))$
B-KUBE for Bounded BwK [174]	$O(K\log(B)/\min_{i\in[K]}\Delta(i))$	$\Omega(\log(B))$
UCB-BV for variable cost [52]	$O(K\log(B)/\min_{i\in[K]}\Delta(i))$	$\Omega(\log(B))$
UCB-MB for multiple plays [247]	$O(K\log(B))$	
EXP3.M.B [247]	$O(\sqrt{K\log(K)B})$	$\Omega((1-1/K)^2\sqrt{KB})$
EXP3.BwK (our contribution)	$O(\sqrt{K\log(K)B})$	$\Omega(\sqrt{KB})$
EXP3++.BwK in Adversarial setting (our contribution)	$O(\sqrt{K\log(K)B})$	
EXP3++.BwK in Stochastic setting (our contribution)	$O(K\log^2(B)/\min_{i\in[K]}\Delta(i))$	

these contributions and compares them with the other results in the literature. In the table, the problem-dependent parameter $\Delta(i)$ represents the difference between the contributions of the optimal action and the action i, and is formally defined in the next section. Finally, we show that if the maximum cost is bounded above by B^α, where $\alpha \in [0,1]$, then the lower bound on the expected regret in the adversarial BwK setup scales at least linearly with the maximum cost, namely it is $\Omega(B^\alpha)$. This implies that when $\alpha > \frac{1}{2}$, it is impossible to achieve a regret bound of $O(\sqrt{B})$, which is order optimal in cases with small costs.

4.2.1 Related Work

In the MAB literature, the problem of finding one algorithm for both the stochastic and the adversarial setting has been referred as "best of both worlds" [34, 18, 195, 194, 139]. SAO, the first algorithm proposed in the literature of this problem, relies on the knowledge of the time horizon T, and performs an irreversible switch to EXP3.P if the beginning of the game is estimated to exhibit an adversarial, or non-stochastic, behavior [34]. The expected regret of SAO in the stochastic MAB setting is $O(\log^3(T))$, and in the adversarial MAB setting is $O(\sqrt{T}\log^2(T))$.

Using ideas from SAO, a new algorithm SAPO was proposed [18]. SAPO exploited some novel criteria for the detection of the adversarial, or non-stochastic, behavior, and performs an irreversible switch to EXP3.P if such a behavior is detected. Thus, both SAO and SAPO initially assume that the rewards are stochastic, and perform an irreversible switch to EXP3.P if this assumption is detected to be incorrect. The expected regret of SAPO is $O(\log^2(T))$ in the stochastic MAB setting, and $O(\sqrt{T\log(T^2)})$ in the adversarial MAB setting. Later, EXP3++ was proposed [195]. Unlike SAO and SAPO, this algorithm starts by assuming the rewards exhibit an adversarial, or non-stochastic, behavior and adapts itself as it encounters stochastic behavior on rewards. The analysis of EXP3++ was improved in [194], showing that the algorithm guarantees an expected regret of $O(\log^2(T))$ and $O(\sqrt{T})$ in the stochastic and the adversarial MAB settings respectively.

The problem of stochastic bandits corrupted with adversarial samples has been studied in the regime of small corruptions [139]. The algorithm proposed in this work utilizes the idea of active arm elimination based on upper and lower confidence bound of the estimated rewards. The work provides the regret analysis of the algorithm as the corruption C is introduced in the rewards, and shows that the decay in performance is order optimal in C.

The "best of both worlds" problem has not been studied before in the BwK setting.

4.3 Problem Formulation

A player can choose from a set of K actions, and has a budget B. At round t, each action $i \in [K]$ is associated with a reward $r_t(i) \in [0,1]$ and a cost $c_t(i) \in [c_{min}, c_{max}]$ with $c_{min} \leq c_{max}$. For now, we assume that $c_{max} = 1$, and will investigate the case of having larger costs in Section 5. At round t, the player performs an action $i_t \in [K]$, pays the cost $c_t(i_t)$ and receives the reward $r_t(i_t)$. The gain of a player's strategy $\mathcal{A}$ is defined as

$$G(\mathcal{A}) = \mathbb{E}\Big[\sum_{t=1}^{\tau(\mathcal{A})} r_t(i_t)\Big], \tag{4.1}$$

where $\tau(\mathcal{A})$ is number of rounds after which the strategy $\mathcal{A}$ terminates. The objective of a player is to design $\mathcal{A}$ such that

$$\begin{aligned} &\max_{\{i_1,i_2,\ldots,i_{\tau(\mathcal{A})}\}} G(\mathcal{A}) \\ &\text{s.t. } \mathbb{P}\Big(\sum_{t=1}^{\tau(\mathcal{A})} c_t(i_t) \leq B\Big) = 1. \end{aligned} \tag{4.2}$$

Note that $\tau(\mathcal{A})$ is dependent on the budget B. Let $\mathcal{A}^*$ be the algorithm that solves (4.2). The expected regret of an algorithm $\mathcal{A}$ is defined as

$$R(\mathcal{A}) = G(\mathcal{A}^*) - G(\mathcal{A}). \tag{4.3}$$

The optimization problem in (4.2) is a knapsack problem, and is known to be NP-hard [107]. Given that the rewards and the costs of all the actions are known and fixed for all T rounds, the greedy algorithm $\mathcal{A}^G$ for solving (4.2) makes an action selection in the decreasing order of the efficiency, defined as $e(i) = r(i)/c(i)$ for an action $i \in [K]$, until the budget constraint in (4.2) is satisfied. It can be shown that [107]

$$G(\mathcal{A}^G) \leq G(\mathcal{A}^*) \leq G(\mathcal{A}^G) + \max_{i\in[K]} e(i). \tag{4.4}$$

In the stochastic setting, for all t and $i \in [K]$, the reward $r_t(i)$ and the cost $c_t(i)$ of an action i are identically and independently distributed according to some unknown distributions. The expected reward and the expected cost of an action i are denoted by $\mu(i)$ and $\rho(i)$ respectively. Thus, in the stochastic setting, the efficiency of an action i can be defined as $e(i) = \mu(i)/\rho(i)$. Using (4.4), the expected regret of an algorithm $\mathcal{A}$ simplifies to

$$\begin{aligned} R(\mathcal{A}) &\leq \max_{i\in[K]} \frac{\mu(i)}{\rho(i)}(\tau(\mathcal{A}^G)+1) - G(\mathcal{A}) \\ &= e(i^*)(\tau(\mathcal{A}^G)+1) - G(\mathcal{A}) \\ &\leq \sum_{i\in[K]/\{i^*\}} \Delta(i)\mathbb{E}[N_T(i)], \end{aligned} \tag{4.5}$$

where $i^* = \text{argmax}_{i\in[K]} e(i)$, $\Delta(i) = e(i^*) - e(i)$, $N_T(i)$ is the number of times an action i is selected in T rounds, and $T = \max\{\tau(\mathcal{A}), \tau(\mathcal{A}^G)\}$. The definition in (4.5) is consistent with the literature of stochastic BwK [52, 213].

In the adversarial setting, for all t, $r_t(i)$ and $c_t(i)$ are chosen by an adversary before the game starts. In this setting, the efficiency of an action i at round t can be defined as $e_t(i) = r_t(i)/c_t(i)$. Therefore, using (4.4), we consider the following expected regret

$$R(\mathcal{A}) = \mathbb{E}\left[z(\mathcal{A})\left(\sum_{t=1}^{T(i^*)} e_t(i^*) - \sum_{t=1}^{\tau(A)} e_t(i_t)\right)\right], \tag{4.6}$$

where $T(i)$ is the number of rounds for which the game is feasible in the budget B when a fixed action $i \in [K]$ is performed, $i^* = \text{argmax}_{i\in[K]} \sum_{t=1}^{T(i)} e_t(i)$ is the optimal action in the hindsight,

$$z(\mathcal{A}) = \max\left\{\frac{B}{T(i^*)}, \frac{B(\mathcal{A})}{\tau(\mathcal{A})}\right\} \tag{4.7}$$

is the maximum cost per round, and $B(\mathcal{A})$ is the budget utilized by the algorithm $\mathcal{A}$. The expected regret is the expectation of the efficiency regret scaled by the maximum of the cost spent per round by the optimal action i^*, and the cost spent per round by the algorithm $\mathcal{A}$, where the efficiency regret is the sum of the rewards per unit cost associated to the optimal action minus the sum of the rewards per unit cost associated to the actions performed by the algorithm $\mathcal{A}$.

4.4 Adversarial BwK

In this section, we propose the algorithm EXP3.BwK for the adversarial BwK setting, and show that it is order optimal.

Similar to EXP3, EXP3.BwK maintains a set of time-varying weights $w_t(i)$ for each action $i \in [K]$. At each round t, an action $i_t = i$ is selected with probability $p_t(i)$ which is dependent on two parameters: the time-varying weights $w_t(i)$ and an exploration constant γ/K. Following the selection of the action i_t, the algorithm pays the cost $c_t(i_t)$. If the cost $c_t(i_t)$

Algorithm 6. EXP3.BwK

Initialization: γ ; For all $i \in [K]$, $w_1(i) = 1$, and $\hat{e}_1(i) = 0$; $t = 1$;
while $B > 0$ **do**
  $W_t = \sum_{j \in [K]} w_t(j)$
  Update $p_t(i) = (1 - \gamma) w_t(i)/W_t + \gamma/K$
  Choose $i_t = i$ with probability $p_t(i)$.
  Observe $(r_t(i_t), c_t(i_t))$
  if $c_t(i_t) > B$ **then**
    exit;
  end if
  $B = B - c_t(i_t)$
  For all $i \in [K]$, $\hat{e}_t(i) = r_t(i)\mathbf{1}(i = i_t)/p_t(i)c_t(i)$.
  $w_{t+1}(i) = w_t(i) \cdot \exp(\gamma c_{min} \cdot \hat{e}_t(i)/K)$
  $t = t + 1$
end while

is greater than the remaining budget of the algorithm, then the algorithm terminates without attempting to find other feasible actions which can be performed using the remaining budget. In EXP3.BwK, the efficiency $e_t(i) = r_t(i)/c_t(i)$ is used as a measure of the contribution from an action $i \in [K]$ at round t. The empirical estimate of the efficiency $\hat{e}_t(i)$ (defined in Algorithm 6) is used to update the weight $w_t(i)$ of the action i. For all $i \in [K]$, the difference in the weights $w_t(i)$ and $w_{t-1}(i)$ is controlled by scaling $\hat{e}_t(i)$ with γc_{min}, which ensures that the $\gamma c_{min} \hat{e}_t(i) \leq 1$. The probability $p_t(i)$ is dependent on $w_t(i)$ and the exploration constant γ/K. In the probability $p_t(i)$, the weight $w_t(i)$ is responsible for the exploitation as it favors the selection of an action with higher cumulative efficiencies i.e. $\sum_{n=1}^{t} \hat{e}_{t-1}(i)$ observed until round $t-1$. On contrary, the exploration constant γ/K ensures that the player is always exploring with a positive probability in search of the optimal action i^*. This balances the trade-off between exploration and exploitation.

In the literature of the adversarial BwK setup [247], it has been assumed that for all actions $i \in [K]$ and for all t, $r_t(i) \geq c_t(i)$. This allows the use of a different efficiency measure $r_t(i) - c_t(i)$, which is linear in both the reward and the cost of an action i, thus simplifying the proofs [247]. In many real life applications, the rewards and the costs are on different scales, and cannot be compared by an inequality operator. For example, in a recommendation system, a recommender is constrained by the total space available on the web page which corresponds

to the budget B, the space occupied by each item corresponds to its cost, and the click rate of each item corresponds to its reward. In this case, the space (cost) of the item and the click rate (reward) of the item are not comparable. Likewise, the efficiency measure $r_t(i) - c_t(i)$ which compares the reward and the cost of an action i on a linear scale, is questionable and provides no intuition about the optimality of an action. In EXP3.BwK, we use a different efficiency measure $r_t(i)/c_t(i)$ for tracking the contributions of each action $i \in [K]$. The use of this measure is motivated from the greedy algorithm $\mathcal{A}^G$, and its performance guarantees with respect to the optimal solution (see (4.4) and (4.6)). The advantages of using this measure are two folds. First, it eliminates the need of the assumption in [247]. Second, it can track $G(\mathcal{A})$ of the algorithm $\mathcal{A}$ irrespective of the measure of the rewards and the costs.

The following theorem provides the performance guarantees of EXP3.BwK in terms of the expected regret, and shows that it is sublinear in the budget B.

Theorem 14. *For $\gamma = \sqrt{c_{min} K \log(K)/B(e-1)}$, the expected regret, as defined in (4.6), of the algorithm EXP3.BwK is at most*

$$R(E) \leq 2\sqrt{\left((e-1) + (e-2)\frac{K}{B}\right)\frac{BK\log(K)}{c_{min}^3}}, \tag{4.8}$$

where E denotes EXP3.BwK.

Proof. We bound

$$\mathbb{E}\left[z(E)\left(\sum_{t=1}^{T(i^*)} \frac{r_t(i^*)}{c_t(i^*)} - \sum_{t=1}^{\tau(E)} \frac{r_t(i_t)}{c_t(i_t)}\right)\right]. \tag{4.9}$$

We show that

$$\mathbb{E}\left[\left(\sum_{t=1}^{T(i^*)} \frac{r_t(i^*)}{c_t(i^*)} - \sum_{t=1}^{\tau(E)} \frac{r_t(i_t)}{c_t(i_t)}\right)\right] \leq 2\sqrt{\left((e-1) + (e-2)\frac{K}{B}\right)\frac{BK\log(K)}{c_{min}^3}}, \tag{4.10}$$

and $z(E) \leq 1$. The detailed version of the proof is in the appendix. □

The key challenge in the proof of Theorem 14 is that the two summations in (4.6) corresponding to the optimal action i^* and the algorithm EXP3.BwK are along the different time scales, $T(i^*)$ and $\tau(E)$ respectively. This requires the analysis to be split into two cases: $T(i^*) \geq \tau(E)$ and $T(i^*) \leq \tau(E)$. The analysis for these cases is based on the inference that $B(E) > B - K$ because the algorithm EXP3.BwK terminates at round t if and only if the remaining budget is insufficient to pay the cost $c_t(i_t) \leq 1$. Hence, we can bound the difference between the two time scales i.e. $T(i^*)$ and $\tau(E)$ as follows:

$$|T(i^*) - \tau(E)| \leq \frac{K}{c_{min}}. \tag{4.11}$$

It follows that the difference between the number of rounds of the optimal action i^* and EXP3.BwK is bounded by a fixed constant independent of the budget B. Hence, the regret of the algorithm due to this difference in (4.11) is at most K/c_{min}^2, and does not introduce any dependency on the budget B.

The following theorem provides the lower bound on the expected regret in the adversarial BwK setting.

Theorem 15. *For any player's strategy $\mathcal{A}$, there exists an adversary for which the expected regret of the algorithm $\mathcal{A}$ is at least $\Omega(\sqrt{KB/c_{min}^2})$.*

Proof. The adversary chooses the optimal action i^* uniformly at random from the set of K actions. For the action i^* and for all t, the reward $r_t(i^*)$ is assigned using an independent Bernoulli random variable with expectation $0.5 + \epsilon$, where $\epsilon = \sqrt{Kc_{min}/B}$. For all $i \in [K]/\{i^*\}$ and for all t, the reward $r_t(i)$ is assigned using an independent Bernoulli random variable with expectation 0.5. For all $i \in [K]$ and for all t, the adversary assigns cost $c_t(i) = c_{min}$. The remaining proof is along the same lines as the lower bound on the expected regret in the MAB setup [17]. □

By comparing the results in Theorem 14 and Theorem 15, the expected regret of the

algorithm EXP3.BwK has an additional factor of $1/\sqrt{c_{min}}$, and is order optimal in the budget B. This also highlights an important feature of an alternate class of algorithms in the BwK setup. Consider a new class of algorithms $\mathcal{G}$ which looks for an alternative action to perform after the algorithm is unable to pay the cost $c_t(i_t)$ at round t in order to utilize the remaining budget effectively. Since EXP3.BwK terminates if it is unable to pay the cost $c_t(i_t)$, EXP3.BwK does not belong to $\mathcal{G}$, and is still order optimal in the budget B. Therefore, the expected regret of this new class of algorithms $\mathcal{G}$ will have same dependency as that of EXP3.BwK on the budget B. Additionally, the difference between the expected regret of EXP3.BwK and the class of algorithms $\mathcal{G}$ will be at most a constant i.e. K/c_{min}^2, independent of B (see (4.11)). The class of algorithms $\mathcal{G}$ faces the additional challenge of designing an appropriate criterion for the termination of the algorithm because the costs are assigned by the adversary.

The ideas developed in EXP3.BwK, particularly the measure of the efficiency $r_t(i)/c_t(i)$ forms form the basis of designing an algorithm which achieves almost optimal performance guarantees in both the stochastic and the adversarial BwK settings.

4.5 One practical algorithm for both stochastic and adversarial BwK

In this section, we propose the algorithm EXP3++.BwK (Algorithm 7), and show that it achieves almost optimal performance guarantees in both the stochastic and the adversarial BwK settings.

Before discussing the algorithm EXP3++.BwK, let us briefly focus on the fundamental difference between the optimal algorithms in the stochastic and the adversarial BwK settings. In the stochastic BwK setting, the algorithms focus on exploration in the initial stage until a reliable estimate of the expected rewards and expected costs is achieved. Then, the algorithms focus on exploitation, and perform exploration with a small probability. For example, in UCB type of algorithms, the probability of exploration decays as $1/t^2$ with round t [211, 52, 174]. In greedy

Algorithm 7. EXP3++.BwK

Initialization: For all $i \in [K]$, $w_1(i) = 1$, $\hat{e}_1(i) = 0$, $\bar{e}_1(i) = 0$, $N_1(i) = 1$ $\delta_1(i) > 0$; $t = 1$, $\gamma_t = 0.5 c_{min}\sqrt{\log(K)/tK}$;

Perform each action once and update for all $i \in [K]$, $\bar{e}_1(i) = r_1(i)/c_1(i)$, $B = B - \sum_{i\in[K]} c_1(i)$ and $t = K + 1$.

while $B > 0$ **do**

For all $i \in [K]$, update:

$\text{UCB}_t(i)$ (see (4.12))

$\text{LCB}_t(i)$ (see (4.13))

$\hat{\Delta}_t(i)$ (see (4.15))

$\delta_t(i) = \beta \log(t)/(t\hat{\Delta}_t(i)^2)$

$\epsilon_t(i) = \min\{1/2K, 0.5\sqrt{\log(K)/t}, \delta_t(i)\}$

$p_t(i) = \frac{\exp(-\gamma_t \hat{L}_{t-1}(i))}{\sum_{j\in[K]} \exp(-\gamma_t \hat{L}_{t-1}(j))}$

$\tilde{p}_t(i) = (1 - \sum_{j\neq i} \epsilon_t(j))p_t(i) + \epsilon_t(i)$

Choose $i_t = i$ with probability $\tilde{p}_t(i)$.

Observe $(r_t(i_t), c_t(i_t))$

if $c_t(i_t) > B$ **then**

exit;

end if

$B = B - c_t(i_t)$

For all $i \in [K]$, update:

$\hat{e}_t(i) = r_t(i)\mathbf{1}(i = i_t)/\tilde{p}_t(i)c_t(i)$.

$\hat{\ell}_t(i) = \mathbf{1}(i = i_t)/c_{min}\tilde{p}_t(i) - \hat{e}_t(i)$.

$\hat{L}_t(i) = \sum_{n=1}^{t} \hat{\ell}_n(i)$

$N_t(i) = N_{t-1}(i) + \mathbf{1}(i = i_t)$.

$\bar{r}_t(i) = \sum_{n=1}^{t} r_n(i)\mathbf{1}(i = i_n)/N_t(i)$

$\bar{c}_t(i) = \sum_{n=1}^{t} c_n(i)\mathbf{1}(i = i_n)/N_t(i)$

$\bar{e}_t(i) = \bar{r}_t(i)/\bar{c}_t(i)$

$t = t + 1$

end while

algorithms, the probability of exploration is zero after a fixed round (or time instance) [210, 213]. On the contrary, in the adversarial regime, the algorithms are always exploring, and looking for the actions with higher contributions [17]. For instance, in EXP3.BwK, the exploration constant γ/K does not change with the round t, and it is dependent on the total number of rounds i.e. $\Theta(B)$ in the BwK setup.

For all action $i \in [K]$, EXP3++.BwK maintains an Upper Confidence Bound (UCB)

$\text{UCB}_t(i)$ and a Lower Confidence Bound (LCB) $\text{LCB}_t(i)$ on the efficiency $e(i)$, where

$$\text{UCB}_t(i) = \min\left\{\frac{1}{c_{min}}, \bar{e}_t(i) + \frac{(1+1/\lambda)\eta_t(i)}{\lambda - \eta_t(i)}\right\}, \tag{4.12}$$

$$\text{LCB}_t(i) = \max\left\{0, \bar{e}_t(i) - \frac{(1+1/\lambda)\eta_t(i)}{\lambda - \eta_t(i)}\right\}, \tag{4.13}$$

$$\eta_t(i) = \sqrt{\frac{\alpha \log(K^{1/\alpha}t)}{2N_t(i)}}, \tag{4.14}$$

$\lambda \leq c_{min}$ and $N_t(i)$ is the number of times an action i has been chosen until round t. The UCB and the LCB on an action i are used to estimate $\Delta(i)$. The estimate of this gap at round t is defined as

$$\hat{\Delta}_t(i) = \max\{0, \max_{j \neq i} \text{LCB}_t(j) - \text{UCB}_t(i)\}. \tag{4.15}$$

It can been shown that for all $i \in [K]$, in the stochastic BwK setting, we have

$$\frac{\Delta(i)}{2} \leq \hat{\Delta}_t(i) \leq \Delta(i),$$

with high probability as $t \to \infty$. Thus, $\hat{\Delta}_t(i)$ is a reliable estimate of $\Delta(i)$. For all $i \in [K]$, the estimate of the gap $\hat{\Delta}_t(i)$ is used to design the exploration parameter $\epsilon_t(i)$ in the sampling probability $\tilde{p}_t(i)$ where $\tilde{p}_t(i)$ is the probability of choosing an action i at round t. In the stochastic BwK setup, since $\Delta(i^*) = 0$, the exploration parameter $\epsilon_t(i^*)$ of the optimal action i^* tends to zero, and favors its selection. Unlike EXP3.BwK, the exploration parameter $\epsilon_t(i)$ varies with t. Additionally, the sampling probability $\tilde{p}_t(i)$ is dependent on both the estimates of the efficiencies $\hat{e}_t(i)$ and $\bar{e}_t(i)$ where $\hat{e}_t(i)$ and $\bar{e}_t(i)$ are crucial in the adversarial BwK setting (see EXP3.BwK) and the stochastic BwK setting respectively. In the sampling probability $\tilde{p}_t(i)$, $\hat{e}_t(i)$ controls the exploitation performed by the algorithm through $p_t(i)$, and $\bar{e}_t(i)$ controls the exploration performed by the algorithm through the exploration parameter $\epsilon_t(i)$.

The following theorem provides the performance guarantees of EXP3++.BwK in the stochastic BwK setting.

Theorem 16. *In the stochastic BwK setting, for $\alpha = 3$ and $\beta = 256/c_{min}^2$, the expected regret of the EXP3++.BwK is at most*

$$R(F) = O\left(\sum_{i:\Delta(i)>0} \frac{\log^2(B/c_{min})}{c_{min}^2 \Delta(i)}\right), \tag{4.16}$$

where F denotes the algorithm EXP3++.BwK.

Proof. The expected regret of the algorithm can be bounded by

$$R(F) \leq \sum_{i\in[K]/\{i^*\}} \Delta(i)\mathbb{E}[N_T(i)], \tag{4.17}$$

where $T \leq B/c_{min}$ is the number of rounds at the termination of the algorithm. We can then bound the expected number of times $\mathbb{E}[N_T(i)]$ an action $i \neq i^*$ is selected by the algorithm. Since the probability of the selection of an action i is $\tilde{p}_t(i)$, we have

$$\mathbb{E}[N_T(i)] \leq \mathbb{E}[\sum_{t=1}^{T} \epsilon_t(i) + p_t(i)]. \tag{4.18}$$

We now bound the two terms in the right hand side of (4.18) in the stochastic BwK setting. First, we show that the estimate $\hat{\Delta}_t(i)$ is a reliable estimate of $\Delta(i)$, i.e.

$$\mathbb{P}(\hat{\Delta}_t(i) \geq \Delta(i)) \leq \frac{1}{t^{\alpha-1}}, \tag{4.19}$$

$$\mathbb{P}\left(\hat{\Delta}_t(i) \leq \frac{\Delta(i)}{2}\right) \leq \left(\frac{\log t}{t c_{min}^2 \Delta(i)^2}\right)^{\alpha-2} + 2\left(\frac{1}{t}\right)^{\frac{\beta c_{min}^2}{8}} + \frac{2}{K t^{\alpha-1}}. \tag{4.20}$$

These results can be used to prove that

$$\mathbb{P}\left(\tilde{\Delta}_t(i) \leq \frac{t\Delta(i)}{2}\right) \leq \left(\frac{\log(t)}{tc_{min}^2\Delta(i)^2}\right)^{\alpha-2} + \frac{1}{t}, \tag{4.21}$$

where $\tilde{\Delta}_t(i) = \sum_{n=1}^{t}(\hat{\ell}_n(i) - \hat{\ell}_n(i^*))$. Since

$$p_t(i) \leq \exp(-\gamma_t\tilde{\Delta}_t(i)), \tag{4.22}$$

(4.21) is used to bound $\sum_{t=1}^{T}\mathbb{E}[p_t(i)]$, and we have

$$\sum_{t=1}^{T}\mathbb{E}[p_t(i)] = O\left(\frac{\log^2(B/c_{min})}{c_{min}^2\Delta(i)^2}\right). \tag{4.23}$$

Using the definition of $\epsilon_t(i)$ and (4.20), we have

$$\sum_{t=1}^{T}\mathbb{E}[\epsilon_t(i)] = O\left(\frac{\log^2(B/c_{min})}{c_{min}^2\Delta(i)^2}\right). \tag{4.24}$$

Hence, the statement of the theorem follows. The detailed version of the proof is in the appendix

□

In Theorem 16, EXP3++.BwK incurs an expected regret of $O(\log^2(B/c_{min}))$, whereas the optimal regret guarantees in the stochastic BwK setting are given by $O(\log(B/c_{min}))$ [211, 52, 174]. Thus, EXP3++.BwK has an additional factor of $\log(B/c_{min})$ in comparison to the results in the literature. This additional factor is also common in the literature of MAB [195, 139]. The following theorem provides the performance guarantees of EXP3++.BwK in the adversarial BwK setting.

Theorem 17. *In the adversarial BwK setting, the expected regret of the EXP3++.BwK is at most*

$$R(F) \leq \sqrt{\frac{6BK\log(K)}{c_{min}^3}}. \tag{4.25}$$

Proof. Similar to the proof of Theorem 14, we bound

$$\mathbb{E}\left[z(E)\left(\sum_{t=1}^{T(i^*)}\frac{r_t(i^*)}{c_t(i^*)}-\sum_{t=1}^{\tau(E)}\frac{r_t(i_t)}{c_t(i_t)}\right)\right]. \tag{4.26}$$

We show that

$$\mathbb{E}\left[\left(\sum_{t=1}^{T(i^*)}\frac{r_t(i^*)}{c_t(i^*)}-\sum_{t=1}^{\tau(E)}\frac{r_t(i_t)}{c_t(i_t)}\right)\right]\leq\sqrt{\frac{6BK\log(K)}{c_{min}^3}}, \tag{4.27}$$

and $z(E)\leq 1$. The detailed version of the proof is in supplementary material. □

Thus, like EXP3.BwK, EXP3++.BwK is order optimal in the adversarial BwK setting. The challenges in the proof of Theorem 16 and Theorem 17 are addressed in a similar way as that of Theorem 14. In conclusion, using Theorem 16 and Theorem 17, the EXP3++.BwK is order optimal in the adversarial BwK setting and has an additional factor of $\log(B/c_{min})$ in the stochastic BwK setting.

4.6 BwK with unbounded cost

Assuming the cost is bounded by unity (i.e., $c_{max}=1$), Theorem 15 provides the dependence of the expected regret on the minimum cost c_{min} in the adversarial BwK setup. In this section, we discuss the scaling of the lower bound on the expected regret with respect to the maximum cost c_{max} in the adversarial BwK setup.

Theorem 18. *Suppose that* $c_{\max}=B^\alpha$. *For any algorithm* $\mathcal{A}$, *there exists an adversary such that the expected regret of the algorithm is at least* $\Omega(B^\alpha)$.

Proof. Let the number of actions be $K=2$, and the actions be i_1,i_2. The adversary chooses the optimal action i^* uniformly at random from these two actions. Let $t^*=B-B^\alpha$. For all $t\leq t^*$ rounds, the adversary assigns $r_t(i_1)=r_t(i_2)=0$ and $c_t(i_1)=c_t(i_2)=1$ to both the actions i_1 and i_2. Now, for rounds $t\geq t^*+1$, the adversary assigns $r_t(i^*)=1$ and $c_t(i^*)=1$

to the optimal action i^*. For the suboptimal action $i \neq i^*$, the adversary assigns $r_{t^*+1}(i) = 0$ and $c_{t^*+1}(i) = B^\alpha$ (since $c_{\max} = B^\alpha$, this is a valid cost assignment), and $r_t(i) = c_t(i) = 1$ for $t > t^* + 1$.

Let S_1 be the case when $i^* = i_1$, and S_2 be the case when $i^* = i_2$. For the first t^* rounds, any algorithm $\mathcal{A}$ would have the same behavior in both the cases S_1 and S_2. Now, at round $t^* + 1$, assume that this algorithm $\mathcal{A}$ selects an action i_1 and i_2 with probability p and $(1 - p)$ respectively. Note that if the suboptimal action is chosen at round $t^* + 1$, then the budget is depleted and the sum of the rewards is 0. On the other hand, if i^* is chosen at $t^* + 1$, the algorithm receives a sum of B^α rewards in the end. Thus, if $i_{t^*+1} \neq i^*$, then the regret of the algorithm is B^α. This implies that the expected regret of the algorithm is $0.5pB^\alpha + 0.5(1 - p)B^\alpha = B^\alpha/2$. The statement of the theorem follows. □

In the literature of BwK, the cost is always considered to be bounded above by a constant independent of the budget B. Here, we consider that the cost is bounded by a function of the budget B. Theorem 18 shows that the lower bound on the expected regret scales at least linearly with the maximum cost c_{max} in the adversarial BwK setup. If $\alpha > 1/2$, then it is impossible to achieve a regret bound of $O(\sqrt{B})$, which is order optimal in cases with small c_{max}.

In the adversarial BwK setup, the adversary can penalize the player in two ways. First, the adversary can control the reward of an action at any round. Second, the adversary can control the cost of an action, which is analogous to penalizing the player on the number of rounds T. For $\alpha > 1/2$, the latter penalty on the number of rounds T becomes significant, and the minimum achievable regret is no longer $\Omega(\sqrt{B})$. In this setting with $\alpha > 1/2$, the design of algorithms which achieve regret of $O(B^\alpha)$ is left as future work.

4.7 Conclusion

The study of BwK has been mostly focused on the stochastic regime. In this work, we considered the adversarial regime and proposed the order optimal algorithm EXP3.BwK for

this setting. We also used ideas from the adversarial BwK setup to design EXP3++.BwK. This algorithm has an expected regret of $O(\sqrt{KB\log(K)})$ and $O(\log^2(B))$ in the adversarial and stochastic settings respectively. Thus, the algorithm is order optimal in the adversarial regime, and has an additional factor of $\log(B)$ in the stochastic regime. It is the first algorithm that provides almost optimal performance guarantees in both stochastic and adversary BwK settings. As part of future work, we are considering designing an algorithm which achieves the optimal regret guarantees with high probability in both the adversarial and the stochastic BwK settings.

All the results in the literature of BwK assume that the maximum cost is bounded by a constant independent of B. We have shown that if the cost is $O(B^\alpha)$, then the expected regret is at least $\Omega(B^\alpha)$. Thus, the minimum expected regret scales at least linearly with the maximum cost of the BwK setup. This setting is of particular interest when $\alpha > 1/2$ because the expected regret of $O(\sqrt{B})$, which is achievable in the setting where cost is bounded by a constant, becomes unachievable. Hence, there is a need to study this BwK setting, and design optimal algorithms whose expected regret is $O(B^\alpha)$, which is left as a future work.

4.8 Acknowledgment

Chapter 4, in full, is a reprint of the material as it appears in Anshuka Rangi, Massimo Franceschetti and Long Tran-Thanh, "Unifying the Stochastic and the Adversarial Bandits with Knapsack", *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI)*, August 2019. The dissertation author was the primary investigator and author of this paper.

4.9 Appendix

4.9.1 Proof of Theorem 1

Proof. Let $T = \max\{T(i^*), \tau(E)\}$, where

$$i^* = \text{argmax}_{i\in[K]} \sum_{t=1}^{T(i)} \frac{r_t(i)}{c_t(i)}. \tag{4.28}$$

Additionally, we have

$$\sum_{i\in[K]} p_t(i)\hat{e}_t(i) = p_t(i_t)\frac{r_t(i_t)}{p_t(i_t)\cdot c_t(i_t)} = \frac{r_t(i_t)}{c_t(i_t)}, \tag{4.29}$$

and

$$\begin{aligned}\sum_{i\in[K]} p_t(i)\hat{e}_t(i)^2 &= p_t(i_t)\frac{r_t(i_t)}{p_t(i_t)\cdot c_t(i_t)}\hat{e}_t(i_t)\\ &\overset{(a)}{\leq} \frac{\hat{e}_t(i_t)}{c_{min}}\\ &= \frac{\sum_{i\in[K]}\hat{e}_t(i)}{c_{min}},\end{aligned} \tag{4.30}$$

where (a) follows from the fact that for all $i \in [K]$, $r_t(i)/c_t(i) \leq 1/c_{min}$. Also, for all $i \in [K]$, we have

$$\mathbf{E}\Big[\hat{e}_t(i)|\{p_t(j)\}_{j\in[K]}\Big] = p_t(i)\cdot\hat{e}_t(i) + (1-p_t(i))\cdot 0 = \frac{r_t(i)}{c_t(i)}. \tag{4.31}$$

Since $W_t = \sum_{j\in[K]} w_t(j)$, we have

$$\begin{aligned}
\frac{W_{t+1}}{W_t} &= \sum_{i\in[K]} \frac{w_{t+1}(i)}{W_t} \\
&= \sum_{i\in[K]} \frac{w_t(i)\exp\left(\gamma c_{min}\cdot \hat{e}_t(i)/K\right)}{W_t} \\
&\overset{(a)}{=} \sum_{i\in[K]} \frac{p_t(i)-\gamma/K}{1-\gamma}\cdot \exp\left(\gamma c_{min}\cdot \hat{e}_t(i)/K\right) \\
&\overset{(b)}{\leq} \sum_{i\in[K]} \frac{p_t(i)-\gamma/K}{1-\gamma}\left(1+\frac{\gamma c_{min}}{K}\hat{e}_t(i)+(e-2)\left(\frac{\gamma c_{min}}{K}\hat{e}_t(i)\right)^2\right) \\
&\overset{(c)}{\leq} 1+\frac{c_{min}\gamma/K}{(1-\gamma)}\sum_{i\in[K]} p_t(i)\hat{e}_t(i)+\frac{(e-2)c_{min}^2(\gamma/K)^2}{(1-\gamma)}\sum_{i\in[K]} p_t(i)\hat{e}_t(i)^2, \qquad (4.32)
\end{aligned}$$

where (a) follows from the definition of $w_t(i)$, (b) follows from the facts that for all $i \in [K]$, we have $p_t(i) > \gamma/K$ and for all $x \leq 1$, we have $e^x \leq 1 + x + (e-2)x^2$, and (c) follows from the fact that $\sum_{i\in[K]} p_t(i) = 1$ and $\gamma/K > 0$.

Now, taking logs on both sides of (4.32), summing over $1, 2, \ldots T+1$, and using $\log(1+x) \leq x$ for all $x > -1$, we get

$$\log\frac{W_{T+1}}{W_1} \leq \frac{c_{min}\gamma/K}{(1-\gamma)}\sum_{t=1}^{T}\sum_{i\in[K]} p_t(i)\hat{e}_t(i)+\frac{(e-2)c_{min}^2(\gamma/K)^2}{(1-\gamma)}\sum_{t=1}^{T}\sum_{i\in[K]} p_t(i)\hat{e}_t(i)^2. \quad (4.33)$$

Additionally, for all $j \in [K]$, we have

$$\begin{aligned}
\log\frac{W_{T+1}}{W_1} &\geq \log\frac{w_{T+1}(j)}{W_1} \\
&= \frac{c_{min}\gamma}{K}\sum_{t=1}^{T}\hat{e}_t(j) - \log(K). \qquad (4.34)
\end{aligned}$$

Combining (4.33) and (4.34), for all $j \in [K]$, we have

$$\frac{c_{min}\gamma}{K}\sum_{t=1}^{T}\hat{e}_t(j) - \log(K) \leq \frac{c_{min}\gamma/K}{(1-\gamma)}\sum_{t=1}^{T}\frac{r_t(i_t)}{c_t(i_t)} + \frac{(e-2)c_{min}^2(\gamma/K)^2}{c_{min}(1-\gamma)}\sum_{t=1}^{T}\sum_{i\in[K]}\hat{e}_t(i), \tag{4.35}$$

where the right hand side of the above equation follows from (4.29) and (4.30). We will split the analysis into two cases: $T(i^*) \leq \tau(E)$ and $T(i^*) > \tau(E)$. For $T(i^*) \leq \tau(E)$, using (4.35), we have

$$\frac{\gamma}{K}\sum_{t=1}^{T(i^*)}\hat{e}_t(i^*) - \frac{\log(K)}{c_{min}} \leq \frac{\gamma/K}{(1-\gamma)}\sum_{t=1}^{\tau(E)}\frac{r_t(i_t)}{c_t(i_t)} + \frac{(e-2)(\gamma/K)^2}{(1-\gamma)}\sum_{t=1}^{\tau(E)}\sum_{i\in[K]}\hat{e}_t(i), \tag{4.36}$$

where the inequality follows by replacing $T = \tau(E)$, and using the facts that $T(i^*) \leq \tau(E)$ and $\hat{e}_t(i^*)$ is non-negative.

Now, for $T(i^*) > \tau(E)$, using (4.35), we have

$$\begin{aligned}
&\frac{\gamma}{K}\sum_{t=1}^{T(i^*)}\hat{e}_t(i^*) - \frac{\log(K)}{c_{min}} \\
&\leq \frac{\gamma/K}{(1-\gamma)}\sum_{t=1}^{T(i^*)}\frac{r_t(i_t)}{c_t(i_t)} + \frac{(e-2)(\gamma/K)^2}{(1-\gamma)}\sum_{t=1}^{T(i^*)}\sum_{i\in[K]}\hat{e}_t(i), \\
&\overset{(a)}{=} \frac{\gamma/K}{(1-\gamma)}\sum_{t=1}^{\tau(E)}\frac{r_t(i_t)}{c_t(i_t)} + \frac{(e-2)(\gamma/K)^2}{(1-\gamma)}\sum_{t=1}^{T(i^*)}\sum_{i\in[K]}\hat{e}_t(i),
\end{aligned} \tag{4.37}$$

where (a) follows from the fact that for all $t > \tau(E)$, we have $r_t(i_t)/c_t(i_t) = 0$. Therefore,

(4.37) can be further simplified as

$$\begin{aligned}
&\frac{\gamma}{K}\sum_{t=1}^{T(i^*)}\hat{e}_t(i^*) - \frac{\log(K)}{c_{min}} \\
&\leq \frac{\gamma/K}{(1-\gamma)}\sum_{t=1}^{\tau(E)}\frac{r_t(i_t)}{c_t(i_t)} + \\
&\qquad \frac{(e-2)(\gamma/K)^2}{(1-\gamma)}\left(\sum_{t=1}^{\tau(E)}\sum_{i\in[K]}\hat{e}_t(i) + \sum_{t=\tau(E)+1}^{T(i^*)}\sum_{i\in[K]}\hat{e}_t(i)\right). \qquad (4.38)
\end{aligned}$$

Combining (4.36) and (4.38), taking expectation on both the sides of (4.38), and using (4.31), we have

$$\begin{aligned}
&\sum_{t=1}^{T(i^*)}\frac{r_t(i^*)}{c_t(i^*)} - \mathbb{E}\left[\sum_{t=1}^{\tau(E)}\frac{r_t(i_t)}{c_t(i_t)}\right] \\
&\leq \frac{K}{c_{min}\gamma}\log(K) + \gamma\sum_{t=1}^{T(i^*)}\frac{r_t(i^*)}{c_t(i^*)} + \frac{(e-2)\gamma}{K}\mathbb{E}\left[\sum_{t=1}^{\tau(E)}\sum_{i\in[K]}\frac{r_t(i)}{c_t(i)}\right] \\
&\qquad + \frac{(e-2)\gamma}{K}\mathbb{P}(T(i^*) > \tau(E))\mathbb{E}\left[\sum_{t=\tau(E)+1}^{T(i^*)}\sum_{i\in[K]}\hat{e}_t(i)\right]. \qquad (4.39)
\end{aligned}$$

Since $B(E) \geq B - K$, we have that $|T(i^*) - \tau(E)| \leq K/c_{min}$. Using $G(\mathcal{A}^*) \leq B/c_{min}^2$ and $T(i^*) - \tau(E) \leq K/c_{min}$, we have that

$$\begin{aligned}
\sum_{t=1}^{T(i^*)}\frac{r_t(i^*)}{c_t(i^*)} - \mathbb{E}\left[\sum_{t=1}^{\tau(E)}\frac{r_t(i_t)}{c_t(i_t)}\right] &\leq \frac{K}{c_{min}\gamma}\log(K) + \gamma\cdot\frac{B}{c_{min}^2} \\
&\quad + (e-2)\gamma\cdot\left(\frac{B}{c_{min}^2} + \frac{K}{c_{min}^2}\right). \qquad (4.40)
\end{aligned}$$

Using $\gamma = \sqrt{c_{min}K\log(K)/(B(e-1)+K(e-2))}$, the right hand side of the above equation is bounded by

$$2\sqrt{\frac{((e-1)B + (e-2)K)K\log(K)}{c_{min}^3}}. \qquad (4.41)$$

Since for all t, we have that $c_t(i^*) \leq 1$, $T(i^*) \geq B$ and $B(E) \geq B - K$. Also, we have $\tau(E) \leq B/c_{min}$. Thus,

$$z(E) \leq 1. \tag{4.42}$$

Combining (4.41) and (4.42), the statement of the theorem follows. □

4.9.2 Proof of Theorem 3

Proof. Let $T = \max\{T(i^*), \tau(E)\}$. The proof of the theorem is split into following results.

In Lemma 9, we show that for all $i \in [K]$, the efficiency $e(i)$ is

$$\text{LCB}_t(i) \leq e(i) \leq \text{UCB}_t(i),$$

with high probability as $t \to \infty$ (see Lemma 9), namely

$$\mathbb{P}(\text{UCB}_t(i) \leq e(i)) \leq \frac{1}{Kt^{\alpha-1}}, \tag{4.43}$$

$$\mathbb{P}(\text{LCB}_t(i) \geq e(i)) \leq \frac{1}{Kt^{\alpha-1}}. \tag{4.44}$$

This is used to show that $\hat{\Delta}_t(i) \leq \Delta(i)$ with high probability as $t \to \infty$ (see Lemma 10), namely

$$\mathbb{P}(\hat{\Delta}_t(i) \geq \Delta(i)) \leq \frac{1}{t^{\alpha-1}}. \tag{4.45}$$

Using Lemma 11 and Lemma 12, we show that (see Lemma 13)

$$\mathbb{P}\left(\hat{\Delta}_t(i) \leq \frac{\Delta(i)}{2}\right) \leq \left(\frac{\log t}{tc_{min}^2\Delta(i)^2}\right)^{\alpha-2} + 2\left(\frac{1}{t}\right)^{\frac{\beta c_{min}^2}{8}} + \frac{2}{Kt^{\alpha-1}}. \tag{4.46}$$

Thus, using (4.45) and (4.46), we have

$$\frac{\Delta(i)}{2} \leq \hat{\Delta}_t(i) \leq \Delta(i),$$

with high probability as $t \to \infty$.

Using Lemma 14 and 15, we have

$$\mathbb{P}\left(\tilde{\Delta}_t(i) \leq \frac{t\Delta(i)}{2}\right) \leq \left(\frac{\log(t)}{tc_{min}^2\Delta(i)^2}\right)^{\alpha-2} + \frac{1}{t}, \tag{4.47}$$

where $\tilde{\Delta}_t(i) = \sum_{n=1}^{t}(\hat{\ell}_n(i) - \hat{\ell}_n(i^*))$. Since $p_t(i) \leq \exp(-\gamma_t\tilde{\Delta}_t(i))$, (4.47) is used to bound $\sum_{t=1}^{T}\mathbb{E}[p_t(i)]$, and we have

$$\sum_{t=1}^{T}\mathbb{E}[p_t(i)] = O\left(\frac{\log^2(B/c_{min})}{c_{min}^2\Delta(i)^2}\right). \tag{4.48}$$

Using the definition of $\epsilon_t(i)$ and (4.47), we have

$$\sum_{t=1}^{T}\mathbb{E}[\epsilon_t(i)] = O\left(\frac{\log^2(B/c_{min})}{c_{min}^2\Delta(i)^2}\right). \tag{4.49}$$

Hence, the statement of the theorem follows. □

Lemma 9. *For all $i \in [K]$ and $t \geq K$, we have*

$$\mathbb{P}(UCB_t(i) \leq e(i)) \leq \frac{1}{Kt^{\alpha-1}}, \tag{4.50}$$

$$\mathbb{P}(LCB_t(i) \geq e(i)) \leq \frac{1}{Kt^{\alpha-1}}, \tag{4.51}$$

Proof. If $\text{UCB}_t(i) \leq e(i)$, then we have

$$\bar{e}_t(i) + \frac{(1+1/\lambda)\eta_t(i)}{\lambda - \eta_t(i)} \leq e(i) = \frac{\mu(i)}{\rho(i)}.$$

Therefore, at least one of the events U_1 and U_2 is true, where

$$U_1 : \bar{r}_t(i) \leq \mu(i) - \eta_t(i),$$

$$U_2 : \bar{c}_t(i) \geq \rho(i) + \eta_t(i).$$

This can be proved by contradiction. Let both U_1 and U_2 are false. Then, we have

$$\begin{aligned}
\frac{\mu(i)}{\rho(i)} - \frac{\bar{r}_t(i)}{\bar{c}_t(i)} &= \frac{\mu(i)\bar{c}_t(i) - \rho(i)\bar{r}_t(i)}{\rho(i)\bar{c}_t(i)} \\
&= \frac{\mu(i)(\bar{c}_t(i) - \rho(i)) + \rho(i)(\mu(i) - \bar{r}_t(i))}{\rho(i)\bar{c}_t(i)} \\
&\overset{(a)}{\leq} \frac{\mu(i)\eta_t(i) + \rho(i)\eta_t(i)}{\rho(i)\bar{c}_t(i)} \\
&\overset{(b)}{\leq} \frac{\eta_t(i)}{\lambda(\lambda - \eta_t(i))} + \frac{\eta_t(i)}{\lambda - \eta_t(i)} \\
&= \frac{(1+1/\lambda)\eta_t(i)}{\lambda - \eta_t(i)},
\end{aligned} \tag{4.52}$$

where (a) follows from the fact that both U_1 and U_2 are false, and (b) follows from the facts that U_1 and U_2 are false, and $\lambda \leq c_{min}$. Hence, at least one of the events U_1 and U_2 is true. Now, using Hoeffding's inequality, we have

$$\mathbb{P}(U_1) \leq \frac{1}{Kt^{\alpha}}, \tag{4.53}$$

and

$$\mathbb{P}(U_2) \leq \frac{1}{Kt^{\alpha}}. \tag{4.54}$$

Thus,

$$\begin{aligned}\mathbb{P}(\text{UCB}_t(i) \leq e(i)) &\leq \mathbb{P}(U_1) + \mathbb{P}(U_2) \\ &\leq \frac{1}{Kt^{\alpha-1}}.\end{aligned} \tag{4.55}$$

Similarly, if $\text{LCB}_t(i) \geq e(i)$, then we have

$$\bar{e}_t(i) - \frac{(1+1/\lambda)\eta_t(i)}{\lambda - \eta_t(i)} \geq e(i) = \frac{\mu(i)}{\rho(i)}. \tag{4.56}$$

Therefore, at least one of the events L_1 and L_2 is true, where

$$L_1 : \bar{r}_t(i) \geq \mu(i) + \eta_t(i),$$

$$L_2 : \bar{c}_t(i) \leq \rho(i) - \eta_t(i).$$

This can be proved by contradiction. Now, using Hoeffding's inequality, we have

$$\mathbb{P}(L_1) \leq \frac{1}{Kt^{\alpha}}, \tag{4.57}$$

and

$$\mathbb{P}(L_2) \leq \frac{1}{Kt^{\alpha}}. \tag{4.58}$$

Thus, we have

$$\begin{aligned}\mathbb{P}(\text{LCB}_t(i) \geq e(i)) &\leq \mathbb{P}(L_1) + \mathbb{P}(L_2) \\ &\leq \frac{1}{Kt^{\alpha-1}}.\end{aligned} \tag{4.59}$$

Hence proved. □

Lemma 10. *For all $i \in [K]$ and $t \geq K$,*

$$\mathbb{P}(\hat{\Delta}_t(i) \geq \Delta(i)) \leq \frac{1}{t^{\alpha-1}}, \tag{4.60}$$

Proof. Since $\Delta(i) = \max_{j\in[K]} e(j) - e(i)$, we have

$$\begin{aligned}\mathbb{P}(\hat{\Delta}_t(i) \geq \Delta(i)) &= \mathbb{P}(\max_{j\neq i} \text{LCB}_t(j) - \text{UCB}_t(i) \geq \Delta(i)) \\ &\leq \sum_{j\neq i} \mathbb{P}(\text{LCB}_t(j) \geq e(j)) + \mathbb{P}(\text{UCB}_t(i) \leq e(i)) \\ &\leq \frac{1}{t^{\alpha-1}}, \end{aligned} \tag{4.61}$$

where the last inequality follows from Lemma 9. Hence proved. □

Lemma 11. *For all $i \in [K]$, let*

$$t_{min}(i) = \min\{t : t \geq 4K\beta(\log t)^2/\Delta(i)^4 \log(K)\}.$$

We define two events $A(i,t)$ and $A(i^*,i,t)$ as

$$A(i,t) = \left\{ \textit{there exists an } n \leq t : \epsilon_n(i) < \frac{\beta \log t}{t\Delta(i)^2} \right\}, \tag{4.62}$$

$$A(i^*,i,t) = \left\{ \textit{there exists an } n \leq t : \epsilon_n(i^*) < \frac{\beta \log t}{t\Delta(i)^2} \right\}. \tag{4.63}$$

For $t > t_{min}(i)$ and $\alpha \geq 3$, we have

$$\mathbb{P}(A(i,t)) \leq \frac{1}{2}\left(\frac{\log t}{tc_{min}^2\Delta(i)^2}\right)^{\alpha-2}, \tag{4.64}$$

$$\mathbb{P}(A(i^*,i,t)) \leq \frac{1}{2}\left(\frac{\log t}{tc_{min}^2\Delta(i)^2}\right)^{\alpha-2}. \tag{4.65}$$

Proof. We start with proving the bound on the probability of the event $A(i,t)$. This proof is divided into two parts. First, for $n \leq tc_{min}^2\Delta(i)^2/\log(t)$, using the Lemma 10, we show that $A(i,t)$ does not occur with high probability as $t \to \infty$. Later, for $n \geq tc_{min}^2\Delta(i)^2/\log(t)$, we bound the probability of the event $A(i,t)$ using the Lemma 10.

For $n \leq tc_{min}^2\Delta(i)^2/\log(t)$, we have

$$\begin{aligned}\frac{\beta\log(n)}{n\hat{\Delta}_n^2(i)} &\overset{(a)}{\geq} \frac{\beta c_{min}^2\log(n)}{n} \\ &\overset{(b)}{\geq} \frac{\beta\log(n)\log(t)}{t\Delta(i)^2} \\ &\geq \frac{\beta\log(t)}{t\Delta(i)^2},\end{aligned} \tag{4.66}$$

where (a) follows from the definition of $\hat{\Delta}_n(i)$, and (b) follows from the range of n. For $t \geq t_{min}$, we have

$$0.5\sqrt{\frac{\log(K)}{tK}} \geq \frac{\beta\log(t)}{t\Delta(i)^2}. \tag{4.67}$$

Additionally, using Lemma 10, we have that $\hat{\Delta}_n(i) \leq \Delta(i)$ w.h.p as $n \to \infty$. Therefore, combining the fact $\hat{\Delta}_n(i) \leq \Delta(i)$ along with (4.67) and (4.66), we have

$$\epsilon_n(i) \geq \frac{\beta\log t}{t\Delta(i)^2}. \tag{4.68}$$

Now, for $n \geq tc_{min}^2\Delta(i)^2/\log(t)$, we have

$$\begin{aligned}&\mathbb{P}\left(\text{There exists } n \in \left[\frac{tc_{min}^2\Delta(i)^2}{\log(t)}, t\right] : \epsilon_n(i) < \frac{\beta\log t}{t\Delta(i)^2}\right) \\ &= \mathbb{P}\left(\text{There exists } n \in \left[\frac{tc_{min}^2\Delta(i)^2}{\log(t)}, t\right] : \hat{\Delta}_n(i) \geq \Delta(i)\right) \\ &\leq \sum_{n=\frac{tc_{min}^2\Delta(i)^2}{\log(t)}}^{t} \frac{1}{n^{\alpha-1}} \leq \frac{1}{2}\left(\frac{\log t}{tc_{min}^2\Delta(i)^2}\right)^{\alpha-2}.\end{aligned} \tag{4.69}$$

Similarly, we can bound the probability of $\mathbb{P}(A(i^*, i, t))$ by using the fact that $\Delta(i^*) = 0 < \Delta(i)$ for $i \neq i^*$. Hence proved. □

Lemma 12. *For all $i \in [K]$ and $t \geq t_{min}(i)$, we have*

$$\mathbb{P}\left(N_t(i) \leq \frac{\beta \log t}{2\Delta(i)^2}\right) \leq \left(\frac{1}{t}\right)^{\frac{\beta c_{min}^2}{8}} + \frac{1}{2}\left(\frac{\log t}{t c_{min}^2 \Delta(i)^2}\right)^{\alpha-2}. \tag{4.70}$$

Additionally,

$$\mathbb{P}\left(N_t(i^*) \leq \frac{\beta \log t}{2\Delta(i)^2}\right) \leq \left(\frac{1}{t}\right)^{\frac{\beta c_{min}^2}{8}} + \frac{1}{2}\left(\frac{\log t}{t c_{min}^2 \Delta(i)^2}\right)^{\alpha-2}. \tag{4.71}$$

Proof. We have

$$\begin{aligned}
\mathbb{P}\left(N_t(i) \leq \frac{\beta \log t}{2\Delta(i)^2}\right) &\leq \mathbb{P}\left(A^C(i,t) \text{ and } N_t(i) \leq \frac{\beta \log t}{2\Delta(i)^2}\right) + \mathbb{P}\left(A(i,t)\right) \\
&\overset{(a)}{\leq} \exp\left(\frac{-\beta \log t}{8\Delta(i)^2}\right) + \frac{1}{2}\left(\frac{\log t}{t c_{min}^2 \Delta(i)^2}\right)^{\alpha-2} \\
&\overset{(b)}{\leq} \left(\frac{1}{t}\right)^{\frac{\beta c_{min}^2}{8}} + \frac{1}{2}\left(\frac{\log t}{t c_{min}^2 \Delta(i)^2}\right)^{\alpha-2},
\end{aligned} \tag{4.72}$$

where $A^C(i,t)$ is the complement of the event $A(i,t)$, (a) follows from the Theorem 8 in [194] and Lemma 11, and (b) follows from the fact that for all $i \in [K]$, $\Delta(i) \leq 1/c_{min}^2$. Similarly, we can bound the probability in (4.71). □

Lemma 13. *For all $i \in [K]$, $t \geq t_{min}(i)$, $\alpha \geq 3$ $\beta \geq 64(\alpha+1)/c_{min}^2 \geq 256/c_{min}^2$, we have*

$$\mathbb{P}\left(\hat{\Delta}_t(i) \leq \frac{\Delta(i)}{2}\right) \leq \left(\frac{\log t}{t c_{min}^2 \Delta(i)^2}\right)^{\alpha-2} + 2\left(\frac{1}{t}\right)^{\frac{\beta c_{min}^2}{8}} + \frac{2}{K t^{\alpha-1}}. \tag{4.73}$$

Proof. Using Lemma 9, we have

$$\mathbb{P}\big((\text{UCB}_t(i^*) \leq e(i^*)) \text{ or } (\text{LCB}_t(i) \geq e(i))\big) \leq 2/Kt^{\alpha-1}. \tag{4.74}$$

Now, assume $\text{UCB}_t(i^*) \geq e(i^*)$ and $\text{LCB}_t(i) \leq e(i)$, we have

$$\begin{aligned}
\hat{\Delta}_t(i) &\geq \max_{j \neq i} \text{LCB}_t(j) - \text{UCB}_t(i) \\
&\geq \text{LCB}_t(i^*) - \text{UCB}_t(i) \\
&= \bar{e}_t(i^*) - \eta_t(i^*) - \bar{e}_t(i) - \eta_t(i) \\
&\geq e(i^*) - 2\eta_t(i^*) - e(i) - 2\eta_t(i) \\
&= \Delta(i) - 2\eta_t(i^*) - 2\eta_t(i).
\end{aligned} \tag{4.75}$$

Similarly, using Lemma 12, we have

$$\begin{aligned}
&\mathbb{P}\left(N_t(i) \leq \frac{\beta \log t}{2\Delta(i)^2} \text{ or } N_t(i^*) \leq \frac{\beta \log t}{2\Delta(i)^2}\right) \\
&\leq 2\left(\frac{1}{t}\right)^{\frac{\beta c_{min}^2}{8}} + \left(\frac{\log t}{t c_{min}^2 \Delta(i)^2}\right)^{\alpha-2}.
\end{aligned} \tag{4.76}$$

Now, assuming $N_t(i) > \beta \log t / 2\Delta(i)^2$ and $N_t(i^*) > \beta \log t / 2\Delta(i)^2$, we have

$$\begin{aligned}
\hat{\Delta}_t(i) &\geq \Delta(i) - 2\eta_t(i^*) - 2\eta_t(i) \\
&\geq \Delta(i) - 4\sqrt{\frac{2\Delta(i)^2 \alpha \log(tK^{1/\alpha})}{2\beta \log(t)}} \\
&\geq \Delta(i)\left(1 - 4\sqrt{\frac{\alpha+1}{c_{min}^2 \beta}}\right) \\
&\geq \Delta/2.
\end{aligned} \tag{4.77}$$

Therefore, combining (4.74),(4.75),(4.76) and (4.77), the statement of the theorem follows.

Hence proved. □

Lemma 14. *For all $i \in [K]$, let $X_t(i) = \Delta(i) - (\hat{\ell}_t(i) - \hat{\ell}_t(i^*))$ be the martingale difference sequence with respect to filtration $\mathcal{F}_1, \ldots, \mathcal{F}_1$ where $\mathcal{F}_t$ is the sigma field based on all the past actions, their rewards and their costs until round t. Then, for $t \geq t_{min}(i)$,we have*

$$\mathbb{P}\left(\max_{1 \leq n \leq t} X_n(i) \geq \frac{1.25t\Delta(i)^2}{c_{min}\beta \log(t)} \right) \leq \frac{1}{2}\left(\frac{\log t}{tc_{min}^2\Delta(i)^2} \right)^{\alpha-2}, \tag{4.78}$$

$$\mathbb{P}\left(\nu_t(i) \geq \frac{2t^2\Delta(i)^2}{c_{min}^3\beta \log(t)} \right) \leq \left(\frac{\log t}{tc_{min}^2\Delta(i)^2} \right)^{\alpha-2}, \tag{4.79}$$

where $\nu_t(i) = \sum_{n=1}^{t} \mathbb{E}[X_n(i)^2 | \mathcal{F}_{n-1}]$.

Proof. We bound the magnitude of $X_n(i)$. For all $i \in [K]$, we have

$$\begin{aligned}
X_n(i) &= \Delta(i) - (\hat{\ell}_n(i) - \hat{\ell}_n(i^*)) \\
&\leq \frac{1}{c_{min}} + \hat{\ell}_n(i^*) \\
&\leq \frac{1}{c_{min}} + \frac{1}{c_{min}\epsilon_n(i^*)} \\
&\leq \frac{1}{c_{min}}\left(1 + \max\left\{ 2K, 2\sqrt{\frac{nK}{\log(K)}}, \frac{n\hat{\Delta}_n(i^*)^2}{\beta \log(n)} \right\} \right) \\
&\leq \frac{1.25}{c_{min}} \max\left\{ 2K, 2\sqrt{\frac{nK}{\log(K)}}, \frac{n\hat{\Delta}_n(i^*)^2}{\beta \log(n)} \right\}.
\end{aligned} \tag{4.80}$$

Similar to the proof of Lemma 11, for $t \geq t_{min}$ and $n \leq tc_{min}^2\Delta(i)^2/\log(t)$, we have $\epsilon_n(i^*) \geq t\Delta(i)^2/\beta \log(t)$ and (see (4.66))

$$\frac{\beta \log(n)}{n\hat{\Delta}_n^2(i)} \geq \frac{\beta \log(t)}{t\Delta(i)^2}. \tag{4.81}$$

Additionally, for $t \geq t_{min}$,

$$0.5\sqrt{\frac{\log(K)}{tK}} \geq \frac{\beta \log(t)}{t\Delta(i)^2}, \tag{4.82}$$

and using Lemma 10, we have that $\hat{\Delta}_n(i) \leq \Delta(i)$ w.h.p as $n \to \infty$. Therefore, using for all $i \in [K]$ $\Delta(i^*) = 0 \leq \Delta(i)$, for $t_1 \leq tc_{min}^2 \Delta(i)^2 / \log(t)$ and $t \geq t_{min}(i)$,

$$\max_{1 \leq n \leq t_1} X_n(i) \leq \frac{1.25t\Delta(i)^2}{c_{min}\beta \log(t)}, \tag{4.83}$$

w.h.p at $t_1 \to \infty$. Now,

$$\begin{aligned}
&\mathbb{P}\left(\max_{1 \leq n \leq t} X_n(i) \geq \frac{1.25t\Delta(i)^2}{c_{min}\beta \log(t)} \right) \\
&\overset{(a)}{=} \mathbb{P}\left(\exists n \in \left[\frac{tc_{min}^2\Delta(i)^2}{\log(t)}, t \right] : X_n(i) \geq \frac{1.25t\Delta(i)^2}{c_{min}\beta \log(t)} \right) \\
&\overset{(b)}{\leq} \mathbb{P}\left(\exists n \in \left[\frac{tc_{min}^2\Delta(i)^2}{\log(t)}, t \right] : \hat{\Delta}_n(i) \geq \Delta(i) \right) \\
&\overset{(c)}{\leq} \frac{1}{2}\left(\frac{\log(t)}{tc_{min}^2\Delta(i)^2} \right)^{\alpha-2},
\end{aligned} \tag{4.84}$$

where (a) follows from (4.83), (b) follows from (4.80), and (c) follows from Lemma 10.

Now, we bound $\nu_t(i) = \sum_{n=1}^{t} \mathbb{E}[X_n(i)^2 | \mathcal{F}_{n-1}]$. For all $i \in [K]$, we have

$$\begin{aligned}
\mathbb{E}[X_n(i)^2|\mathcal{F}_{n-1}] &\leq \mathbb{E}[(\hat{\ell}_n(i^*) - \hat{\ell}_n(i))^2 | \mathcal{F}_{n-1}] \\
&\overset{(a)}{=} \mathbb{E}[\hat{\ell}_n(i^*)^2|\mathcal{F}_{n-1}] + \mathbb{E}[\hat{\ell}_n(i)^2|\mathcal{F}_{n-1}] \\
&= \tilde{p}_n(i)\left(\frac{\ell_n(i)}{\tilde{p}_n(i)} \right)^2 + \tilde{p}_n(i^*)\left(\frac{\ell_n(i^*)}{\tilde{p}_n(i^*)} \right)^2 \\
&\leq \frac{1}{c_{min}^2 \tilde{p}_n(i)} + \frac{1}{c_{min}^2 \tilde{p}_n(i^*)} \\
&\overset{(b)}{\leq} \frac{1}{c_{min}^3}\left(\max\left\{ 2K, 2\sqrt{\frac{nK}{\log(K)}}, \frac{n\hat{\Delta}_n(i)^2}{\beta \log(n)} \right\} \right) + \\
&\qquad \max\left\{ 2K, 2\sqrt{\frac{nK}{\log(K)}}, \frac{n\hat{\Delta}_n(i^*)^2}{\beta \log(n)} \right\} \Bigg),
\end{aligned} \tag{4.85}$$

where (a) follows from the fact that for all $i \in [K]$ and $n \leq t$, we have $\hat{\ell}_n(i^*)\hat{\ell}_n(i) = 0$, and (b)

follows from (4.80).

Similar to (4.84), we bound the $\nu_t(i)$ as follows

$$\begin{aligned}\mathbb{P}\left(\nu_t(i) \geq \frac{2t^2\Delta(a)^2}{c_{min}^3\beta\log(t)}\right) &\overset{(a)}{\leq} \mathbb{P}\left(\exists n \in \left[\frac{tc_{min}^2\Delta(i)^2}{\log(t)}, t\right] : \hat{\Delta}_n(i) \geq \Delta(i)\right) \\ &\quad + \mathbb{P}\left(\exists n \in \left[\frac{tc_{min}^2\Delta(i)^2}{\log(t)}, t\right] : \hat{\Delta}_n(i^*) \geq 0\right) \\ &\overset{(b)}{\leq} \left(\frac{\log(t)}{tc_{min}^2\Delta(i)^2}\right)^{\alpha-2}, \end{aligned} \tag{4.86}$$

where (a) can be implied in a similar way as (b) of (4.84), and (b) follows from Lemma 10. □

Lemma 15. *For all $t \geq t_{min}(i)$ and $\beta \geq 256/c_{min}^2$, we have*

$$\mathbb{P}\left(\tilde{\Delta}_t(i) \leq \frac{t\Delta(i)}{2}\right) \leq \left(\frac{\log(t)}{tc_{min}^2\Delta(i)^2}\right)^{\alpha-2} + \frac{1}{t}. \tag{4.87}$$

where $\tilde{\Delta}_t(i) = \sum_{n=1}^{t}(\hat{\ell}_n(i) - \hat{\ell}_n(i^*))$.

Proof. We have

$$\begin{aligned}\mathbb{P}\left(\tilde{\Delta}_t(i) \leq \frac{t\Delta(i)}{2}\right) &= \mathbb{P}\left(t\Delta(i) - \tilde{\Delta}_t(i) \geq \frac{t\Delta(i)}{2}\right) \\ &\leq \mathbb{P}(M_1(t)) + \mathbb{P}(M_2(t)) + \mathbb{P}(M_3(t)), \end{aligned} \tag{4.88}$$

where

$$M_1(t) = \left\{\max_{1\leq n\leq t} X_n(i) \geq \frac{1.25t\Delta(i)^2}{c_{min}\beta\log(t)}\right\}, \tag{4.89}$$

$$M_2(t) = \left\{\nu_t(i) \geq \frac{2t^2\Delta(a)^2}{c_{min}^2\beta\log(t)}\right\}, \tag{4.90}$$

and

$$M_3(t) = \left\{t\Delta(i) - \tilde{\Delta}_t(i) \geq \frac{t\Delta(i)}{2} \text{ and } M_1(t) \text{ and } M_2(t)\right\}. \tag{4.91}$$

The probability of the events $M_1(t)$ and $M_2(t)$ can be bound using Lemma 14, and using the fact that $c_{min} \leq 1$.

Let $w_1 = 2t^2\Delta(a)^2/c_{min}^2\beta\log(t)$, $w_2 = 1.25t\Delta(i)^2/c_{min}\beta\log(t)$, and $w_3 = 1/t$. For all $t \geq t_{min}(i)$ and $\beta \geq 256/c_{min}^2$, we have

$$\begin{aligned}\sqrt{2w_1\log\frac{1}{w_3}} + \frac{w_2}{3}\log\frac{1}{w_3} &= \sqrt{\frac{4t^2\Delta(i)^2\log t}{c_{min}^2\beta\log t}} + \frac{1.25t\Delta(i)^2\log t}{c_{min}\beta\log t} \\ &\leq t\Delta(i)\left(\frac{2t}{\sqrt{c_{min}^2\beta}} + \frac{1.25}{3c_{min}^2\beta}\right) \\ &\leq \frac{1}{2}t\Delta(i). \end{aligned} \tag{4.92}$$

Thus, using Bernstein's inequality for martingales and (4.92), we can bound the probability of $M_3(t)$ as follows

$$\mathbb{P}(M_3(t)) \leq \frac{1}{t}. \tag{4.93}$$

Thus, combining the bounds over the probabilities of the events $M_1(t)$, $M_2(t)$ and $M_3(t)$, the statement of the lemma follows.

Lemma 16. *For all $i \in [K]$, $\tau(E) \geq t_{min}(i)$, $T = \max\{\tau(E), T(i^*)\}$, $\alpha = 3$ and $\beta = 256/c_{min}^2$, we have*

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}[\epsilon_t(i)] \leq t_{min}(i) + \frac{4\beta(\log^2(T)+\log(T))}{\Delta(i)^2} + \frac{\log^2(T)+\log(T)}{c_{min}^2\Delta(i)^2} \\ + \frac{2}{K}(\log(T)+1) + \frac{2\pi^2}{3}. \end{aligned} \tag{4.94}$$

Proof: We have

$$\begin{aligned}
\sum_{t=1}^{T}\mathbb{E}[\epsilon_t(i)] &= \sum_{t=1}^{T}\mathbb{E}\left[\min\left\{\frac{1}{2K},\frac{1}{2}\sqrt{\frac{\log(t)}{tK}},\frac{\beta\log t}{t\hat{\Delta}_t(i)^2}\right\}\right] \\
&\leq \sum_{t=1}^{T}\mathbb{E}\left[\frac{\beta\log t}{t\hat{\Delta}_t(i)^2}\right] \\
&\overset{(a)}{\leq} t_{min}(i) + \frac{4\beta(\log^2(T)+\log(T))}{\Delta(i)^2} \\
&\quad + \sum_{t=t_{min}(i)}^{T}\left(\left(\frac{\log t}{tc_{min}^2\Delta(i)^2}\right)^{\alpha-2} + \frac{2}{Kt^{\alpha-1}} + 2\left(\frac{1}{t}\right)^{\frac{\beta c_{min}^2}{8}}\right) \\
&\leq t_{min}(i) + \frac{4\beta(\log^2(T)+\log(T))}{\Delta(i)^2} \\
&\quad + \frac{\log^2(T)+\log(T)}{c_{min}^2\Delta(i)^2} + \frac{2}{K}(\log(T)+1) + \frac{2\pi^2}{3},
\end{aligned} \tag{4.95}$$

where (a) follows from Lemma 13. □

Lemma 17. *For all $i \in [K]$, $\tau(E) \geq t_{min}(i)$, $\gamma \geq c_{min}^2\sqrt{K\log(K)/B(1+(e-2)/c_{min}^2)}$ and $\alpha \geq 3$, there exists a constant m_2 such that*

$$\sum_{t=1}^{T}\mathbb{E}[p_t(i)] \leq t_{min}(i) + m_2\frac{\log^2(T)}{c_{min}^2\Delta(i)^2}. \tag{4.96}$$

Proof. We have

$$\begin{aligned}
\sum_{t=1}^{T}\mathbb{E}[p_t(i)] &\leq \sum_{t=1}^{T}\mathbb{E}\Big[\exp(-\gamma_t\tilde{\Delta}_t(i))\Big] \\
&\overset{(a)}{\leq} t_{min}(i) + \sum_{t=t_{min}(i)}^{T}\Bigg(e^{-\sqrt{\frac{\log(K)}{tK}}\frac{t\Delta(i)}{4K}} + \frac{1}{t} + 2\left(\frac{1}{t}\right)^{\frac{\beta c_{min}^2}{8}} \\
&\quad + \left(\frac{\log(t)}{tc_{min}^2\Delta(i)^2}\right)^{\alpha-2} + \left(\frac{\log t}{tc_{min}^2\Delta(i)^2}\right)^{\alpha-2} + \frac{2}{Kt^{\alpha-1}}\Bigg) \\
&\overset{(b)}{\leq} t_{min}(i) + O\left(\frac{\log^2(T)}{c_{min}^2\Delta(i)^2}\right),
\end{aligned} \tag{4.97}$$

where (a) follows from the Lemma 13 , and (b) follows from bounds over the summation of sequences via integration. □

4.9.3 Proof of Theorem 4

Proof. For all $i \in [K]$,

$$p_t(i) = \frac{\exp(-\gamma_t \sum_{n=1}^{t-1} \hat{\ell}_n(i))}{\sum_{i\in[K]} \exp(-\gamma_t \sum_{n=1}^{t-1} \hat{\ell}_n(i))}, \tag{4.98}$$

and $\gamma_t = 0.5\sqrt{c_{min}^2 \log(K)/Kt}$. Therefore, using Lemma 7 of [195], we have

$$\sum_{t=1}^{T} \sum_{i\in[K]} p_t(i)\hat{\ell}_t(i) - \min_{j\in[K]} \sum_{t=1}^{T} \hat{\ell}_t(j) \leq \frac{1}{2} \sum_{t=1}^{T} \gamma_t \sum_{i\in[K]} p_t(i)(\hat{\ell}_t(i))^2 + \frac{\log(K)}{\gamma_T}, \tag{4.99}$$

where $T = \max\{T(i^*), \tau(E)\}$. We have

$$\begin{aligned} &\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}[\sum_{i\in[K]} p_t(i)\hat{\ell}_t(i)|\mathcal{F}_{t-1}]\right] - \mathbb{E}\left[\sum_{t=1}^{T} \hat{\ell}_t(i)\right] \\ &\leq \frac{\log(K)}{\gamma_T} + \mathbb{E}\left[\sum_{t=1}^{T} \frac{\gamma_t}{2} \mathbb{E}\left[\sum_{i\in[K]} p_t(i)\hat{\ell}_t^2(i)|\mathcal{F}_{t-1}\right]\right], \end{aligned} \tag{4.100}$$

where $\mathcal{F}_t$ is the sigma field with respect to the entire past until round t.

Now, let us bound the terms in (4.100). We have

$$\begin{aligned} \mathbb{E}[\sum_{i\in[K]} p_t(i)\hat{\ell}_t(i)|\mathcal{F}_{t-1}] &\geq \mathbb{E}\left[\sum_{i\in[K]} (\tilde{p}_t(i) - \epsilon_t(i))\hat{\ell}_t(i)|\mathcal{F}_{t-1}\right], \\ &\geq \frac{1}{c_{min}} - \mathbb{E}\left[\frac{r_t(i_t)}{c_t(i_t)}\middle|\mathcal{F}_{t-1}\right] - \sum_{i\in[K]} \frac{\epsilon_t(i)}{c_{min}}. \end{aligned} \tag{4.101}$$

Also, we have

$$\mathbb{E}\left[\sum_{t=1}^{T} \hat{\ell}_t(i^*)\right] = \sum_{t=1}^{T} \frac{1}{c_{min}} - \sum_{t=1}^{T} \frac{r_t(j)}{c_t(j)}. \tag{4.102}$$

Additionally, we have

$$\begin{aligned}\mathbb{E}\left[\sum_{i\in[K]} p_t(i)\hat{\ell}_t^2(i)|\mathcal{F}_{t-1}\right] &\leq \mathbb{E}\left[\sum_{i\in[K]} \frac{p_t}{c_{min}^2\tilde{p}_t^2}\middle|\mathcal{F}_{t-1}\right], \\ &\leq \sum_{i\in[K]} \frac{p_t}{c_{min}^2\tilde{p}_t}, \\ &\overset{(a)}{\leq} \frac{2K}{c_{min}^2}, \end{aligned} \tag{4.103}$$

where last inequality follows from the definition of $\tilde{p}_t(i)$, and the fact that for all $i \in [K]$ and t,$(1 - \sum_{j\neq i} \epsilon_t(j)) \geq 0.5$.

Using (4.39),(4.100), (4.101),(4.102) and (4.103), we have that the expected regret of the algorithm is at most

$$\begin{aligned}\frac{\log(K)}{\gamma_{n'}} + \frac{K}{c_{min}^2}\sum_{t=1}^{n'}\gamma_t + \sum_{t=1}^{\tau(E)}\sum_{i\in[K]}\frac{\epsilon_t(i)}{c_{min}} &\overset{(a)}{\leq} \frac{\log(K)}{\gamma_{n'}} + \frac{K}{c_{min}^2}\sum_{t=1}^{n'}\gamma_t + \sum_{t=1}^{n'}\sum_{i\in[K]}\frac{\gamma_t}{c_{min}^2} \\ &\overset{(b)}{\leq} 6\sqrt{\frac{BK\log(K)}{c_{min}^3}}, \end{aligned} \tag{4.104}$$

where $n' = \tau(E) + K/c_{min}$, (a) follows from the value of γ, and from the fact that $\epsilon_t(i) \leq 0.5c_{min}\sqrt{\log(K)/tK}$,and (b) follows from the concavity of $\sqrt{x}$. □

Chapter 5

Online learning with Feedback Graphs and Switching Costs

5.1 Introduction

Online learning has a wide variety of applications like classification, estimation, and ranking, and it has been investigated in different areas, including learning theory, control theory, operations research, and statistics. The problem can be viewed as a one player game against an adversary. The game runs for T rounds and at each round the player chooses an action from a given set of K actions. Every action $k \in [K]$ performed at round $t \in [T]$ carries a loss, that is a real number in the interval $[0, 1]$. The losses for all pairs (k, t) are assigned by the adversary before the game starts. The player also incurs a fixed and known Switching Cost (SC) every time he changes his action, that is an arbitrary real number $c > 0$. The expected regret is the expectation of the sum of losses associated to the actions performed by the player plus the SCs minus the losses incurred by the best fixed action in hindsight. The goal of the player is to minimize the expected regret over the duration of the game.

Based on the feedback information received after each action, online learning can be divided into three categories: Multi-Armed Bandit (MAB), Partial Information (PI), and Expert setting. In a MAB setting, at any given round the player only incurs the loss corresponding to the selected action, which implies the player only observes the loss of the selected action. In a PI setting, the player incurs the loss of the selected action $k \in [K]$, as well as observes the losses

that he would have incurred in that round by taking actions in a subset of $[K]\backslash\{k\}$. This feedback system can be viewed as a time-varying directed graph G_t with K nodes, where a directed edge $k \to j$ in G_t indicates that performing an action k at round t also reveals the loss that the player would have incurred if action j was taken at round t. In an Expert setting, taking an action reveals the losses that the player would have incurred by taking any of the other actions in that round. In this extremal case, the feedback system G_t corresponds to a time-invariant, undirected clique.

Online learning with PI has been used to design a variety of systems [70, 105, 252, 179]. In these applications, feedback captures the idea of side information provided to the player during the learning process. For example, the performance of an employee can provide information about the performance of other employees with similar skills, or the rating of a web page can provide information on ratings of web pages with similar content. In most of these applications, switching between the actions is not free. For example, a company incurs a cost associated to the learning phase while shifting an employee among different tasks, or switching the content of a web page frequently can exasperate users and force them to avoid visiting it. Similarly, re-configuring the production line in a factory is a costly process, and changing the stock allocation in an investment portfolio is subject to certain fees. Despite the many applications where both SC and PI are an integral part of the learning process, the study of online learning with SC has been limited only to the MAB and Expert settings. In the MAB setting, it has been shown that the expected regret of any player is at least $\tilde{\Omega}(c^{1/3}K^{1/3}T^{2/3})$ [49], and that Batch EXP3 is an order optimal algorithm [11]. In the Expert setting, it has been shown that the expected regret is at least $\tilde{\Omega}(\sqrt{\log(K)T})$ [41], and order optimal algorithms have been proposed in [71, 79]. The PI setup has been investigated only in the absence of SC, and for any fixed feedback system $G_t = G$ with independence number $\alpha(G) > 1$, it has been shown that the expected regret is at least $\tilde{\Omega}(\sqrt{\alpha(G)T})$ [145].

5.1.1 Contributions

We provide a lower bound on the expected regret for any sequence of feedback graphs $G_1, \dots G_T$ in the PI setting with SC. We show that for any sequence of feedback graphs $G_{1:T} = \{G_1, \dots G_T\}$ with independence sequence number $\beta(G_{1:T}) > 1$, the expected regret of any player is at least $\tilde{\Omega}(c^{1/3}\beta(G_{1:T})^{1/3}T^{2/3})$. We then show that for $G_{1:T}$ with $\alpha(G_t) > 1$ for all $t \leq T$, the expected regret of any player is at least $\tilde{\Omega}(c^{1/3}\sum_{G_j \in \mathcal{G}} \alpha(G_j)^{1/3}N(G_j)^{2/3})$, where $\mathcal{G}$ is the set of unique feedback graphs in the sequence $G_{1:T}$, and $N(G_j) = \sum_{t=1}^{T} \mathbf{1}(G_t = G_j)$ is the number of rounds for which the feedback graph G_j is seen in T rounds. These results introduce a new figure of merit $\beta(G_{1:T})$ in the PI setting, which can also be used to generalize the lower bound given in the PI setting without SC [145]. A consequence of these results is that the presence of SC changes the asymptotic regret by at least a factor $T^{1/6}$. Additionally, these results also recover the lower bound on the expected regret in the MAB setting [49].

We also show that in the PI setting for any algorithm that is order optimal without SC, there exists an assignment of losses from the adversary that forces the algorithm to make at least $\tilde{\Omega}(T)$ switches, thus increasing its asymptotic regret by at least a factor $T^{1/2}$. This shows that any algorithm that is order optimal in the PI setting without SC, is necessarily sub-optimal in the presence of SC, and motivates the development of new algorithms in the PI setting and in the presence of SC.

We propose two new algorithms for the PI setting with SC: Threshold-Based EXP3 and EXP3.SC. Threshold-Based EXP3 requires the knowledge of T in advance, whereas EXP3.SC does not. The performance of these algorithms is given for different scenarios in Table 5.1. The algorithms are order optimal in T and $\beta(G_{1:T})$ for two special cases of feedback information system: symmetric PI setting i.e. the feedback graph $G_t = G$ is fixed and un-directed, and MAB. In these two cases, $\beta(G_{1:T})$ equals $\alpha(G)$ and K respectively. The state-of-art algorithm EXP3 SET in PI setting without SC is known to be order optimal only for these cases as well [8]. Threshold Based EXP3 is order optimal in the SC c as well, while EXP3.SC

Table 5.1. Comparison of Threshold based EXP3 and EXP3.SC.

Scenarios	Threshold based EXP3	EXP3.SC	Lower Bound
For all t, $G_t = G$	$\tilde{O}(c^{1/3}(\text{mas}(G))^{1/3}T^{2/3})$	$\tilde{O}(c^{4/3}(\text{mas}(G))^{2/3}T^{2/3})$	$\tilde{\Omega}(c^{1/3}\alpha(G)^{1/3}T^{2/3})$
Symmetric PI	$\tilde{O}(c^{1/3}\alpha(G)^{1/3}T^{2/3})$	$\tilde{O}(c^{4/3}\alpha(G)^{2/3}T^{2/3})$	$\tilde{\Omega}(c^{1/3}\alpha(G)^{1/3}T^{2/3})$
MAB	$\tilde{O}(c^{1/3}K^{1/3}T^{2/3})$	$\tilde{O}(c^{4/3}K^{2/3}T^{2/3})$	$\tilde{\Omega}(c^{1/3}K^{1/3}T^{2/3})$
$G_{1:T}$	$\tilde{O}\left(c\sum_{t=1}^{t^*}\frac{\text{mas}(G_{(t)})}{\text{mas}(G_{(T)})}\right)$	$\tilde{O}\left(\sum_{t=1}^{n^*}\frac{\text{mas}(G_{(t)})}{\text{mas}(G_{(T)})}\right)$	$\tilde{\Omega}(c^{1/3}\beta(G_{1:T})^{1/3}T^{2/3})$
Equi-informational	$\tilde{O}(c^{1/3}\alpha(G_1)^{1/3}T^{2/3})$	$\tilde{O}(c^{4/3}\alpha(G_1)^{2/3}T^{2/3})$	$\tilde{\Omega}(c^{1/3}\beta(G_{1:T})^{1/3}T^{2/3})$

has an additional factor of c in its expected regret. In the time-varying case, for sequence $G_{1:T}$, the expected regret is dependent on the worst t^* and n^* instances of the ratio of mas(G_t) and mas$(G_{(T)})$, where $\{\text{mas}(G_{(1)}), \text{mas}(G_{(2)}), \ldots, \text{mas}(G_{(T)})\}$ are the sizes of the maximal acyclic subgraphs of $G_{1:T}$ arranged in non-increasing order, $t^* = \lceil T^{2/3}c^{-2/3}\text{mas}^{1/3}(G_{(T)})\rceil$ and $n^* = 0.5\text{mas}^{1/3}(G_{(T)})T^{2/3}c^{1/3}$. Finally, Table 5.1 also provides the performance in the equi-informational setting, namely when G_t is undirected and all the maximal acyclic subgraphs in $G_{1:T}$ have the same size.

Numerical comparison shows that Threshold Based EXP3 outperforms EXP3 SET in the presence of SCs. Threshold Based EXP3 also outperforms Batch EXP3, which is another order optimal algorithm for the MAB setting with SC [11].

5.1.2 Related Work

In the absence of SC, the lower bound on the expected regret is known for all three categories of online learning problems. In the MAB setting, the expected regret is at least $\tilde{\Omega}(\sqrt{KT})$ [17, 41, 180]. In the PI setting with fixed feedback graph G, the expected regret is at least $\tilde{\Omega}(\sqrt{\alpha(G)T})$ [145]. In the Expert setting, the expected regret is at least $\tilde{\Omega}(\sqrt{\log(K)T})$ [41]. All three cases present an asymptotic regret factor $T^{1/2}$. In contrast, in the presence of SC the expected regrets for MAB and Expert settings present different factors, namely $T^{2/3}$ and $T^{1/2}$ respectively. The expected regret is at least $\tilde{\Omega}(c^{1/3}K^{1/3}T^{2/3})$ in the MAB setting and

$\tilde{\Omega}(\sqrt{\log(K)T})$ in the Expert setting [49]. This work provides the lower bound on the expected regret $\tilde{\Omega}(c^{1/3}\beta(G_{1:T})^{1/3}T^{2/3})$ for the PI setting in the presence of SC. For the case without SC, this work establishes that the lower bound in PI setting is $\tilde{\Omega}(\sqrt{\beta(G_{1:T})T})$.

The PI setting was first considered in [9, 145], and many of its variations have been studied without SC [7, 9, 38, 178, 126, 118, 177, 231, 174]. In the adversarial setting we described, all of these algorithms are order optimal in the MAB and symmetric PI settings, but they also require the player to have knowledge of the graph G_t before performing an action. The algorithm EXP3 SET does not require such knowledge [8]. We show that all of these algorithms are sub-optimal in the PI setting with SC, and propose new algorithms that are order optimal in the MAB and symmetric PI settings.

In the expert setting with SC, there are two order optimal algorithms with expected regret $\tilde{O}(\sqrt{\log(K)T})$ [71, 79]. In the MAB setting with SC, Batch EXP3 is an order optimal algorithm with expected regret $\tilde{O}(c^{1/3}K^{1/3}T^{2/3})$ [11]. This algorithm has also been used to solve a variant of the MAB setting [62]. In the MAB setting, our algorithm has the same order of expected regret as Batch EXP3 but it numerically outperforms Batch EXP3.

There is a large literature on a continuous variation of the MAB setting, where the number of actions K depends on the number of rounds T. In this setting, the case without the SC was investigated in [19, 33, 116, 238]. Recently, the case including SC has also been studied in [120, 121]. In [120], the algorithm Slowly Moving Bandits (SMB) has been proposed and in [121], it has been extended to different settings. These algorithms incur an expected regret linear in T when applied in our discrete setting.

5.2 Problem Formulation

Before the game starts, the adversary fixes a loss sequence $\ell_1, \ldots, \ell_T \in [0,1]^K$, assigning a loss in $[0,1]$ to K actions for T rounds. At round t, the player performs an action $i_t \in [K]$, and incurs the loss $\ell_t(i_t)$ assigned by the adversary. If $i_t \neq i_{t-1}$, then the player also incurs a cost

$c > 0$ in addition to the loss $\ell_t(i_t)$.

In the PI setting, the feedback system can be viewed as a time-varying directed graph G_t with K nodes, where a directed edge $k \to j$ indicates that choosing action k at round t also reveals the loss that the player would have incurred if action j were taken at round t. Let $S_t(i) = \{j : i \to j \text{ is a directed edge in } G_t\}$. Following the action i_t, the player observes the losses he would have incurred in round t by performing actions in the subset $S_t(i_t) \subseteq [K]$. Since the player always observes its own loss, $i_t \in S_t(i_t)$. In a MAB setup, the feedback graph G_t has only self loops, i.e. for all $t \leq T$ and $i \in [K]$, $S_t(i) = \{i\}$. In an Expert setup, G_t is a undirected clique i.e. for all $t \leq T$ and $i \in [K]$, $S_t(i) = [K]$. The expected regret of a player's strategy δ is defined as

$$R^{\delta}(\ell_{1:T}, c) = \mathbf{E}\left[\sum_{t=1}^{T} \ell_t(i_t) + \sum_{t=2}^{T} c \cdot \mathbf{1}(i_{t-1} \neq i_t)\right] - \min_{k \in [K]} \sum_{t=1}^{T} \ell_t(k). \tag{5.1}$$

In words, the expected regret is the expectation of the sum of losses associated to the actions performed by the player plus the SCs minus the losses incurred by the best fixed action in the hindsight, and the objective of the player is to minimize the expected regret.

5.3 Lower Bound in PI setting with SC

We start by defining the independence sequence number for a sequence of graphs $G_{1:T}$.

Definition 1. *Given $G_{1:T}$, let $P(G_t)$ be the set of all the possible independent sets of the graph G_t. The independence sequence number $\beta(G_{1:T})$ is the largest cardinality among all intersections of the independent sets $s_1 \cap s_2 \cap \ldots \cap s_T$, where $s_t \in P(G_t)$. Namely,*

$$\beta(G_{1:T}) = \max_{s_1 \in P(G_1), \ldots s_T \in P(G_T)} |s_1 \cap s_2 \cap \ldots \cap s_T|. \tag{5.2}$$

Definition 2. *The independence sequence set $\mathcal{I}(G_{1:T})$ is the set $s_1 \cap s_2 \cap \ldots s_T$ attaining the maximum in* (5.2)*.*

Algorithm 8. Adversary's strategy

Input:$T > 0$, $G_{1:T}$ with $\beta(G_{1:T}) > 1$;
Set $\epsilon_1 = \epsilon_2 = c^{1/3}\beta(G_{1:T})^{1/3}T^{-1/3}/9\log_2(T)$ and $\sigma = 1/9\log_2(T)$.
Choose an arm $X \in \mathcal{I}(G_{1:T})$ uniformly at random
Draw T variables such that $\forall t \leq T$, $y_t \sim \mathcal{N}(0, \sigma^2)$.
For all $1 \leq t \leq T$ and $i \in [K]$, assign

$$\ell_t(i) = W_t + 0.5 - \epsilon_1 \mathbf{1}(X = i) + \epsilon_2 \mathbf{1}(i \notin \mathcal{I}(G_{1:T})),$$

$$\ell_t(i) = clip(\ell_t(i)),$$

where $clip(a) = \min\{\max\{a, 0\}, 1\}$, For all $t \leq T$ $W_t = W_{\rho(t)} + y_t$,, $W_0 = 0$, $\rho(t) = t - 2^{\delta(t)}$ and $\delta(t) = \max\{i \geq 0 : 2^i \text{ divides } t\}$.
Output: loss sequence $\ell_{1:T}$.

We use the notion of $\beta(G_{1:T})$ to provide a lower bound on the expected regret in the PI setting with SC.

Theorem 19. *For any $G_{1:T}$ with $\beta(G_{1:T}) > 1$, there exists a constant $b > 0$ and an adversary's strategy (Algorithm 8) such that for all $T \geq 27c\log_2^{3/2}(T)/\beta(G_{1:T})^2$, and for any player's strategy $\mathcal{A}$, the expected regret of $\mathcal{A}$ is*

$$R^{\mathcal{A}}(\ell_{1:T}, c) \geq b\, c^{1/3}\beta(G_{1:T})^{1/3}T^{2/3}/\log T. \tag{5.3}$$

The proof of Theorem 19 relies on Yao's minimax principle [237]. A randomized adversary strategy is constructed such that the expected regret of a player, whose action at any round is a deterministic function of his past observations, is at least $b\, c^{1/3}\beta(G_{1:T})^{1/3}T^{2/3}/\log T$. This adversary strategy is described in Algorithm 8, and is a generalization of the one proposed to establish similar bounds in the MAB setup [49]. The generalization is different than the one proposed for the PI setting without SC [145]. Since $G_{1:T}$ is known to the adversary, it computes the independence sequence set $\mathcal{I}(G_{1:T})$, and the cardinality of this set is $\beta(G_{1:T})$. For all $t \leq T$ and $i, j \in \mathcal{I}(G_{1:T})$, there exists no edge in the graph G_t between the actions i and j. Thus, the selection of any action in $\mathcal{I}(G_{1:T})$ provides no information about the losses of the other actions in $\mathcal{I}(G_{1:T})$. The adversary selects the optimal action uniformly at random from $\mathcal{I}(G_{1:T})$, and

assigns an expected loss of $1/2 - \epsilon_1$. The remaining actions in $\mathcal{I}(G_{1:T})$ are assigned an expected loss of $1/2$. On the other hand, since $i \in [K] \backslash \mathcal{I}(G_{1:T})$ provides information about the losses of actions in $\mathcal{I}(G_{1:T})$, action i is assigned an expected loss of $1/2 + \epsilon_2$ to compensate for this additional information. In practice, even a small bias ϵ_2 compensates for the extra information provided by an action in $[K] \backslash \mathcal{I}(G_{1:T})$.

In the PI setup without SC, for a fixed feedback graph $G_t = G$, the expected regret is at least $\tilde{\Omega}(\sqrt{\alpha(G)T})$ [8]. The lower bound is provided only for a fixed feedback system, and the lower bound for a general time-varying feedback system $G_{1:T}$ is left as an open question [8]. This also motivates the investigation of different graph theoretic measures to study the PI setting [8]. Theorem 19 provides a lower bound for a general time-varying feedback system $G_{1:T}$ for the PI setting in presence of SC. The lower bound is dependent on the independence sequence number $\beta(G_{1:T})$ of $G_{1:T}$. Thus, the ideas introduced in Theorem 19 can be extended to close this gap in the literature of PI setting without SC.

Lemma 18. *In the PI setting without SC, for any $G_{1:T}$ with $\beta(G_{1:T}) > 1$, there exists a constant $b > 0$ and an adversary's strategy such that for any player's strategy $\mathcal{A}$, the expected regret of $\mathcal{A}$ is at least $b\sqrt{\beta(G_{1:T})T}$.*

Using Theorem 19 and Lemma 18, it can be concluded that the presence of SC changes the asymptotic regret by at least a factor $T^{1/6}$. In the MAB setup, $\beta(G_{1:T}) = K$, and Theorem 19 recovers the bounds provided in [49].

We now focus on the assumption in Theorem 19, i.e. $\beta(G_{1:T}) > 1$. This is satisfied in many networks of practical interest. For example, networks modeled as p-random graphs where p is the probability of having edge between two nodes. The expected independence number of these graphs is $2\log(Kp)/p$ [46]. Since the probability of each node being in independent set is same, the expected value of $\beta(G_{1:T})$ is $K(2\log(Kp)/Kp)^T$, and Kp is the expected node degree which is usually a constant as p is inversely proportional to K. This is greater than one for large values of K, and small values of T.

Algorithm 8 depends on the independence sequence set $\mathcal{I}(G_{1:T})$ whose cardinality is non-increasing in T. In such cases, the adversary can split the sequence of feedback graphs $G_{1:T}$ into multiple sub-sequences i.e. say M sub-sequences such that $U_1 = \{G_t : t \in T_1 \subseteq [T]\} \ldots U_M = \{G_t : t \in T_M \subseteq [T]\}$, $[T] = \cup_{m\in[M]} T_m$, and for all $m_1, m_2 \in [M]$, $T_{m_1} \cap T_{m_2}$ is an empty set. For each sub-sequence U_m, compute the independence sequence set and assign losses independently of other sub-sequences according to Algorithm 8. This adversary's strategy, which we call Algorithm 1.1, gives the following bound on the expected regret.

Theorem 20. *For any split of $G_{1:T}$ into disjoint sub-sequences $U_1, \ldots U_M$ with $\beta(U_m) > 1$ and $N(U_m) \geq 27c \log_2^{3/2}(N(U_m))/\beta(U_m)^2 \; \forall m \in [M]$, there exists a constant $b > 0$ and an adversary's strategy (Algorithm 1.1) such that for any player's strategy $\mathcal{A}$, the expected regret of $\mathcal{A}$ is*

$$R^{\mathcal{A}}(\ell_{1:T}, c) \geq b\, c^{1/3} \sum_{m\in[M]} \beta(U_m)^{1/3} N(U_m)^{2/3} / \log T, \tag{5.4}$$

where $N(U_m) = \sum_{t=1}^{T} \mathbf{1}(G_t \in U_m)$ is the length of sub-sequence U_m.

With the insight provided by Theorem 20, the regret can be made large with an appropriate split of $G_{1:T}$ into sub-sequences. This can be formulated as a sub-modular optimization problem where the objective is:

$$\max_{\{U_1,\ldots,U_M\}} c^{1/3} \sum_{m\in[M]} \beta(U_m)^{1/3} N(U_m)^{2/3} / \log T \tag{5.5}$$

$$\text{subject to} \sum_{m\in[M]} N(U_m) = T,$$

$$\forall m_1, m_2 \in [M], U_{m_1} \cap U_{m_2} = \phi. \tag{5.6}$$

This can be solved using greedy algorithms developed in the context of sub-modular maximization.

Until now, we have been focusing on designing an adversary's strategy for maximizing the regret for a given sequence of feedback graphs $G_{1:T}$. Now, we briefly discuss the case when $G_{1:T}$ can also be chosen by the adversary. If the adversary is not constrained about the choice of feedback graphs, then the feedback graph that maximizes the expected regret would be a feedback graph with only self loops, as this reveals the least amount of information. If the adversary is constrained by the choice of independence number, i.e. for all $t \leq T$, $\alpha(G_t) \leq H$, then the optimal value of (5.5) is achieved for a sequence of fixed feedback graphs i.e. for all $t \leq T$, $\alpha(G_t) = H$, which implies $\beta(G_{1:T}) = H$.

We now discuss the trade-off between the loss incurred and the number of switches performed by the player.

Lemma 19. *If the expected regret computed ignoring the switching cost of any algorithm $\mathcal{A}$ is $\tilde{O}((\beta(G_{1:T})^{1/2}T)^{\beta})$, then there exists a loss sequence $\ell_{1:T}$ such that $\mathcal{A}$ makes at least $\tilde{\Omega}[(\beta(G_{1:T})^{1/2}T)^{2(1-\beta)}]$ switches.*

Along the same lines of Lemma 4, it can also be shown that if the expected number of switches of $\mathcal{A}$ is $\tilde{O}[(\beta(G_{1:T})^{1/2}T)^{2(1-\beta)}]$, then the expected regret without SC is at least $\tilde{\Omega}((\beta(G_{1:T})^{1/2}T)^{\beta})$. This provides the lower bound on the expected regret given the SC is constrained by a fixed budget. Using Lemma 19, if the expected regret without SC of $\mathcal{A}$ is $\tilde{O}(\sqrt{\beta(G_{1:T})T})$, then there exists a loss sequence that forces $\mathcal{A}$ to make at least $\tilde{\Omega}(T)$ switches. This implies the regret of $\mathcal{A}$ with the SC is linear in T. Thus, any algorithm that is order optimal without SC, is necessarily sub-optimal in the presence of SC, which motivates the design of new algorithms in our setting.

5.4 Algorithms in PI setting with SC

In this section, we introduce the two algorithms Threshold Based EXP3 and EXP3.SC for an uninformed setting where G_t is only revealed after the action i_t has been performed. This is common in a variety of applications. For instance, a user's selection of some product

Algorithm 9. Threshold based EXP3

Initialization: $\eta \in (0,1]$; For all $i \in [K]$, $w_{i,1} = 1$, $\hat{\ell}_0(i) = 0$ and $\ell'_0(i) = 0$; $r = 1$;
for $t = 1, \ldots, T$ **do**
  if E_1^t or E_2^t or E_3^t **then**
    if $t \neq 1$ **then**
      $\hat{\ell}_t(i) = \hat{\ell}_{t-1}(i) + \ell'_{t-1}(i)$
      $w_{i,t} = w_{i,t-1} \exp\left(-\eta \ell'_{t-1}(i)\right)$
    end if
    Update $p_{i,t} = w_{i,t} / \sum_{j \in [K]} w_{j,t}$.
    Choose $i_t = i$ with probability $p_{i,t}$.
    Set $r = 1$ and for all $i \in [K]$, set $\ell'_t(i) = 0$
  else
    For all $i \in [K]$, $p_{i,t} = p_{i,t-1}$, $\hat{\ell}_t(i) = \hat{\ell}_{t-1}(i)$
    and $w_{i,t} = w_{i,t-1}$; $i_t = i_{t-1}$;$r = r + 1$
  end if
  For all $i \in S_t(i_t)$, observe the pair $(\ell_t(i), i)$.
  For all $i \in [K]$, $\ell'_t(i) = \ell'_{t-1}(i) + \ell_t(i)\mathbf{1}(i \in S_t(i_t))/q_{i,t}$, where $q_{i,t} = \sum_{j:j \to i} p_{j,t}$
end for

allows to infer that the user might be interested in similar products. However, no action on the recommended products may mean that user might not be interested in the product, does not need it or did not check the products. Thus, the feedback is revealed only after the action has been performed.

In Threshold Based EXP3 (Algorithm 9), each action $i \in [K]$ is assigned a weight $w_{i,t}$ at round t. When the loss of action i is observed at round t, i.e. $i \in S_t(i_t)$, $w_{i,t}$ is computed by penalizing $w_{i,t-1}$ exponentially by the empirical loss $\ell_t(i)\mathbf{1}(i \in S_t(i_t))/q_{i,t}$. At round t, $p_t = \{p_{1,t}, \ldots, p_{K,t}\}$ is the sampling distribution where $p_{i,t} = w_{i,t} / \sum_{i \in [K]} w_{i,t}$. At round t, action i_t is selected with probability $p_{i,t}$ if the threshold event $E^t = E_1^t \cup E_2^t \cup E_3^t$ is true, where

$$E_1^t = \{t = 1\}, \tag{5.7}$$

$$E_2^t = \{r > \gamma_t, \text{ where } \gamma_t = T^{1/3} c^{2/3} / \text{mas}(G_{(T)})^{1/3}\}, \tag{5.8}$$

$$E_3^t = \{\forall i \in [K]\backslash\{i_t\}, \hat{\ell}_{t-1}(i) + \ell'_{t-1}(i) > \epsilon_t/\eta + 1/q_{i_t,t-1}, \text{ and there exists an } i \in [K]\backslash\{i_t\} \text{ such that } \hat{\ell}_{t-1}(i) + \ell'_{t-1}(i) - \ell'_{t-1}(i_t) \leq \epsilon_t/\eta + 1/q_{i_t,t-1}\}, \tag{5.9}$$

and $\epsilon_t \geq \log(tc^2/\text{mas}(G_{(T)}))/3$. The event E^t contains two threshold conditions, one on the variable r and the other on the empirical losses. The threshold event E^t is critical in balancing the trade-off between the number of switches and the loss incurred by the player. E_1^t corresponds to the first selection of action, and incurs no SC. In E_2^t, the variable r tracks the number of rounds (or time instances) since the event E^t occurred last time. If the choice of a new action has not been considered for past γ_t rounds, then E_2^t forces the player to choose an action according to the updated sampling distribution p_t at round t. The threshold condition in E_2^t ensures that the regret incurred due to the selection of a sub-optimal action does not grow continuously while trying to save on the SC between the actions. The event E_2^t is independent of the observed losses, and will occur at most $O(T^{2/3})$ times. Unlike event E_2^t, the event E_3^t is dependent on the losses $\hat{\ell}_t(i)$ and $\ell'_t(i)$, for all $i \in [K]$. Each loss $\hat{\ell}_t(i)$ tracks the total empirical loss of action i observed until round $\sigma(t) - 1$, namely

$$\hat{\ell}_t(i) = \sum_{k=1}^{\sigma(t)-1} \ell_k(i)\mathbf{1}(i \in S_k(i_k))/q_{i,k}, \tag{5.10}$$

where $\sigma(t) = \max\{k \leq t : E^k \text{ is true }\}$ is the latest round $k^* \leq t$ at which E^{k^*} is true. On the other hand, each loss $\ell'_t(i)$ represents the total empirical loss of action i observed between rounds $\sigma(t)$ and t, namely

$$\ell'_t(i) = \sum_{k=\sigma(t)}^{t} \ell_k(i)\mathbf{1}(i \in S_k(i_k))/q_{i,k}. \tag{5.11}$$

This loss tracks the total empirical loss observed after the selection of an action at time instance $\sigma(t)$. The event E_3^t balances exploration and exploitation while taking into account the SC. In E_3^t, the first condition ensures that the player has sufficient amount of information about the

losses of all other actions before exploitation is considered. Given sufficient exploration has been performed, the second condition triggers the exploitation. The selection of a new action is considered when the empirical loss $\ell'_t(i_t)$ incurred by the current action i_t, following its selection at $\sigma(t)$, becomes significant in comparison to the total empirical loss $\hat{\ell}_t(i) + \ell'_t(i)$ incurred by the other actions $i \in [K]\backslash\{i_t\}$. Since the total empirical loss of an action i increases with t, it is desirable that the threshold $\epsilon_t/\eta + 1/q_{i_t,t-1}$ increases with t as well. Since the increment in $\ell'_{t-1}(i_{t-1})$ is bounded above by $1/q_{i,t-1}$ at round t, for all $i \in [K]\backslash\{i_t\}$, E_3^t implies that

$$\hat{\ell}_{t-1}(i) + \ell'_{t-1}(i) - \ell'_{t-1}(i_{t-1}) \geq \epsilon_t/\eta. \tag{5.12}$$

Thus, E_3^t ensures that the player reconsiders the action selection if the loss incurred due to the current selection becomes significant in comparison to the total empirical loss of other actions. The event also ensures that the loss incurred due to the current selection is sufficiently smaller than the total empirical loss of other actions (see (5.12)). The event ensures that the sampling distribution p_t has changed significantly from the previous sampling distribution $p_{\sigma(t-1)}$ before selecting the action again. Thus, E_3^t balances exploration and exploitation based on the observed losses.

Batch EXP3, the order optimal algorithm in MAB with SC, is EXP3 performed in batches of $O(T^{1/3})$. A similar strategy to design an algorithm for the PI setting with SC will fail because unlike MAB setting, the feedback graph G_t can change at every round t, and this requires an update of empirical losses based on G_t at every round. In our algorithm, the computation of empirical loss is dependent on G_t via $q_{i,t}$. Additionally, Batch EXP3 does not utilize the information about the observed losses, which is captured in E_3^t. The following theorem presents the performance guarantees of our algorithm.

Theorem 21. *The following statements hold for Threshold Based EXP3:*

(i)*The expected regret without accounting for SC is*

$$\mathbb{E}\left[\sum_{t=1}^{T}\ell_t(i_t) - \min_{k\in[K]}\sum_{t=1}^{T}\ell_t(k)\right] \leq \frac{\log(K)}{\eta} + \frac{\eta}{2}\sum_{t=1}^{t^*}\frac{T^{2/3}c^{4/3}\text{mas}(G_{(t)})}{(1-1/e)\text{mas}^{2/3}(G_{(T)})}, \tag{5.13}$$

where $t^* = \lceil T^{2/3}c^{-2/3}\text{mas}^{1/3}(G_{(T)})\rceil$.

(ii) *The expected number of switches is*

$$\mathbb{E}\left[\sum_{t=2}^{T}\boldsymbol{1}(i_{t-1}\neq i_t)\right] \leq 2T^{2/3}c^{-2/3}\text{mas}^{1/3}(G_{(T)}). \tag{5.14}$$

(iii) *Letting* $\eta = \log(K)/T^{2/3}c^{1/3}\text{mas}^{1/3}(G_{(T)})$*, the expected regret (5.1) is at most*

$$R^{EXP3.T}(\ell_{1:T}, c) \leq 3T^{2/3}c^{1/3}\text{mas}^{1/3}(G_{(T)}) \tag{5.15}$$

$$+ \frac{ec\log(K)}{2(e-1)\text{mas}(G_{(T)})}\sum_{t=1}^{t^*}\text{mas}(G_{(t)}). \tag{5.16}$$

(iv) *In a symmetric PI setting i.e. for all* $t \leq T$ G_t *is un-directed and fixed, the expected regret (5.1) is at most*

$$R^{EXP3.T}(\ell_{1:T}, c) \leq 4T^{2/3}c^{1/3}\alpha^{1/3}(G_1)\log(K). \tag{5.17}$$

In the PI setting, $\text{mas}(G_t)$ captures the information provided by the feedback graph G_t. As $\text{mas}(G_t)$ increases, the information provided by G_t about the losses of actions decreases. The regret of the algorithm depends on the $O(T^{2/3})$ instances of $\text{mas}(G_{(t)})$ (see Theorem 21 (i)). This is because the algorithm makes a selection of a new action $O(T^{2/3})$ times in expectation (see Theorem 21 (ii)), and G_t is not available in advance to influence the selection of the action. Also, the ratio $\text{mas}(G_{(t)})/\text{mas}(G_{(T)})$ is bounded above by K and has no affect on order of T. The bounds of the algorithm on the expected regret are tight in two special cases. In the symmetric PI setting, the expected regret of Threshold Based EXP3 is $\tilde{O}(T^{2/3}c^{1/3}\alpha^{1/3}(G_1))$ (see Theorem 21 (iii)), hence, the algorithm is order optimal. In the MAB setting, the expected

Algorithm 10. EXP3.SC

Initialization: For all $i \in [K]$, $\hat{\ell}_1(i) = 0$; $t = 1$, $\epsilon_t = 0.5c^{1/3}\text{mas}^{1/3}(G_{(T)})/t^{1/3}$, $\eta_t = \log(K)/t^{2/3}c^{1/3}\text{mas}^{1/3}(G_{(T)})$

for $t = 1, \ldots, T$ **do**

For all $i \in [K]$, update:

$$p_t(i) = \frac{\exp(-\eta_t \hat{L}_{t-1}(i))}{\sum_{j\in[K]} \exp(-\eta_t \hat{L}_{t-1}(j))}$$

Choose $i_t = i_{t-1}$ with probability $1 - \epsilon_t$,

else, $i_t = i$ with probability $\epsilon_t p_{i,t}$.

For all $i \in S_t(i_t)$, observe the pair $(\ell_t(i), i)$.

For all $i \in [K]$, update $\hat{L}_t(i) = \sum_{n=1}^{t} \hat{\ell}_n(i)$,

where $\hat{\ell}_t(i) = \ell_t(i)\mathbf{1}(i \in S_t(i_t))/q_{i,t}$ and

$q_{i,t} = \sum_{j:j\to i} p_{j,t}$.

end for

regret of Threshold Based EXP3 is $\tilde{O}(T^{2/3}c^{1/3}K^{1/3})$, hence, the algorithm is order optimal. The state-of-art algorithm for the case without SCs is known to be order optimal only for these cases as well, and the key challenges for closing this gap are highlighted in the literature[8].

EXP3.SC (Algorithm 10) is another algorithm in PI setting with SC. The key differences between Threshold based EXP3 and EXP3.SC are highlighted here. Unlike Threshold based EXP3, EXP3.SC does not require the knowledge of the number of rounds T. Threshold based EXP3 favors the selection of action at regular intervals based on the event E^t. On contrary, EXP3.SC chooses a new action with probability ϵ_t which is decreasing in t. Thus, the algorithm favors exploration in the initial rounds, and favors exploitation as t increases. In Threshold based EXP3, the scaling exponent η is a constant dependent on T. On contrary, in EXP3.SC, the scaling exponent η_t is time-varying, and is decreasing in t. The following theorem provides the performance guarantees of EXP3.SC.

Theorem 22. *The expected regret (5.1) of EXP3.SC is at most*

$$R^{EXP3.SC}(\ell_{1:T}, c) \leq 1.5c^{4/3}\text{mas}^{1/3}(G_{(T)})T^{2/3} + \frac{2\log(K)}{\text{mas}^{2/3}(G_{(T)})}\sum_{j=1}^{n^*}\text{mas}(G_{(j)}), \tag{5.18}$$

where $n^* = 0.5\text{mas}^{1/3}(G_{(T)})T^{2/3}c^{1/3}$.

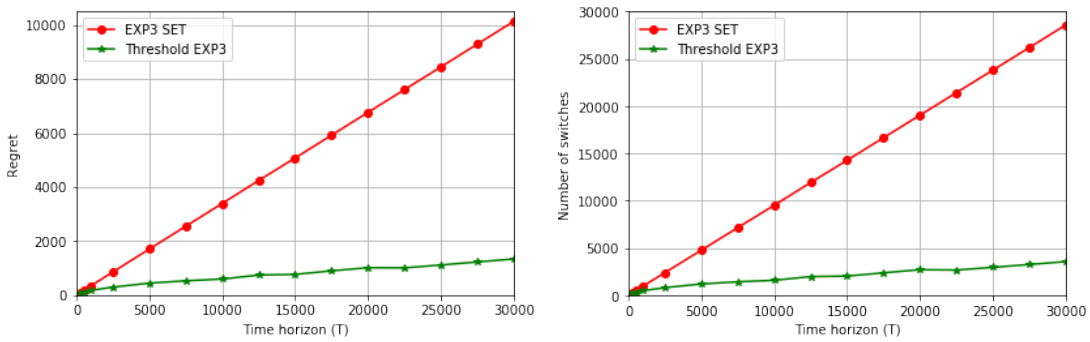
In symmetric PI and MAB settings, the expected regret of EXP3.SC is bounded above by $\tilde{O}(c^{4/3}\alpha^{2/3}(G_1)T^{2/3})$ and $\tilde{O}(c^{4/3}K^{2/3}T^{2/3})$ respectively. Hence, the algorithm is order optimal in T and $\beta(G_{1:T})$, and has an additional factor of c in the performance guarantees. In EXP3.SC, the dependency on T is removed at the expense of an additional factor of c in its performance.

In an alternative setting where the number of switches are constraint to be $O(T^{2(1-\beta)})$, it can be shown using Lemma 19 that the expected regret without SC is at least $\tilde{\Omega}((\beta(G_{1:T})^{1/2}T)^{\beta})$. The two algorithms in this setting are also simple variations of our two algorithms: Threshold based EXP3 and EXP3.SC. Threshold based EXP3 can be adapted by using threshold $\gamma_t = O(T^{2\beta-1})$, $\epsilon_t = O(\log(t)/2\beta - 1)$ and $\eta = O(T^{-\beta})$. EXP3.SC can be adapted by using $\epsilon_t = O(t^{-(2\beta-1)})$ and $\eta_t = O(t^{-\beta})$. These adapted algorithms would be order optimal in MAB and symmetric PI settings as well.

5.5 Performance Evaluation

In this section, we numerically compare the performance of Threshold based EXP3 with EXP3 SET and Batch EXP3 in PI and MAB setups with SC respectively. We do not compare the performance of our algorithm with the ones proposed in the Expert setting with SC because in MAB and PI setups, the player needs to balance the exploration-exploitation trade-off, while in the Expert setting the player is only concerned about the exploitation. Hence, there is a fundamental discontinuity in the design of algorithms as we move from the Expert to the PI setting. This gap is also evident from the discontinuity in the lower bounds in these settings, for the Expert setting the expected regret is at least $\tilde{\Omega}(\sqrt{\log(K)T})$, while for the PI setting the expected regret is at least $\tilde{\Omega}(\beta(G_{1:T})^{1/3}T^{2/3})$, for $\beta(G_{1:T}) > 1$ which excludes the clique feedback graph.

We evaluate these algorithms by simulations because in real data sets, the adversary's strategy is not necessarily unfavorable for the players. Hence, the trends in the performance can vary widely across different data sets. For this reason, in the literature only algorithms in



(a) For $\alpha(G) = 5$ **(b)** For $\alpha(G) = 5$

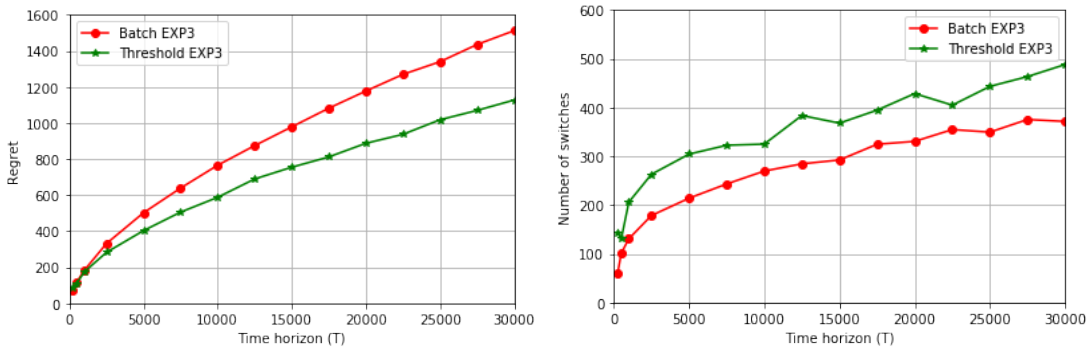
Figure 5.1. Performance evaluation of EXP3 SET and Threshold based EXP3 for K=25

stochastic setups rather than adversarial setups are typically evaluated on real data sets [105, 252]. In our simulations, the adversary uses the Algorithm 8, and $c = 0.35$.

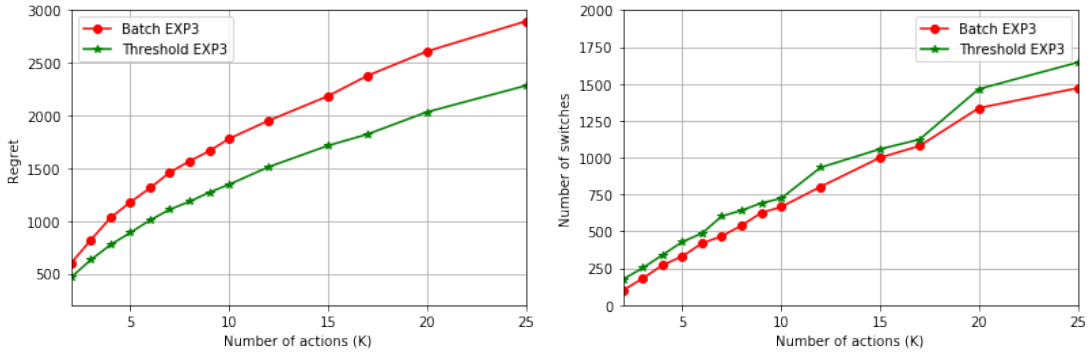
Figure 5.1 shows that the Threshold based EXP3 outperforms EXP3 SET in the presence of SC. Additionally, the expected regret and the number of switches of EXP3 SET grow linearly with T. These observations are in line with our theoretical results presented in Lemma 19. The results presented here are for $G_t = G$, $\alpha(G) = 5$ and $K = 25$. Similar trends were observed for different value of $\alpha(G)$ and K.

Figure 5.2 shows that Threshold based EXP3 outperforms Batch EXP3 in MAB setup with SC. The gap in the performance of these algorithm increases with T (see Figure 5.2(a)). Additionally, the number of switches performed by threshold based EXP3 is larger than the number of switches performed by Batch EXP3 (see Figure 5.2(b) and (d)). The former algorithm utilizes the information about the observed losses via E_3^t to balance the trade off between the regret and the number of switches. On contrary, Batch EXP3 does not utilize any information from the observed losses, and switches the action only after playing an action $\tilde{O}(T^{1/3})$ times. Note that MAB setup reveals the least information about the losses, and performance gap due to utilization of this information is significant (see Figure 5.2). This gap in performance grows as $\beta(G_{1:T})$ decreases.

In summary, Threshold Based EXP3 outperforms both EXP3 SET and Batch EXP3 in PI



(a) For $K = 5$ **(b)** For $K = 5$



(c) For $T = 20000$ **(d)** For $T = 20000$

Figure 5.2. Performance evaluation of Batch EXP3 and Threshold based EXP3 in MAB setting

and MAB settings with SC respectively. Threshold Based EXP3 fills a gap in the literature by providing a solution for the PI setting with SC, and improves upon the existing literature in the MAB setup.

5.6 Conclusion

This work focuses on online learning in the PI setting with SC in the presence of an adversary. The lower bound on the expected regret is presented in the PI setup in terms of independence sequence number. There is a need to design new algorithms in this setting because any algorithm that is order optimal without SC is necessarily sub-optimal in the presence of SC. Two algorithms, Threshold Based EXP3 and EXP3.SC, are proposed and their performance is evaluated in terms of expected regret. These algorithms are order optimal in T in two cases: symmetric PI and MAB setup. Numerical comparisons show that the Threshold Based EXP3

outperforms EXP3 SET and Batch EXP3 in PI setting with SC.

As future work, algorithms can be designed in a partially informed setting and a fully informed setting. In the partially informed setting, the feedback graph G_t at round t is revealed following the action at round $t-1$. Thus, the feedback graphs are revealed one at a time in advance at the beginning of each round. In the fully informed setting, the entire sequence of feedback graphs $G_{1:T}$ is revealed before the game starts. Since the adversary is aware of $G_{1:T}$, these settings are important to study from the player's end as well. Note that without SC, the algorithms in both the partially informed and fully informed settings can exploit the feedback graphs at every round in a greedy manner, and perform an action accordingly. Hence, the algorithm in partially informed setting is also optimal in a fully informed setting in the absence of SC. On the contrary, in the presence of SC, a greedy exploitation of the feedback structure is not possible at every round. Hence, in fully informed setting with SC, the player chooses an action based on $G_{1:T}$ such that the selected action balances the trade off between the regret and the SC. Thus, the partially informed and fully informed settings of PI are of particular interest in the presence of SC, and is an interesting area for further study.

5.7 Acknowledgement

Chapter 5, in full, is a reprint of the material as it appears in Anshuka Rangi and Massimo Franceschetti, "Online learning with feedback graphs and switching costs", *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, April 2019. The dissertation author was the primary investigator and author of this paper.

5.8 Appendix

5.8.1 Proof of Theorem 1

Proof. Without loss of generality, let the independent sequence set $\mathcal{I}(G_{1:T})$ formed of actions (or "arms") from 1 to $\beta(G_{1:T})$. Given the sequence of feedback graphs $G_{1:T}$, let T_i be the number of

times the action $i \in \mathcal{I}(G_{1:T}) = [\beta(G_{1:T})]$ is selected by the player in T rounds. Let T_Δ be the total number of times the actions are selected from the set $[K]\backslash\mathcal{I}(G_{1:T})$. Let $\mathbb{E}_i$ denote expectation conditioned on $X = i$, and $\mathbb{P}_i$ denote the probability conditioned on $X = i$. Additionally, we define $\mathbb{P}_0$ as the probability conditioned on event $\epsilon_1 = 0$. Therefore, under $\mathbb{P}_0$, all the actions in the independent sequence set, i.e. $i \in \mathcal{I}(G_{1:T})$, incur an expected regret of $1/2$, whereas, the expected regret of actions $i \in [K]\backslash\mathcal{I}(G_{1:T})$ is $1/2 + \epsilon_2$. Let $\mathbb{E}_0$ be the corresponding conditional expectation. For all $i \in [K]$ and $t \leq T$, $\ell_t(i)$ and $\ell_t^c(i)$ denote the unclipped and clipped loss of the action i respectively. Assuming the unclipped losses are observed by the player, then $\mathcal{F}$ is the sigma field generated by the unclipped losses, and $S_t(i_t)$ is the set of actions whose losses are observed at time t, following the selection of i_t, according to the feedback graph G_t. The observed sequence of unclipped losses will be referred as $\ell_{1:T}^o$. Additionally, $\mathcal{F}'$ is the sigma field generated by the clipped losses, for all $t \in [T]$, $\ell_t'(i)$ where $i \in S_t(i_t)$, and the observed sequence of clipped losses will be referred as $\ell_{1:T}'^o$. By definition, $\mathcal{F}' \subseteq \mathcal{F}$.

Let $i_1, \ldots, i_T$ be the sequence of actions selected by a player over the time horizon T. Then, the regret R^c of the player corresponding to clipped loses is

$$R^c = \sum_{t=1}^{T} \ell_t^c(i_t) + c \cdot M_s - \min_{i\in[K]} \sum_{t=1}^{T} \ell_t^c(i), \tag{5.19}$$

where M_s is the number of switches in the action selection sequence $i_1, \ldots, i_T$, and c is the cost of each switch in action. Now, we define the regret R which corresponds to the unclipped loss function in Algorithm 1 as following

$$R = \sum_{t=1}^{T} \ell_t(i_t) + c \cdot M_s - \min_{i\in[K]} \sum_{t=1}^{T} \ell_t(i). \tag{5.20}$$

Using [49, Lemma 4], we have

$$\mathbb{P}\left(\text{For all } t \in [T], \frac{1}{2} + W_t \in \left[\frac{1}{6}, \frac{5}{6}\right]\right) \geq \frac{5}{6}. \tag{5.21}$$

Thus, for all $T > \max\{\beta(G_{1:T}), 6\}$, we have $\epsilon_1 = \epsilon_2 < 1/6$. If $B = \{\text{For all } t \in [T] : 1/2 + W_t \in [1/6, 5/6]\}$ occurs and $\epsilon_1 = \epsilon_2 < 1/6$, then for all $i \in [K]$, $\ell_t^c(i) = \ell_t(i)$ which implies $R^c = R$ (see (5.19) and (5.20)). Now, if the event B does not occur, then the losses at any time t satisfy

$$\ell_t(i) - \ell_t^c(i) \leq (\epsilon_1 + \epsilon_2).$$

Therefore, we have

$$c \cdot M_s \leq R^c \leq R \leq c \cdot M_s + (\epsilon_1 + \epsilon_2)T.$$

Now, for $T > \max\{\beta(G_{1:T}), 6\}$, we have

$$\mathbb{E}[R] - \mathbb{E}[R^c] = (1 - \mathbb{P}(B))\mathbb{E}[R - R^c | B \text{ does not occur}] \leq \frac{(\epsilon_1 + \epsilon_2)T}{6}. \tag{5.22}$$

Thus, (5.22) lower bounds the actual regret R^c in terms of regret R. Now, we derive the lower bound on regret R corresponding to the unclipped loses. Using the definition of R, we have

$$\begin{aligned}
\mathbb{E}[R] &= \max_{i \in [K]} \mathbb{E}[\sum_{t=1}^{T} \ell_t(i_t) - \sum_{t=1}^{T} \ell_t(i)] + \mathbb{E}[M_s] \\
&= \frac{1}{\beta(G_{1:T})} \sum_{i=1}^{\beta(G_{1:T})} \mathbb{E}_i[\sum_{t=1}^{T} \ell_t(i_t) - \min_{i \in [K]} \sum_{t=1}^{T} \ell_t(i)] + \mathbb{E}[M_s] \\
&= \frac{1}{\beta(G_{1:T})} \sum_{i=1}^{\beta(G_{1:T})} \mathbb{E}_i\Bigg[\sum_{j \in \mathcal{I}(G_{1:T}) \setminus \{i\}} \frac{T_j}{2} + \left(\frac{1}{2} - \epsilon_1\right) T_i + \left(\frac{1}{2} + \epsilon_2\right) T_\Delta - \left(\frac{1}{2} - \epsilon_1\right) T \Bigg] + \mathbb{E}[M_s] \\
&= \frac{1}{\beta(G_{1:T})} \sum_{i=1}^{\beta(G_{1:T})} \mathbb{E}_i\Bigg[\sum_{j=1}^{\beta(G_{1:T})} \frac{1}{2} T_j + \left(\frac{1}{2} + \epsilon_2\right) T_\Delta - \epsilon_1 T_i - \left(\frac{1}{2} - \epsilon_1\right) T \Bigg] + \mathbb{E}[M_s] \\
&\overset{(a)}{=} \frac{1}{\beta(G_{1:T})} \sum_{i=1}^{\beta(G_{1:T})} \mathbb{E}_i\Big[\epsilon_2 T_\Delta + \epsilon_1 (T - T_i) \Big] + \mathbb{E}[M_s] \\
&\overset{(b)}{\geq} \epsilon_1 \left(T - \frac{1}{\beta(G_{1:T})} \sum_{i=1}^{\beta(G_{1:T})} \mathbb{E}_i\big[T_i\big] + \mathbb{E}\big[T_\Delta\big] \right) + \mathbb{E}[M_s],
\end{aligned} \tag{5.23}$$

where (a) follows from the fact that $\sum_{j=1}^{\beta(G_{1:T})} T_j + T_\Delta = T$, and (b) follows from $\epsilon_1 = \epsilon_2$.

Now, we upper bound the term $\mathbb{E}_i\big[T_i\big]$ in (5.23) to obtain the lower bound on the expected regret $\mathbb{E}[R]$. Since the player is deterministic, the event $\{i_t = i\}$ is $\mathcal{F}'$ measurable. Therefore, we have

$$\mathbb{P}_i(i_t = i) - \mathbb{P}_0(i_t = i) \leq d_{TV}^{\mathcal{F}'}(P_0, P_i) \overset{(a)}{\leq} d_{TV}^{\mathcal{F}}(P_0, P_i), \tag{5.24}$$

where $d_{TV}^{\mathcal{F}}(P_0, P_i) = \sup_{A\in\mathcal{F}}|\mathbb{P}_0(A) - \mathbb{P}_i(A)|$ is the total variational distance between the two probability measures, and (a) follows from $\mathcal{F}' \subseteq \mathcal{F}$. Summing the above equation over $t \in [T]$ and $i \in \mathcal{I}(G_{1:T})$ yields

$$\sum_{i=1}^{\beta(G_{1:T})} \big(\mathbb{E}_i[T_i] - \mathbb{E}_0[T_i]\big) \leq T \sum_{i=1}^{\beta(G_{1:T})} d_{TV}^{\mathcal{F}}(P_0, P_i). \tag{5.25}$$

Rearranging the above equation and using $\sum_{i=1}^{\beta(G_{1:T})} \mathbb{E}_0[T_i] = \mathbb{E}_0[\sum_{i=1}^{\beta(G_{1:T})} T_i] = T$, we have

$$\sum_{i=1}^{\beta(G_{1:T})} \mathbb{E}_i[T_i] \leq T \sum_{i=1}^{\beta(G_{1:T})} d_{TV}^{\mathcal{F}}(P_0, P_i) + T. \tag{5.26}$$

Combining the above equation with (5.23), we have

$$\begin{aligned}
&\mathbb{E}[R] \\
&\geq \epsilon_1 T - \frac{\epsilon_1 T}{\beta(G_{1:T})} \sum_{i=1}^{\beta(G_{1:T})} d_{TV}^{\mathcal{F}}(P_0, P_i) - \frac{\epsilon_1 T}{\beta(G_{1:T})} + \frac{\epsilon_1}{\beta(G_{1:T})} \sum_{i=1}^{\beta(G_{1:T})} \mathbb{E}_i\big[T_\Delta\big] + \mathbb{E}[M_s] \\
&\overset{(a)}{\geq} \frac{\epsilon_1 T}{2} - \frac{\epsilon_1 T}{\beta(G_{1:T})} \sum_{i=1}^{\beta(G_{1:T})} d_{TV}^{\mathcal{F}}(P_0, P_i) + \epsilon_1 \mathbb{E}\big[T_\Delta\big] + \mathbb{E}[M_s],
\end{aligned} \tag{5.27}$$

where (a) uses the fact that $\beta(G_{1:T}) > 1$. Next, we upper bound the second term in the right hand side of (5.27). Using Pinsker's inequality, we have

$$d_{TV}^{\mathcal{F}}(P_0, P_i) \leq \sqrt{\frac{1}{2} D_{KL}(\mathbb{P}_0(\ell_{1:T}^o) || \mathbb{P}_i(\ell_{1:T}^o))}, \tag{5.28}$$

where $\ell_{1:T}^o$ are the losses observed by the player over the time horizon T. Using the chain rule of

relative entropy to decompose $D_{KL}(\mathbb{P}_0(\ell^o_{1:T})||\mathbb{P}_i(\ell^o_{1:T}))$, we get

$$\begin{aligned} D_{KL}(\mathbb{P}_0(\ell^o_{1:T})||\mathbb{P}_i(\ell^o_{1:T})) &= \sum_{t=1}^{T} D_{KL}(\mathbb{P}_0(\ell^o_t|\ell^o_{1:t-1})||\mathbb{P}_i(\ell^o_t|\ell^o_{1:t-1})) \\ &= \sum_{t=1}^{T} D_{KL}(\mathbb{P}_0(\ell^o_t|\ell^o_{\rho^*(t)})||\mathbb{P}_i(\ell^o_t|\ell^o_{\rho^*(t)})), \end{aligned} \tag{5.29}$$

where $\rho^*(t)$ is the set of time instances $0 \leq k \leq t$ encountered when operation $\rho(.)$ in Algorithm 1 is applied recursively to t. Now, we deal with each term $D_{KL}(\mathbb{P}_0(\ell^o_t|\ell^o_{\rho^*(t)})||\mathbb{P}_i(\ell^o_t|\ell^o_{\rho^*(t)}))$ in the summation individually. For $i \in \mathcal{I}(G_{1:T})$, we separate this computation into four cases: i_t is such that loss of action i is observed at both time instances t and $\rho(t)$ i.e. $i \in S_t(i_t)$ and $i \in S_t(i_{\rho(t)})$; i_t is such that loss of action i is observed at time instance t but not at time instance $\rho(t)$ i.e. $i \in S_t(i_t)$ and $i \notin S_t(i_{\rho(t)})$; i_t is such that loss of action i is not observed at time instance t but is observed at time instance $\rho(t)$ i.e. $i \notin S_t(i_t)$ and $i \in S_t(i_{\rho(t)})$; i_t is such that loss of action i is not observed at both time instances t and $\rho(t)$ i.e. $i \notin S_t(i_t)$ and $i \notin S_t(i_{\rho(t)})$.

Case 1: Since the loss of action i is observed from the independent sequence set $\mathcal{I}(G_{1:T})$ at both the time instances, the loss distribution for the action i is $\ell^o_t(i)|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(i), \sigma^2)$ for both $\mathbb{P}_0$ and $\mathbb{P}_i$. For all $j \in \mathcal{I}(G_{1:T})\backslash\{i\}$, the loss distribution is $\ell^o_t(j)|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(i)+\epsilon_1, \sigma^2)$ under both $\mathbb{P}_0$ and $\mathbb{P}_i$. For all $j \in [K]\backslash\mathcal{I}(G_{1:T})$, the loss distribution is $\ell^o_t(j)|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(i)+\epsilon_1+\epsilon_2, \sigma^2)$ under both $\mathbb{P}_0$ and $\mathbb{P}_i$.

Case 2: The loss of action i is observed from the independent sequence set $\mathcal{I}(G_{1:T})$ at time instance t but not at $\rho(t)$. Let $k' \in \mathcal{I}(G_{1:T})\backslash\{i\}$ be the action from the independent sequence set which was observed at time instance $\rho(t)$. Then, the loss distribution for the action i is $\ell^o_t(i)|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(k'), \sigma^2)$ under $\mathbb{P}_0$, and $\ell^o_t(i)|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(k') - \epsilon_1, \sigma^2)$ under $\mathbb{P}_i$. For all $j \in \mathcal{I}(G_{1:T})\backslash\{i\}$, the loss distribution is $\ell^o_t(j)|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(k'), \sigma^2)$ under both $\mathbb{P}_0$ and $\mathbb{P}_i$. For all $j \in [K]\backslash\mathcal{I}(G_{1:T})$, the loss distribution is $\ell^o_t(j)|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(k') + \epsilon_2, \sigma^2)$ under both $\mathbb{P}_0$ and $\mathbb{P}_i$. If no such k' exists, then there exists a $k \in [K]\backslash\mathcal{I}(G_{1:T})$ that was observed at $\rho(t)$. Then, the loss distribution for the action i is $\ell^o_t(i)|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(k) - \epsilon_2, \sigma^2)$ under $\mathbb{P}_0$, and

$\ell_t^o(i)|\ell_{\rho^*(t)}^o \sim \mathcal{N}(\ell_{\rho(t)}(k) - \epsilon_1 - \epsilon_2, \sigma^2)$ under $\mathbb{P}_i$. For all $j \in \mathcal{I}(G_{1:T})\backslash\{i\}$, the loss distribution is $\ell_t^o(j)|\ell_{\rho^*(t)}^o \sim \mathcal{N}(\ell_{\rho(t)}(k) - \epsilon_2, \sigma^2)$ under both $\mathbb{P}_0$ and $\mathbb{P}_i$. For all $j \in [K]\backslash\mathcal{I}(G_{1:T})$, the loss distribution is $\ell_t^o(j)|\ell_{\rho^*(t)}^o \sim \mathcal{N}(\ell_{\rho(t)}(k), \sigma^2)$ under both $\mathbb{P}_0$ and $\mathbb{P}_i$.

Case 3: The action i is observed from the independent sequence set $\mathcal{I}(G_{1:T})$ at time instance $\rho(t)$ but not at t. Then, $\ell_t^o(i)|\ell_{\rho^*(t)}^o \sim \mathcal{N}(\ell_{\rho(t)}(i), \sigma^2)$ under both $\mathbb{P}_0$ and $\mathbb{P}_i$. Let $k' \in \mathcal{I}(G_{1:T})\backslash\{i\}$ be the action from the independent sequence set which was observed at time instance t. Then, the loss distribution for the arm k' is $\ell_t^o(k')|\ell_{\rho^*(t)}^o \sim \mathcal{N}(\ell_{\rho(t)}(i), \sigma^2)$ under $\mathbb{P}_0$, and $\ell_t^o(k')|\ell_{\rho^*(t)}^o \sim \mathcal{N}(\ell_{\rho(t)}(i) + \epsilon_1, \sigma^2)$ under $\mathbb{P}_i$. For all $j \in [K]\backslash\mathcal{I}(G_{1:T})$, the loss distribution is $\ell_t^o(j)|\ell_{\rho^*(t)}^o \sim \mathcal{N}(\ell_{\rho(t)}(i) + \epsilon_1 + \epsilon_2, \sigma^2)$ under both $\mathbb{P}_0$ and $\mathbb{P}_i$.

Case 4: The loss of i is not observed at $\rho(t)$ and t. Then the distribution of all action $[K]\backslash\{i\}$ is same under both $\mathbb{P}_0$ and $\mathbb{P}_i$.

Therefore, we have

$$\begin{aligned}
&D_{KL}(\mathbb{P}_0(\ell_t^o|\ell_{\rho^*(t)}^o)||\mathbb{P}_i(\ell_t^o|\ell_{\rho^*(t)}^o)) \\
&\leq \mathbb{P}_0(i \in S_t(i_t), i \notin S_{\rho(t)}(i_{\rho(t)})) \cdot D_{KL}(\mathcal{N}(0,\sigma^2)||\mathcal{N}(-\epsilon_1,\sigma^2)) \\
&\quad + \mathbb{P}_0(i \notin S_t(i_t), i \in S_{\rho(t)}(i_{\rho(t)})) \cdot D_{KL}(\mathcal{N}(0,\sigma^2)||\mathcal{N}(\epsilon_1,\sigma^2)) \\
&= \frac{\epsilon_1^2}{2\sigma^2}\mathbb{P}_0(B_t), \qquad (5.30)
\end{aligned}$$

where $B_t = \{i \in S_t(i_t), i \notin S_{\rho(t)}(i_{\rho(t)}) \cup i \notin S_t(i_t), i \in S_{\rho(t)}(i_{\rho(t)})\}$. The event B_t implies that at least one of the following events are true:

(i) The player has switched at least once between the feedback systems $S_t(k_1)$ and $S_{\rho(t)}(k_2)$ such that $i \in S_t(k_1)$ but $i \notin S_{\rho(t)}(k_2)$ or vice-versa.

(ii) The player did not change the selection of action, however, the feedback system has changed between $\rho(t)$ and t such that i has become observable or vice versa. This can occur only if the fixed action belongs to $[K] \setminus \mathcal{I}(G_{1:T})$.

Let N_i be the number of times a player switches from the feedback system which includes i to the feedback system which does not include i and vice-versa. Then, using (5.29) and (5.30), we

have

$$D_{KL}(\mathbb{P}_0(\ell^o_{1:T})||\mathbb{P}_i(\ell^o_{1:T})) \leq \frac{\epsilon_1^2\omega(\rho)}{\sigma^2}\mathbb{E}_0[N_i + T_\Delta], \tag{5.31}$$

where $\omega(\rho)$ is the width of process $\rho(.)$ (see Definition 2 in [49]) and is bounded above by $2\log_2(T)$. Combining (5.28) and (5.31), we have

$$\sup_{A\in\mathcal{F}}(\mathbb{P}_0(A) - \mathbb{P}_i(A)) \leq \frac{\epsilon_1}{\sigma}\sqrt{\log_2(T)\mathbb{E}_0[N_i + T_\Delta]}. \tag{5.32}$$

If $M_s \geq \epsilon_1 T$, then $\mathbb{E}[R'] > \epsilon_1 T$. Thus, the claimed lower bound follows. Now, let us assume $M_s \leq \epsilon_1 T$. For all $i \in \mathcal{I}(G_{1:T})$, we have

$$\begin{aligned}\mathbb{E}_0[M_s] - \mathbb{E}_i[M_s] &= \sum_{m=1}^{\lfloor\epsilon_1 T\rfloor} \mathbb{P}_0(M_s \geq m) - \mathbb{P}_i(M_s \geq m)) \\ &\leq \epsilon_1 T \cdot d^{\mathcal{F}}_{TV}(\mathbb{P}_0, \mathbb{P}_i).\end{aligned} \tag{5.33}$$

Using the above equation, we have

$$\begin{aligned}\mathbb{E}_0[M_s] - \mathbb{E}[M_s] &= \frac{1}{\beta(G_{1:T})}\sum_{i=1}^{\beta(G_{1:T})}(\mathbb{E}_0[M_s] - \mathbb{E}_i[M_s]) \\ &\leq \frac{\epsilon_1 T}{\beta(G_{1:T})}\sum_{i=1}^{\beta(G_{1:T})} d^{\mathcal{F}}_{TV}(\mathbb{P}_0, \mathbb{P}_i).\end{aligned} \tag{5.34}$$

Now, combining (5.22), (5.27), (5.32)and (5.34), we obtain

$$\begin{aligned}
\mathbb{E}[R'] &\geq \frac{\epsilon_1 T}{6} - \frac{\epsilon_1 T}{\beta(G_{1:T})} \sum_{i=1}^{\beta(G_{1:T})} \frac{\epsilon_1}{\sigma} \sqrt{\log_2(T)\mathbb{E}_0[N_i + T_\Delta]} + \epsilon_1 \mathbb{E}\big[T_\Delta\big] + c \cdot \mathbb{E}_0[M_s] \\
&\overset{(a)}{\geq} \frac{\epsilon_1 T}{6} - \frac{\epsilon_1^2 T}{\sigma\sqrt{\beta(G_{1:T})}} \sqrt{2\log_2(T)\mathbb{E}_0[M_s + T_\Delta]} + \epsilon_1 \mathbb{E}\big[T_\Delta\big] + c \cdot \mathbb{E}_0[M_s] \\
&\overset{(b)}{\geq} \frac{\epsilon_1 T}{6} - \frac{\epsilon_1^4 T^2 \log_2(T)}{c \cdot \sigma^2 \beta(G_{1:T})} + \epsilon_1 \mathbb{E}\big[T_\Delta\big] + c \cdot \left(\frac{\epsilon_1^4 T^2 \log_2(T)}{2c^2 \cdot \sigma^2 \beta(G_{1:T})} - \mathbb{E}_0\big[T_\Delta\big] \right) \\
&\geq \frac{c^{1/3}\beta(G_{1:T})^{1/3} T^{2/3}}{54 \log_2(T)} - \frac{4c^{1/3}\beta(G_{1:T})^{1/3} T^{2/3}}{162 \log_2(T)} + (\epsilon_1 - c)\mathbb{E}_0\big[T_\Delta\big] \\
&\overset{(c)}{\geq} \frac{c^{1/3}\beta(G_{1:T})^{1/3} T^{2/3}}{81 \log_2(T)},
\end{aligned} \tag{5.35}$$

where (a) follows from the concavity of $\sqrt{x}$ and $\sum_i^{\beta(G_{1:T})} N_i \leq 2M_s$, (b) follows from the fact that the right hand side is minimized for $\sqrt{\mathbb{E}_0[M_s + T_\Delta]} = \epsilon^2 T\sqrt{\log_2(T)}/2c\sigma\sqrt{\beta(G_{1:T})}$, and (c) follows from the assumption

$$T \geq 27c \log_2^{3/2}(T)/\beta(G_{1:T})^2, \tag{5.36}$$

which implies $\epsilon_1 \geq c$. The claim of the theorem now follows. □

5.8.2 Proof of Lemma 2

We have that $\beta(G_{1:T})$ actions are non adjacent in the entire sequence of feedback graphs $G_{1:T}$. Let $1, 2, \ldots \beta(G_{1:T})$ belong to the $\mathcal{I}(G_{1:T})$. Then, the adversary selects an action uniformly at random from the set $\mathcal{I}(G_{1:T})$ say j, and assigns the loss sequence to action j using independent Bernoulli random variable with parameter $0.5 - \epsilon$, where $\epsilon = \sqrt{\beta(G_{1:T})/T)}$. For all $i \in \mathcal{I}(G_{1:T})/\{j\}$, losses are assigned using independent Bernoulli random variable with parameter 0.5. For all $i \notin \mathcal{I}(G_{1:T})$, the losses are assigned using independent Bernoulli random variable with parameter 1. The proof of the lemma follows along the same lines as in Theorem 5 in ([8]).

5.8.3 Proof of Theorem 3

Proof of this theorem uses the results from Theorem 1. Since the loss sequence is assigned independently to each sub-sequence U_m where $m \in [M]$. Using Theorem 1, there exists a constant b_m such that

$$\mathbb{E}\left[\sum_{t=1}^{T}(\ell_t(i_t)\mathbf{1}(G_t \in U_m) + cW_m\right] - \min_{i \in U_m}\sum_{t=1}^{T}(\ell_t(i)\mathbf{1}(G_t \in U_m)$$
$$\geq b_m c^{1/3}\beta(U_m)^{1/3}N(U_m)^{2/3}/\log(T), \tag{5.37}$$

where W_m is number of switches performed within the sequence U_m. Since

$$\sum_{m\in[M]} W_m \leq \sum_{t=1}^{T}\mathbf{1}(i_t \neq i_{t-1}),$$

there exist a constant b such that the expected regret of any algorithm $\mathcal{A}$ is at least

$$b\, c^{1/3}\sum_{m\in[M]}\beta(U_m)^{1/3}N(U_m)^{2/3}/\log T.$$

5.8.4 Proof of Lemma 4

Proof. The proof follows from contradiction and is along the same lines as the proof of Theorem 4 in [49]. Let $\mathcal{A}$ performs at most $\tilde{O}((\beta(G_{1:T})^{1/2}T)^{\alpha})$ switches for any sequence of loss function over T rounds with $\beta + \alpha/2 < 1$. Then, there exists a real number γ such that $\beta < \gamma < 1 - \alpha/2$. Then, assign $c = (\beta(G_{1:T})^{1/2}T)^{3\gamma-2}$. Thus, the expected regret, including the switching cost, of the algorithm is

$$\tilde{O}((\beta(G_{1:T})^{1/2}T)^{\beta} + (\beta(G_{1:T})^{1/2}T)^{3\gamma-2}(\beta(G_{1:T})T)^{\alpha}) = \tilde{o}(\beta(G_{1:T})^{1/2}T)^{\gamma}, \tag{5.38}$$

over a sequence of losses assigned by the adversary because $\beta < \gamma$ and $\alpha < 2 - 2\gamma$. However, according to Theorem 1, the expected regret is at least

$$\tilde{\Omega}(\beta(G_{1:T})^{1/3}(\beta(G_{1:T})^{1/2}T)^{(3\gamma-2)/3}T^{2/3}) = \tilde{\Omega}((\beta(G_{1:T})T)^{\gamma}). \tag{5.39}$$

Hence, by contradiction, the proof of the lemma follows.

□

5.8.5 Proof of Theorem 5

Proof. Let $t_1, t_2 \ldots, t_{\sigma(T)}$ be the sequence of time instances at which the event E^t occurs during the duration T of the game. We define $\{r_j = t_{j+1} - t_j\}_{1 \le j \le T}$ as the sequence of inter-event times between the events E^t. Let $\text{mas}(G_{(1)}), \ldots, \text{mas}(G_{(T)})$ denote the sequence in the decreasing order of size of maximal acyclic graphs, i.e. $\text{mas}(G_{(1)})$ (or $\text{mas}(G_{(T)})$) is the maximum (or minimum) size of maximal acyclic graph observed in sequence $G_{1:T} = \{G_1, \ldots G_T\}$. Using the definition of E^t, note that r_j is a random variable bounded by $T^{1/3}c^{2/3}/\text{mas}(G_{(T)})^{1/3}$. For all $1 \le j \le \sigma(T)$, the ratio of total weights of actions at round t_j and t_{j+1} is

$$\begin{aligned}
\frac{W_{t_{j+1}}}{W_{t_j}} &= \sum_{i \in [K]} \frac{w_{i,t_{j+1}}}{W_{t_j}} \\
&= \sum_{i \in [K]} \frac{w_{i,t_j} \exp(-\eta \ell'_{t_j + r_j - 1}(i))}{W_{t_j}} \\
&= \sum_{i \in [K]} p_{i,t_j} \exp(-\eta \ell'_{t_j + r_j - 1}(i)) \\
&\overset{(a)}{\le} \sum_{i \in [K]} p_{i,t_j} \left(1 - \eta \ell'_{t_j + r_j - 1}(i) + \frac{1}{2}\eta^2 \ell'^2_{t_j + r_j - 1}(i)\right) \\
&= 1 - \eta \sum_{i \in [K]} p_{i,t_j} \cdot \ell'_{t_j + r_j - 1}(i) + \frac{\eta^2}{2} \sum_{i \in [K]} p_{i,t_j} \cdot \ell'^2_{t_j + r_j - 1}(i),
\end{aligned} \tag{5.40}$$

where (a) follows from the fact that, for all $x \geq 0$, $e^{-x} \leq 1 - x - x^2/2$. Now, taking logs on both sides of (5.40), summing over $t_1, t_2, \ldots t_{\sigma(T)}$, and using $\log(1+x) \leq x$ for all $x > -1$, we get

$$\log \frac{W_{t_{\sigma(T)+1}}}{W_1} \leq -\eta \sum_{j=1}^{\sigma(T)} \sum_{i \in [K]} p_{i,t_j} \cdot \ell'_{t_j+r_j-1}(i) + \frac{\eta^2}{2} \sum_{j=1}^{\sigma(T)} \sum_{i \in [K]} p_{i,t_j} \cdot \ell'^2_{t_j+r_j-1}(i). \tag{5.41}$$

For all actions $k' \in [K]$, we also have

$$\log \frac{W_{t_{\sigma(T)+1}}}{W_1} \geq \log \frac{w_{k',t_{\sigma(T)+1}}}{W_1} \geq -\eta \sum_{j=1}^{\sigma(T)} \ell'_{t_j+r_j-1}(k') - \log(K). \tag{5.42}$$

Combining (5.41) and (5.42), for all $k' \in [K]$, we obtain

$$\sum_{j=1}^{\sigma(T)} \sum_{i \in [K]} p_{i,t_j} \cdot \ell'_{t_j+r_j-1}(i) - \sum_{j=1}^{\sigma(T)} \ell'_{t_j+r_j-1}(k') \leq \frac{\log(K)}{\eta} + \frac{\eta}{2} \sum_{j=1}^{\sigma(T)} \sum_{i \in [K]} p_{i,t_j} \cdot \ell'^2_{t_j+r_j-1}(i). \tag{5.43}$$

Now, for all $i \in [K]$, the conditional expectation of $\ell'_{t_j+r_j-1}(i)$ is

$$\begin{aligned}
\mathbb{E}\left[\ell'_{t_j+r_j-1}(i) \middle| p_{t_j}, r_j\right] &= \sum_{t=t_j}^{t_j+r_j-1} \sum_{k': i \in S_t(k')} p_{k',t_j} \cdot \frac{\ell_t(i)}{q_{i,t}} \\
&= \sum_{t=t_j}^{t_j+r_j-1} \frac{\ell_t(i)}{q_{i,t}} \cdot \sum_{k': i \in S_t(k')} p_{k',t_j} \\
&= \sum_{t=t_j}^{t_j+r_j-1} \ell_t(i).
\end{aligned} \tag{5.44}$$

Therefore, we have that for all $i \in [K]$, the conditional expectation

$$\mathbb{E}\left[\sum_{j=1}^{\sigma(T)} \ell'_{t_j+r_j-1}(i) \middle| \{p_{t_j}, r_j\}_{1 \leq j \leq \sigma(T)}]\right] = \sum_{j=1}^{\sigma(T)} \sum_{t=t_j}^{t_j+r_j-1} \ell_t(i) = \sum_{t=1}^{T} \ell_t(i). \tag{5.45}$$

Now, the expectation of second term in right hand side of (5.43) is

$$\mathbb{E}\left[\sum_{j=1}^{\sigma(T)}\sum_{i\in[K]} p_{i,t_j}\cdot \ell'^2_{t_j+r_j-1}(i)\right] = \mathbb{E}\left[\sum_{j=1}^{\sigma(T)}\mathbb{E}\Big[\sum_{i\in[K]} p_{i,t_j}\ell'^2_{t_j+r_j-1}(i)|\{p_{t_j},r_j\}_{1\le j\le\sigma(T)}\Big]\right]$$
$$\overset{(a)}{\le} \mathbb{E}\left[\sum_{j=1}^{\sigma(T)} \text{mas}(G_{t_j:t_j+r_j-1})r_j^2\right], \tag{5.46}$$

where $\text{mas}(G_{t_j:t_j+r_j-1}) = \max_{n\in[t_j,t_j+r_j-1]}\text{mas}(G_n)$, and (a) follows from the fact that, for all $i\in[K]$ and $t\le T$, $\ell_t(i)\le 1$, and $\sum_{i\in[K]} p_{i,t}/q_{i,t}\le \text{mas}(G_t)$[8, Lemma 10].

Now, we bound $\sum_{j=1}^{\sigma(T)}\text{mas}(G_{t_j:t_j+r_j-1})r_j^2$. We write the following optimization problem:

$$\max_{\{r_j\}_{1\le j\le T}} \sum_{j=1}^{T}\text{mas}(G_{t_j:t_j+r_j-1})r_j^2, \text{ subject to} \tag{5.47}$$

$$\sum_{j=1}^{T} r_j = T,$$

$$0\le r_j\le \frac{T^{1/3}c^{2/3}}{\text{mas}^{1/3}(G_{(T)})}.$$

Since the objective function is submodular and the constraints are linear, the ratio of the solution of the greedy algorithm and the optimal solution is at most $(1-1/e)$ ([163]). Therefore, the optimal solution o^* of the above optimization problem is

$$o^*\le \sum_{t=1}^{t^*}\frac{T^{2/3}\text{mas}(G_{(t)})c^{4/3}}{(1-1/e)\text{mas}^{2/3}(G_{(T)})}, \tag{5.48}$$

where $t^* = \lceil T^{2/3}c^{-2/3}\text{mas}^{1/3}(G_{(T)})\rceil$. Using (5.43), (5.44), (5.45), (5.46) and (5.48), we have

$$\mathbb{E}\left[\sum_{j=1}^{\sigma(T)}\sum_{i\in[K]} p_{i,k_j}\sum_{t=k_j}^{k_j+r_j-1}\ell_t(i) - \sum_{j=1}^{T}\ell_t(k')\right]$$
$$\le \frac{\log(K)}{\eta} + \frac{\eta}{2}\sum_{t=1}^{t^*}\frac{T^{2/3}c^{4/3}\text{mas}(G_{(t)})}{(1-1/e)\text{mas}^{2/3}(G_{(T)})}. \tag{5.49}$$

Additionally, the player switches its action only if E^t is true. Thus, using (5.49) and $c(i,j) = c$, for all $i, j \in [K]$, we have

$$R^{\mathcal{A}}(l_{1:T}, \mathcal{C}) \leq \frac{\log(K)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{t^*} \frac{T^{2/3} c^{4/3} \text{mas}(G_{(t)})}{(1-1/e)\text{mas}^{2/3}(G_{(T)})} + c\mathbb{E}[\sum_{t=2}^{T} \mathbf{1}(i_t \neq i_{t-1})]. \tag{5.50}$$

Now, we bound $\mathbb{E}[\sum_{t=2}^{T} \mathbf{1}(i_t \neq i_{t-1})]$. E_1^t occurs with probability 1, and does not contribute to any SC. E_2^t can lead to at most $\lceil T^{2/3} c^{-2/3} \text{mas}^{1/3}(G_{(T)}) \rceil$ switches. Now, let E_3^t causes N_T switches. Then, we have

$$\begin{aligned}
&\mathbb{E}[N_T] \\
&= \mathbb{E}\left[\sum_{j=1}^{\sigma(T)} \mathbf{1}(i_{t_{j+1}} \neq i_{t_j}, E_3^{t_j} \text{ is true})\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{\sigma(T)} \mathbb{E}\left[\mathbf{1}(i_{t_{j+1}} \neq i_{t_j}, E_3^{t_j} \text{ is true}) \middle| \{p_{t_j}, r_j\}_{1 \leq j \leq \sigma(T)}\right]\right] \\
&\leq \mathbb{E}\left[\sum_{j=1}^{\sigma(T)} \mathbb{E}\left[\sum_{\substack{i \in [K], \\ k' \in [K] \setminus \{i\}}} \mathbb{P}(i_{t_j} = i \big| E_3^{t_j} \text{ is true}) \mathbb{P}(i_{t_{j+1}} = k' \big| i_{t_j} = i) \middle| \{p_{t_j}, r_j\}_{1 \leq j \leq \sigma(T)}\right]\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{\sigma(T)} \sum_{i \in [K], k' \in [K] \setminus \{i\}} p_{i,t_j} p_{k',t_{j+1}}\right] \\
&\overset{(a)}{\leq} \sum_{t=1}^{T} c^{-2/3} \text{mas}^{1/3}(G_{(T)}) t^{-1/3} = c^{-2/3} \text{mas}^{1/3}(G_{(T)}) T^{2/3},
\end{aligned} \tag{5.51}$$

where (a) follows from Lemma 20 in this section. Thus, the number of switches are bounded above by $2c^{-2/3}\text{mas}^{1/3}(G_{(T)})T^{2/3}$, and the SC is $2c^{1/3}\text{mas}^{1/3}(G_{(T)})T^{2/3}$.

Part (iii) of the theorem follows by combining the results from (i) and (ii). Part (iv) follows from the fact that if G_t is undirected, $\text{mas}(G_t) = \alpha(G_t)$. □

Lemma 20. *Given $i \in [K]$ is chosen at time instance t_j, for all $k' \in [K] \setminus \{i\}$, we have*

$$p_{i,t_j} \cdot p_{k',t_{j+1}} \leq (t_{j+1})^{-1/3}.$$

Proof. Given i is chosen at time instance t_j, for all $k' \in [K]\backslash\{i\}$, we have

$$\begin{aligned}
\frac{p_{k',t_{j+1}}}{p_{i,t_{j+1}}} &= \frac{p_{k',1}\exp(-\eta\hat{\ell}_{t_{j+1}}(k'))}{p_{i,t_j}\exp(-\eta\ell'_{t_j+r_j-1}(i))} \\
&\overset{(a)}{=} \frac{p_{k',1}\exp(-\eta(\hat{\ell}_{t_j}(k') + \ell'_{t_j+r_j-1}(k')))}{p_{i,t_j}\exp(-\eta\ell'_{t_j+r_j-1}(i))} \\
&\overset{(b)}{\leq} \frac{\exp\big(-\eta(\hat{\ell}_{t_j}(k') + \ell'_{t_j+r_j-1}(k') - \ell'_{t_j+r_j-1}(i))\big)}{p_{i,t_j}} \\
&\overset{(c)}{\leq} \frac{\exp\big(-\eta(\epsilon_{t_{j+1}}/\eta)\big)}{Kp_{i,t_j}} \\
&= \frac{\exp\big(-\epsilon_{t_{j+1}}\big)}{p_{i,t_j}},
\end{aligned} \tag{5.52}$$

where (a) follows from the fact that $\hat{\ell}_{t_{j+1}}(k') = \hat{\ell}_{t_j}(k') + \ell'_{t_j+r_j-1}(k')$; (b) follows from $p_{k',1} = 1/K$; (c) follows from the fact that for all $k \in [K]\backslash\{i\}$, $\hat{\ell}_{k,t-1} - \ell'_{i,t-1} > \epsilon_t/\eta$ as the increment in $\ell'_{i,t-1}$ is bounded by $1/q_{i,t-1}$. Now, replacing $\epsilon_t \geq \log(tc^2/\text{mas}(G_{(T)}))/3$ in (5.52), we have

$$p_{i,t_j} \cdot p_{k',t_{j+1}} \leq c^{-2/3}\text{mas}^{1/3}(G_{(T)})t_{j+1}^{-1/3}. \tag{5.53}$$

□

5.8.6 Proof of Theorem 6

Proof. We borrow the notations from the proof of Theorem 5. Using the fact that η_t is decreasing in t and (5.43), we have

$$\begin{aligned}
&\sum_{j=1}^{\sigma(T)}\sum_{i\in[K]} p_{i,t_j} \cdot \ell'_{t_j+r_j-1}(i) - \min_{k'\in[K]}\sum_{j=1}^{\sigma(T)} \ell'_{t_j+r_j-1}(k') \\
&\leq \frac{\log(K)}{\eta_T} + \sum_{j=1}^{\sigma(T)} \frac{\eta_{t_j}}{2}\sum_{i\in[K]} p_{i,t_j} \cdot \ell'^2_{t_j+r_j-1}(i).
\end{aligned} \tag{5.54}$$

Now, taking expectation on both the sides and using the fact that expectation of the min(.) is smaller than the min(.) of the expectation, we have

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{j=1}^{\sigma(T)}\sum_{i\in[K]} p_{i,t_j}\cdot \ell'_{t_j+r_j-1}(i)\right] - \min_{k'\in[K]}\mathbb{E}\left[\sum_{j=1}^{\sigma(T)}\ell'_{t_j+r_j-1}(k')\right] \\
&\leq \frac{\log(K)}{\eta_T} + \mathbb{E}\left[\sum_{j=1}^{\sigma(T)}\frac{\eta_{t_j}}{2}\epsilon_{t_j}\mathbb{E}\left[\sum_{i\in[K]} p_{i,t_j}\cdot \ell'^2_{t_j+r_j-1}(i)|p_{t_j}, r_j, \mathbf{1}(i_t \text{ is selected using } p_t)\right]\right] \\
&\overset{(a)}{\leq} \frac{\log(K)}{\eta_T} + \mathbb{E}\left[\sum_{j=1}^{\sigma(T)}\frac{\eta_{t_j}}{2}\epsilon_{t_j}\mathbb{E}[\text{mas}(G_{t_j:t_j+r_j-1})r_j^2|\mathbf{1}(i_t \text{ is selected using } p_t)]\right] \\
&\overset{(b)}{\leq} \frac{\log(K)}{\eta_T} + \mathbb{E}\left[\sum_{j=1}^{\sigma(T)}\frac{\eta_{t_j}}{2}\epsilon_{t_j}\frac{2\cdot\text{mas}(G_{t_j:t_j+r_j-1})}{\epsilon_{t_j}^2}\right] \\
&= \frac{\log(K)}{\eta_T} + \mathbb{E}\left[\sum_{j=1}^{\sigma(T)}\frac{\eta_{t_j}}{2}\frac{2\cdot\text{mas}(G_{t_j:t_j+r_j-1})}{\epsilon_{t_j}}\right] \\
&\overset{(c)}{\leq} \frac{\log(K)}{\eta_T} + \mathbb{E}\left[\sum_{j=1}^{\sigma(T)}\frac{2\log(K)}{\text{mas}^{2/3}(G_{(T)})}\text{mas}(G_{(j)})\right] \\
&\overset{(d)}{\leq} \frac{\log(K)}{\eta_T} + \sum_{j=1}^{\mathbb{E}[\sigma(T)]}\frac{2\log(K)}{\text{mas}^{2/3}(G_{(T)})}\text{mas}(G_{(j)}),
\end{aligned}
\tag{5.55}
$$

where (a) follows from (5.46), (b) follows from the fact that since the probability of selecting a new action is at most ϵ_{t_j}, the mean and the variance of the geometric random variable r_j is bounded by $1/\epsilon_{t_j}^2$ and $(1-\epsilon_{t_j})/\epsilon_{t_j}^2$ respectively, (c) follows from the value of η_t and ϵ_t, and (d) follows from the fact that $\text{mas}(G_{(j)})/\text{mas}(G_{(T)})$ is a monotonic non increasing sequence in j, therefore the summation is a concave function and the inequality follows from the Jensen's inequality.

Now, we bound the $\mathbb{E}[\sigma(T)]$ in (5.55). This also gives a bound on the number of switches

performed by the algorithm. We have

$$\begin{aligned}\mathbb{E}[\sigma(T)] &= \sum_{t=1}^{T} \mathbb{E}[\mathbf{1}(i_t \neq i_{t-1})] \\ &\leq \sum_{t=1}^{T} \epsilon_t \\ &\leq 0.5\text{mas}^{1/3}(G_{(T)})T^{2/3}c^{1/3}. \end{aligned} \tag{5.56}$$

□

Chapter 6

Attacks and Security of Multi-Armed Bandits

6.1 Introduction

Multi Armed Bandits (MAB) algorithms are often used in web services [2, 129], sensor networks [212], medical trials [21, 180], and crowdsourcing systems [174]. The distributed nature of these applications makes these algorithms prone to third party attacks. For example, in web services decision making critically depends on reward collection, and this is prone to attacks that can impact observations and monitoring, delay or temper rewards, produce link failures, and generally modify or delete information through hijacking of communication links [2] [37]. Making these systems secure requires an understanding of the regime where the systems can be attacked, as well as designing ways to mitigate these attacks. In this paper, we study both of these aspects in a stochastic MAB setting.

We consider a data poisoning attack, also referred as man in the middle (MITM) attack. In this attack, there are three agents: the environment, the learner (MAB algorithm), and the attacker. At each discrete time-step t, the learner selects an action i_t among K choices, the environment then generates a reward $r_t(i_t) \in [0, 1]$ corresponding to the selected action, and attempts to communicate it to the learner. However, an adversary intercepts $r_t(i_t)$ and can contaminate it by adding noise $\epsilon_t(i_t) \in [-r_t(i_t), 1 - r_t(i_t)]$. It follows that the learner observes the contaminated reward $r_t^o(i_t) = r_t(i_t) + \epsilon_t(i_t)$, and $r_t^o(i_t) \in [0, 1]$. Hence, the adversary acts

as a "man in the middle" between the learner and the environment. We present an upper bound on both the amount of contamination, which is the total amount of additive noise injected by the attacker, and the number of attacks, which is the number of times the adversary contaminates the observations, sufficient to ensure that the regret of the algorithm is $\Omega(T)$, where T is the total time of interaction between the learner and the environment. Additionally, we establish that this upper bound is order-optimal by providing a lower bound on the number of attacks and the amount of contamination required by a specific algorithm to suffer regret $\Omega(T)$.

A typical way to protect a distributed system from a MITM attack is to employ a secure channel between the learner and the environment [14, 202, 36]. These secure channels ensure the CIA triad: confidentiality, integrity, and availability [72, 53, 76]. Various ways to establish these channels have been explored in the literature [14, 202, 81, 36]. An alternative way to provide security is by auditing, namely perform data verification [104]. Establishing a secure channel or an effective auditing method is generally costly [202]. Hence, it is crucial to design algorithms that achieve security, namely the performance of the algorithm is unaltered in presence of attack, while limiting the usage of these additional resources.

Motivated by these observations, we consider a *reward verification* model in which the learner can access verified (i.e. uncontaminated) rewards from the environment. This verified access can be implemented through a secure channel between the learner and the environment, or using auditing. At any round t, the learner can decide whether to access the possibly contaminated reward $r_t^o(i_t) = r_t(i_t) + \epsilon_t(i_t)$, or to access the verified reward $r_t^o(i_t) = r_t(i_t)$. Since verification is costly, the learner faces a tradeoff between its performance in terms of regret, and the number of times access to a verified reward occurs. Second, the learner needs to decide when to access a verified reward during the learning process. We design an order-optimal bandit algorithm which strategically plans the verification, and makes no assumptions on the attacker's strategy or capabilities.

6.2 Contributions

Our first contribution is a tight characterization of the regret-contamination trade-off in poisoning attacks. We show that for any $\alpha \geq 1$ and any bandit algorithm, if the expected regret of an algorithm in the absence of attacks is $O((\log T)^{\alpha})$, then there exists an attack that uses $O((\log T)^{\alpha})$ expected amount of contamination and is successful, namely it forces the algorithm to suffer $\Omega(T)$ regret.

In the absence of attacks, it is known that the order optimal regret for any bandit algorithm is $O(\log T)$ (see e.g. [16]). It then follows letting $\alpha = 1$ in our results that any bandit algorithm achieving order optimal regret can be forced to suffer $\Omega(T)$ regret by injecting $O(\log T)$ expected amount of contamination. In this case, we also show that our upper bound $O(\log T)$ on the expected amount of contamination for a successful attack is tight. Namely, we show that there exists an order optimal algorithm, the classical Upper Confidence Bound (UCB) algorithm, which requires at least $\Omega(\log T)$ amount of contamination to be attacked successfully. Our results complement recent works on poisoning attacks on bandit algorithms with *unbounded* rewards [135, 97, 253]. In this case, an $\Theta(\sqrt{\log T})$ amount of contamination has been recently proved to be order-optimal to carry a successful attack [253]. Compared to our results, the main difference is that when rewards are unbounded the contamination at a single round can be arbitrarily large, and can have an indefinite effect [253]. In contrast, in our setting of *bounded* rewards the contamination at each round is also bounded, and the attack on a single round has a limited effect.

Our second contribution is to propose a novel algorithm, called Secure-Upper Confidence Bound (Secure-UCB) — a variant of the classical UCB, that overcomes poisoning attacks using verification. We show that the regret of Secure-UCB is $O(\log T)$ irrespective of the adversary's strategy. Additionally, since verification is costly, we show that the expected number of verifications performed by Secure-UCB is $O(\log T)$. Finally, we show that $\Omega(\log T)$ number of verifications are necessary for any algorithm to have $O(\log T)$ regret irrespective of the

adversary's strategy. Therefore, Secure-UCB is order-optimal in terms of both the expected regret and the expected number of verifications. We also note that in the absence of verification, any bandit algorithm that is attacked with an amount of contamination bounded by C, must experience a regret $\Omega(C)$, see, e.g., [139, 27] for more details. Our algorithm can break the barrier of this lower bound and obtain a regret $O(\log T)$, irrespective of the amount of contamination C, by using verification in an optimal way.

6.3 Related Work

The MITM attack has been previously studied in a stochastic MAB setting with *unbounded* rewards [97, 135, 253]. The work in [97] focuses on two bandits algorithms, UCB and ϵ-greedy, and shows that these algorithms can be successfully attacked using $O(\log T)$ amount of contamination. The work in [135] consider both online and offline MITM attacks. In the online setting, it shows that any order-optimal bandit algorithm can be attacked in $O(\log T)$ amount of contamination. Recently, [253] shows that the adversary needs $\Theta(\sqrt{\log T})$ amount of contamination to successfully attack the UCB algorithm. It is worthwhile to compare our results developed in a *bounded* reward setting to the ones in [253] developed in *unbounded* reward setting. There is an interesting contrast between the lower bounds in the two settings. This can be explained as follows. In the unbounded reward setting, the contamination at each round could be arbitrarily large. Therefore, the attack can drag the reward of any action to an arbitrarily negative value [253]. In contrast, in our *bounded* reward setting, the contamination at each round is bounded. It follows that each contamination has limited effect and hence it is more difficult for the attacker to be successful.

Extending the work in [97, 135], the MITM attack has also been studied in linear contextual bandits, and in this case $O(\log T)$ amount of contamination is sufficient to successfully attack the LinUCB algorithm [69]. A study regarding the feasibility of a successful attack has been performed in [141] for contextual bandits. Recently, in [236], considers a MITM attack in

the context of adversarial bandits. It shows that the regret of any bandit algorithm can be $\Omega(T)$ in the presence of $o(T)$ contamination.

Other variants of MITM attack have also been considered. The work in [136] considers action-manipulation attacks where an adversary can manipulate the action of the learner instead of the rewards, and shows that $O(\log T)$ manipulations are sufficient to successfully attack the UCB algorithm. The work [63] studies a special case of data poisoning attacks where each action of the algorithm introduces contamination in its own observation, and the algorithm unaware of the contamination introduced by other actions.

We also point out that our attacker's model significantly differs from the attacks considered by recent works on robust stochastic bandit algorithms [139, 78]. The attacker in [139, 78] has to prepare the attack observations before the action i_t is selected, and the contamination at each round t is the largest manipulation over the actions, irrespective of the selected action. In this model of weak attack, it is indeed possible to design robust stochastic bandit algorithms that achieve sub-linear regret if the total amount of contamination is $o(T)$. The adversarial attack model we consider is better aligned with the recent line of research on adversarial attacks on stochastic bandits [97, 135]. In this case, the attack occurs *after* the action i_t is selected by the algorithm. This subtle difference turns out to make the attacker significantly more powerful, as we show that any $O(\log T)$-regret stochastic algorithms will suffer $\Omega(T)$ regret with at most $O(\log T)$ expected amount of contamination.[1] The design of robust algorithms in [139, 78] has been extended to episodic reinforcement learning and learning product ranking [140, 73].

More recently, [27] considers a strong attacker model similar to ours in the linear bandit setting with contextual features, and designs robust bandit algorithms in the presence of an amount of contamination C, whose *instant-independent* regret is $O(C)$. In contrast, our objective is to use limited reward verification to obtain the an *instant-dependent* order-wise optimal $O(\log T)$ regret that is independent of the amount of contamination C. It is worth noting that their robust algorithm crucially relies on the assumption that the amount of contamination is *almost surely*

[1] This has also been proved in [135] for the unbounded rewards setting.

bounded by C. In our work, the amount of contamination is bounded only *in expectation* (see Section 6.7 for a detailed comparison).

The adversarial attacks has been studied in supervised learning [74, 89]. Additionally, there has been efforts towards analyzing the robustness of neural networks to these adversarial attacks [228, 224, 218]. Finally, differential privacy is considered as a defensive mechanism from MITM attack in supervised learning [143].

The MITM attacks has also been studied in Reinforcement Learning (RL) [142, 243, 242, 250]. These works study the feasibility of these attacks in RL, and provide an upper bound on the attack cost, which varies with the attacker's objective, for an attacker's strategy. Related to RL, these attacks have also been studied in linear control systems [152, 192, 204, 170, 182]. These works focus on detecting the attacks, develop methodologies to mitigate the attacks, and provide both upper bound and lower bound on the attacker's cost.

6.4 Preliminaries and Problem Statement

6.4.1 Poisoning Attacks on Stochastic Bandits

We consider the classical stochastic bandit setting under data poisoning attacks. In this setting, a learner can choose from a set of K actions for T rounds. At each round t, the learner chooses an action $i_t \in [K]$, triggers a reward $r_t(i_t) \in [0, 1]$ and observes a possibly corrupted (and thus altered) reward $r_t^o(i_t) \in [0, 1]$ corresponding to the chosen action. The reward $r_t(i)$ of action i is sampled independently from a fixed unknown distribution of action i. Let μ_i denote the expected reward of action i and $i^* = \text{argmax}_{i\in[K]}\mu_i$.[2] Also, let $\Delta(i) = \mu_{i^*} - \mu_i$ denote the difference between the expected reward of actions i^* and i. Finally, we assume that $\{\mu_i\}_{i\in[K]}$ are unknown to both the *learner* and the *attacker*.

The reward $r_t^o(i_t)$ observed by the learner and the true reward $r_t(i_t)$ satisfy the following

[2]For convenience, we assume i^* is unique though all our conclusions hold when there are multiple optimal actions.

relation

$$r_t^o(i_t) = r_t(i_t) + \epsilon_t(i_t), \tag{6.1}$$

where the contamination $\epsilon_t(i_t)$ added by the attacker can be a function of $\{i_n\}_{n=1}^t$ and $\{r_n(i_n)\}_{n=1}^t$. Additionally, since $r_t^o(i_t) \in [0,1]$, we have that $\epsilon_t(i_t) \in [-r_t(i_t), 1 - r_t(i_t)]$. If $\epsilon_t(i_t) \neq 0$, then the round t is said to be *under attack*. Hence, the *number of attacks* is $\sum_{t=1}^T \mathbf{1}(\epsilon_t(i_t) \neq 0)$ and the *amount of contamination* is $\sum_{t=1}^T |\epsilon_t(i_t)|$.

The regret $R^{\mathcal{A}}(T)$ of a learning algorithm $\mathcal{A}$ is the difference between the total expected true reward from the best fixed action and the total expected *true* reward over T rounds, namely

$$R^{\mathcal{A}}(T) = T\mu_{i^*} - \mathbb{E}[\sum_{t=1}^T r_t(i_t)], \tag{6.2}$$

The objective of the learner is to minimize the regret $R^{\mathcal{A}}(T)$. In contrast, the objective of the attacker is to increase the regret to at least $\Omega(T)$. As a convention, we say the attack is "successful" only when it leads to $\Omega(T)$ regret [97, 135]. The first question we address is the following.

Question 1: *Is there a* tight characterization *of the amount of contamination and the number of attacks leading to a regret of* $\Omega(T)$ *in stochastic bandits?*

6.4.2 Remedy via Limited Reward Verification

It is well known that no stochastic bandit algorithm can be resilient to data poisoning attacks if the attacker has sufficiently large amount of contamination [135]. Therefore, to guarantee sub-linear regret when the attacker has an unbounded amount of contamination it is necessary for the bandit algorithm to exploit additional (and possibly costly) resources. We consider one of the most natural resource — *verified rewards*. Namely, we assume that at any round t, the learner can choose to access the true, uncontaminated reward of the selected action i_t, namely, when *round t is verified* we have $r_t^o(i_t) = r_t(i_t)$. This process of accessing true rewards is referred to as *verification*. If the learner performs verification at each round, then it is clear that the regret of any bandit algorithm is unaltered in the presence of attacker. Unfortunately,

this is unrealistic because verification is costly in practice. Therefore, the learner has to carefully balance the regret and the number of verifications. This naturally leads to the second question that we aim to answer in this paper:

Question 2: *Is there a* tight characterization *of the number of verifications needed by the learner to guarantee the optimal $O(\log T)$ regret for* any *poisoning attack?*

In this paper we answer both the above questions in the affirmative.

6.5 Tight Characterization for the Cost of Poisoning Attack on Stochastic Bandits

We now design a data poisoning attack that with $O(\log T)$ expected number of attacks leads to $\Omega(T)$ regret for any order-optimal bandit algorithm, namely the algorithm which has $O(\log T)$-regret in the absence of attack. Since $r_t^o(i_t) \in [0, 1]$, this also implies that the attack would require at most $O(\log T)$ expected amount of contamination. Moreover, we show that both the expected number of attacks $O(\log T)$ and the expected amount of contamination $O(\log T)$ are order-wise optimal. Specifically, there exists an order-optimal stochastic bandit algorithm (the UCB algorithm) which cannot be successfully attacked with $o(\log T)$ expected amount of contamination (or equivalently with $o(\log T)$ expected number of attacks). The key technical aspect of this section lies in our second result showing that any poisoning attack must use at least $\Omega(\log T)$ amount of contamination in order to force UCB to suffer $\Omega(T)$ regret. En route, we prove an novel "convervativeness" property of the UCB algorithm which may be of independent interest.

6.5.1 Upper Bound on the Contaminations

We consider an attack where the attacker tries to ensure that a sub-optimal action $i_A \neq i^*$ will be selected by the bandit algorithm at least $\Omega(T)$ times in expectation. As a consequence, this would imply that the expected regret of the bandit algorithm is $\Omega(T)$. It suffices to consider the following simple attack, which pulls the observed reward down to 0 whenever the target

suboptimal action i_A is not selected. Namely,

$$r_t^o(i_t) = \begin{cases} r_t(i_t) & \text{if } i_t = i_A, \\ 0 & \text{if } i_t \neq i_A. \end{cases} \tag{6.3}$$

Equivalently, the attacker adds $\epsilon_t(i_t) = -r_t(i_t)\mathbf{1}(i_t \neq i_A)$ to the true reward $r_t(i_t)$. Unlike the attacks in [97, 135], the attack in (6.3) can also be considered oblivious, since it overwrites the rewards observation by zero irrespective of $r_t(i_t)$. The following proposition establishes an upper bound on the expected number of attacks sufficient to be successful.

Proposition 1. *For any stochastic bandit algorithm $\mathcal{A}$ with expected regret in the* absence *of attack given by*

$$R^{\mathcal{A}}(T) = O\bigg(\sum_{i \neq i^*} \frac{\log^{\alpha}(T)}{(\Delta(i))^{\beta}}\bigg), \tag{6.4}$$

where $\alpha \geq 1$ and $\beta \geq 1$; and for any sub-optimal target action $i_A \in [K] \setminus i^*$; if an attacker follows strategy (6.3)*, then it will use an expected number of attacks*

$$\mathbb{E}[\sum_{t=1}^{T} \mathbf{1}(\epsilon_t(i_t) \neq 0)]] = O\bigg(\frac{(K-1)\log^{\alpha}(T)}{\mu_{i_A}^{\beta+1}}\bigg), \tag{6.5}$$

and it will force $\mathcal{A}$ to select the action i_A at least $\Omega(T)$ times in expectation, namely $\mathbb{E}[\sum_{t=1}^{T} \mathbf{1}(i_t = i_A)] = \Omega(T)$.

Proposition 1 provides a relationship between the regret of the algorithm without attack and the number of attacks sufficient to ensure that the target action i_A is selected $\Omega(T)$ times, which also implies $R^{\mathcal{A}}(T) = \Omega(T)$. Additionally, since $\epsilon_t(i_t) \leq 1$, we have that using (6.5), the expected amount of contamination is $O((\log T)^{\alpha})$ as well, namely $\mathbb{E}[\sum_{t=1}^{T} |\epsilon_t(i_t)|] = O((\log T)^{\alpha})$. In (6.5), the number of attacks is inversely proportional to the mean μ_{i_A} of the target action. This is because the reward observation of all other actions $i \neq i_A$ are zero, and i_A is the optimal action for the algorithm, which implies that $\Delta(i)$ in (6.4) equals μ_{i_A} for all $i \neq i_A$

during the attack. Another important consequence of the proposition is that for an order optimal algorithm such as UCB, we have that $\alpha = 1$ and $\beta = 1$ in (6.4). Thus, the expected number of attacks and the expected amount of contamination are $O(\log T)$.

A small criticism to the attack strategy (6.3) might be that it pulls down the reward "too much". This turns out to be fixable. In Appendix 6.11.2, we prove that a different type of attack that pulls the reward of any action $i \neq i_A$ down by an *estimated* gap $\Delta = 2\max\{\mu_i - \mu_{i_A}, 0\}$ (similar to the ACE algorithm in [141]) will also succeed. However, the number of attacks now will be inversely proportional to $\min_{i \neq i_A} |\mu_i - \mu_{i_A}|^{\beta+1}$, while not $\mu_{i_A}^{\beta+1}$ as in Proposition 1.

6.5.2 Matching Lower Bound on the Contaminations

We now show that the simple attack strategy analyzed in Proposition 1 is essentially order-optimal. That is, there exists an order-optimal bandit algorithm — in fact, the classical UCB algorithm — which cannot be attacked with $o(\log T)$ amount of contamination by *any* poisoning attack strategy. This implies that if an attacking strategy is required to be successful for all order-optimal bandit algorithms, then the amount of contamination needed is at least $\Omega(\log T)$. Since the amount of contamination is bounded above by the number of attacks, this also implies that any attacker requires at least $\Omega(\log T)$ number of attacks to be successful.

Here we briefly describe the well-known UCB algorithm [16], and defer its details to Algorithm 11. At each round $t \leq K$, UCB selects an action in round robin manner. At each round $t > K$, the selected action i_t has the maximum *upper confidence bound*, namely

$$i_t = \text{argmax}_{i \in [K]} \left(\hat{\mu}_{t-1}(i) + \sqrt{\frac{8 \log t}{N_{t-1}(i)}} \right), \tag{6.6}$$

where $N_t(i) = \sum_{n=1}^{t} \mathbf{1}(i_n = i)$ is the number of rounds action i is selected until (and including) round t, and

$$\hat{\mu}_t(i) = \frac{\sum_{n=1}^{t} r_n^o(i_n) \mathbf{1}(i_n = i)}{N_t(i)}, \tag{6.7}$$

is the empirical mean of action i until round t. Note that the algorithm uses the *observed* rewards.

Algorithm 11. (Classical) Upper Confidence Bound

For all $i \in [K]$, initialize $\hat{\mu}_0(i) = 0$, $N_0(i) = 0$.
for $t \leq K$ **do**
  Choose action $i_t = t$, and observe $r_t(i_t)$.
  Update $\hat{\mu}_t(i_t) = r_t^o(i_t)$, $N_t(i_t) = N_{t-1}(i_t) + 1$
  For all $i \neq i_t$, $\hat{\mu}_t(i) = \hat{\mu}_{t-1}(i)$, $N_t(i) = N_{t-1}(i)$.
end for
for $K + 1 \leq t \leq T$ **do**
  Choose action i_t such that

$$i_t = \text{argmax}_{i \in [K]} \left[\hat{\mu}_{t-1}(i) + \sqrt{\frac{8 \log t}{N_{t-1}(i)}} \right]. \tag{6.8}$$

  Update $N_t(i_t) = N_{t-1}(i_t) + 1$, and

$$\hat{\mu}_t(i_t) = \frac{\hat{\mu}_{t-1}(i_t) \cdot N_{t-1}(i_t) + r_t^o(i_t)}{N_{t-1}(i_t) + 1}. \tag{6.9}$$

  For all $i \in [K] \setminus i_t$, $\hat{\mu}_t(i) = \hat{\mu}_{t-1}(i)$ and $N_t(i) = N_{t-1}(i)$.
end for

The following Theorem 23 establishes that the UCB algorithm will have sublinear regret $o(T)$ under any poisoning attack if the amount of contamination is $o(\log T)$. The proof of Theorem 23 crucially hinges on the following "conservativeness" property about the UCB algorithm, which might be of independent interest.

Lemma 21 (Conservativeness of UCB)**.** *Let t_0 be such that $t_0/(\log(t_0))^2 \geq 36K^2$. Then for all $t \geq t_0$ and any sequence of rewards $\{r_n^o(i)\}_{i \in [K], n \leq t}$ in $[0, 1]$ (can even be adversarial), UCB will select every action at least $\log(t/2)$ times up until round t.*

Lemma 21 is inherently due to the design of the UCB algorithm. Its proof does *not* rely on the rewards being stochastic, and it holds deterministically — i.e., at any time $t \geq t_0$, UCB will pull each action at least $\log(t/2)$ times. This lemma leads to the following theorem.

Theorem 23. *For all $0 < \epsilon < 1$ and $\alpha > 0$ such that $0 < \epsilon\alpha \leq 1/2$, and for all $T >$*

$\max\{(t_0)^{\frac{1}{1-\alpha\epsilon}}, \exp(4^\alpha)\}$*, if the total* amount *of contamination by the attacker is*

$$\sum_{n=1}^{T} |\epsilon_n(i_n)| \leq (\log T)^{1-\epsilon}, \tag{6.10}$$

then there exists a constant c_1 *such that the expected regret of UCB algorithm is*

$$R^{UCB}(T) \leq c_1\big(T^{1-\alpha\epsilon} \max_i \Delta(i) + \sum_{i \neq i^*} \log T/\Delta(i)\big). \tag{6.11}$$

The constant α in Theorem 23 is an adjustable *parameter* to control the tradeoff between the scale of time horizon T ($T \geq \max\{(t_0)^{\frac{1}{1-\alpha\epsilon}}, \exp(4^\alpha)\}$) and the dominating term ($T^{1-\alpha\epsilon} \max_i \Delta(i)$) in the regret. If ϵ is small, then the larger α leads to a smaller regret, however T should be sufficiently large in order for us to see such a regret.

The upper bound on the expected regret in Theorem 23 holds if the total amount of contamination in (6.10) is at most $(\log T)^{1-\epsilon}$. Furthermore, if the total number of attacks is at most $(\log T)^{1-\epsilon}$, then using $|\epsilon_t(i_t)| \leq 1$, we have that (6.10) holds. Hence, Theorem 23 also establishes that if the total number of attacks is $o(\log T)$, then the expected regret of UCB is $o(T)$. In other words, the attacker requires at least $\Omega(\log T)$ amount of contamination (or number of attacks) to ensure its success.

The lower bound on the amount of contamination in Theorem 23 cannot be directly compared with the upper bound in Proposition 1 since the former assumes that the amount of contamination is bounded above by $o(\log T)$ *almost surely*, while the latter is a bound on the *expected* amount of contamination. Instead, we consider the following corollary, which can be easily derived from Theorem 23 using Markov's inequality, and establishes the lower bound on the expected amount of contamination necessary for a successful attack.

Corollary 23.1. *For all* $0 < \epsilon < 1$ *and sufficiently large* T *such that the conditions in Theorem 23 are satisfied, if the expected amount of contamination by the attacker is at most* $(\log T)^{1-\epsilon}$*, then the regret of UCB is* $o(T)$.

6.6 Secure Upper Confidence Bound

In this section, we propose the *Secure Upper Confidence Bound* (Secure-UCB) algorithm which utilizes verification, and is robust to *any* data poisoning attack. Specifically, Secure-UCB uses only $O(\log T)$ reward verifications and exhibits $O(\log T)$ regret, *irrespective of the amount of contamination and the number of attacks* . Moreover, we prove that $\Omega(\log T)$ verifications are necessary for any bandit algorithm to have $O(\log T)$ regret. Therefore, Secure-UCB uses an order-optimal number of verifications $O(\log T)$, and guarantees the order-optimal regret $O(\log T)$.

The details of Secure-UCB are presented in Algorithm 12. At each round $t \leq K$, Secure-UCB selects an action $i \in [K]$ in round-robin manner, verifies all the reward observations and updates the corresponding parameters, see Algorithm 12. At each round $t > K$, Secure-UCB selects an action i_t with the largest upper confidence bound of similar format as the classical UCB. However, Secure-UCB differs from UCB in the following three crucial aspects:

1. The confidence interval $\hat{\mu}_t(i) + \sqrt{8 \log t / N_t(i)}$ of the classical UCB algorithm depends on the total number of rounds the action i is *selected* until round t, namely $N_t(i)$. However, in Secure-UCB, the confidence interval uses the total number of rounds the action i is *verified* until round t, namely $N_t^s(i)$. Note that, like classical UCB, Secure-UCB also uses the empirical mean $\hat{\mu}_t(i)$ of the *observed* rewards.

2. At each round t, Secure-UCB takes an additional step to decide whether to verify the reward of the current action i_t or not, based on a carefully designed criterion.

3. If the algorithm decides to not verify the reward observation $r_t^o(i_t)$, then it will additionally decide, based on another carefully designed criterion, whether to ignore the current unverified rewards $r_t^o(i_t)$ by *not* updating both the empirical mean $\hat{\mu}_{t-1}(i_t)$ and the number of rounds the current action is selected $N_{t-1}(i_t)$.

The first and second deviations from UCB, as described above, are to guarantee that at

any round, the algorithm always has the correct and sufficient confidence level. This is done through: (1) using the number of verified rewards for the confidence interval since the algorithm is only certain about these observations; (2) dynamically requesting additional verifications as the algorithm proceeds to guarantee the desirable confidence level. The third deviation is designed to control the integration of unverified rewards into the empirical mean estimation so that it does not contain too many attacked (or unverified) rewards.

Next, we give more details about the Secure-UCB algorithm. The first aspect described above of using number of verifications in the confidence interval is easy to implement, and is presented in (6.14). Our descriptions below are primarily focused on the second and third key differences.

Criterion for Performing Verification.

Secure-UCB maintains a count $N_t^s(i)$ of the number of verification performed until round t for each action $i \in [K]$. Additionally, it also maintains a "secured" empirical mean $\hat{\mu}_t^s(i)$, which is the empirical estimate of the mean of action i using all the verified reward observations until round t. Secure-UCB uses this mean $\hat{\mu}_t^s(i)$ in the criterion to decide whether to perform additional verification. Specifically, it performs a verification at round t if the following criterion holds:

$$N_{t-1}^s(i_t) \leq 1200 \log T / \hat{\Delta}_{t-1}^{*2}, \tag{6.12}$$

where $\hat{\Delta}_t^*$ is intuitively the estimation of $\min_{i \neq i^*} \Delta(i)$, which is the difference between the largest expected reward and the second largest expected reward. In Secure-UCB, this estimation $\hat{\Delta}_t^*$ is based on the verified rewards and is defined as the difference between the largest lower confidence bound (obtained by, say action a_t^*) and the largest upper confidence bound among all actions excluding a_t^*, namely

$$\hat{\Delta}_t^* = \max \left\{ 0, \hat{\mu}_t^s(a_t^*) - \sqrt{\frac{3 \log T}{N_t^s(a_t^*)}} - \hat{\mu}_t^s(\tilde{a}_t) - \sqrt{\frac{3 \log T}{N_t^s(\tilde{a}_t)}} \right\}, \tag{6.13}$$

where

$$a_t^* = \text{argmax}_{a\in[K]}\left[\hat{\mu}_t^s(a) - \sqrt{3\log T/N_t^s(a)}\right],$$

$$\tilde{a}_t = \text{argmax}_{a\in[K]\setminus a^*}\left[\hat{\mu}_t^s(a) + \sqrt{3\log T/N_t^s(a)}\right].$$

We show in Lemma 25 in Appendix 6.11.5 that with high probability, we have $\hat{\Delta}_t^* \leq \min_{i\neq i^*}\Delta(i)$. Note that, after verification, the algorithm will observe the true reward, namely $r_t^o(i_t) = r_t(i_t)$ at round t. Also, (6.12)- (6.13) depend on the time horizon T. This is for convenience of our analysis — if T is unknown, the doubling trick can be used in conjunction with Secure-UCB [24].

Criterion for Integrating Unverified Rewards into Empirical Mean.

The UCB term of Secure-UCB, presented in (6.14), relies on the empirical estimate $\hat{\mu}_t(i)$ of the mean of action i which is estimated using all the verified reward observations and some unverified observations of action i. Specifically, given that the reward is *not* verified at current round t, namely (6.12) does not hold, the algorithm will include the unverified observation $r_t^o(i)$ in the estimate $\hat{\mu}_t(i)$ if the following event $\mathcal{S}_t(i)$ occurs

$$\mathcal{S}_t(i) = \left\{\frac{P_t(i) + 1 + L_t(i) + r_t^o(i)}{N_{t-1}(i)+1} \leq \frac{\max\{\hat{\Delta}_{t-1}(i), \hat{\Delta}_{t-1}^*\}}{20}\right\}, \tag{6.18}$$

where $N_t(i)$ is the total number of observations (verified and un-verified) used to calculate $\hat{\mu}_t(i)$, $P_t(i) = N_{t-1}(i) - N_{t-1}^s(i)$ is the total number of unverified observations used to calculate $\hat{\mu}_{t-1}(i)$,

$$L_t(i) = \hat{\mu}_{t-1}(i)N_{t-1}(i) - \hat{\mu}_{t-1}^s(i)N_{t-1}^s(i), \tag{6.19}$$

is the total unverified reward observations previously included in the estimate of $\hat{\mu}_{t-1}(i)$, and

$$\hat{\Delta}_t(i) = \max_{a\in[K]}\left\{0, \hat{\mu}_t^s(a) - \sqrt{\frac{3\log T}{N_t^s(a)}} - \hat{\mu}_t^s(i) - \sqrt{\frac{3\log T}{N_t^s(i)}}\right\}. \tag{6.20}$$

Algorithm 12. Secure Upper Confidence Bound

For all $i \in [K]$, initialize $\hat{\mu}_0(i) = 0$, $N_0(i) = 0$, $\hat{\mu}_0^s(i) = 0$, $N_0^s(i) = 0$, $t = 1$.
for $t \leq K$ **do**
  Choose action $i_t = t$.
  Verify the observed reward, i.e., $r_t^o(i_t) = r_t(i_t)$.
  Update $\hat{\mu}_t(i_t) = r_t(i_t)$, $N_t(i_t) = N_{t-1}(i_t) + 1$, $\hat{\mu}_t^s(i_t) = r_t(i_t)$, $N_t^s(i_t) = N_{t-1}^s(i_t) + 1$.
  For all $i \in [K] \setminus i_t$, $\hat{\mu}_t(i) = \hat{\mu}_{t-1}(i)$, $N_t(i) = N_{t-1}(i)$, $\hat{\mu}_t^s(i) = \hat{\mu}_{t-1}^s(i)$, $N_t^s(i) = N_{t-1}^s(i)$.
end for
For all $i \in [K]$, update $\hat{\Delta}_K(i)$ in (6.20) and $\hat{\Delta}_K^*$ in (6.13).
for $K + 1 \leq t \leq T$ **do**
  Choose action i_t such that

$$i_t = \text{argmax}_{i \in [K]} \big(\hat{\mu}_{t-1}(i) + \sqrt{400 \log T / N_{t-1}^s(i)} \big). \tag{6.14}$$

  if $N_{t-1}^s(i_t) \leq 1200 \log T / \hat{\Delta}_{t-1}^{*2}$ **then**
    Verify the observed reward, i.e., $r_t^o(i_t) = r_t(i_t)$.
    Update $N_t(i_t) = N_{t-1}(i_t) + 1$, $N_t^s(i_t) = N_{t-1}^s(i_t) + 1$,

$$\hat{\mu}_t^s(i_t) = \big(\hat{\mu}_{t-1}^s(i_t) \cdot N_{t-1}^s(i_t) + r_t(i_t) \big) / \big(N_{t-1}^s(i_t) + 1 \big), \tag{6.15}$$

$$\hat{\mu}_t(i_t) = \big(\hat{\mu}_{t-1}(i_t) \cdot N_{t-1}(i_t) + r_t^o(i_t) \big) / \big(N_{t-1}(i_t) + 1 \big), \tag{6.16}$$

    Update $\hat{\Delta}_t(i)$ in (6.20) and $\hat{\Delta}_t^*$ in (6.13), $\forall i \in [K]$.
  else
    Observe reward $r_t^o(i_t)$.
    if $\mathcal{S}_t(i_t)$ defined in Equation (6.18) is true **then**
      Update $N_t(i_t) = N_{t-1}(i_t) + 1$, $N_t^s(i_t) = N_{t-1}^s(i_t)$, $\hat{\mu}_t^s(i_t) = \hat{\mu}_{t-1}^s(i_t)$,

$$\hat{\mu}_t(i_t) = \big(\hat{\mu}_{t-1}(i_t) \cdot N_{t-1}(i_t) + r_t^o(i_t) \big) \big(N_{t-1}(i_t) + 1 \big). \tag{6.17}$$

    else
      Update $\hat{\mu}_t(i_t) = \hat{\mu}_{t-1}(i_t)$, $N_t(i_t) = N_{t-1}(i_t)$, $\hat{\mu}_t^s(i_t) = \hat{\mu}_{t-1}^s(i_t)$, $N_t^s(i_t) = N_{t-1}^s(i_t)$.
    end if
  end if
  For all $i \neq i_t$, $\hat{\mu}_t(i) = \hat{\mu}_{t-1}(i)$, $N_t(i) = N_{t-1}(i)$, $\hat{\mu}_t^s(i) = \hat{\mu}_{t-1}^s(i)$, $N_t^s(i) = N_{t-1}^s(i)$.
end for

Intuitively, $\hat{\Delta}_t(i)$ in (6.20) is an estimation of the gap $\Delta(i) = \mu_{i^*} - \mu_i$, namely the difference between the expected reward of the optimal action i^* and the action i. Like the estimate $\hat{\Delta}_t^*$ in (6.13), this estimate is also computed using only the verified mean $\hat{\mu}_t^s(i)$ and $N_t^s(i)$. We show in Lemma 24 in the appendix that with high probability, we have $\hat{\Delta}_t(i) \leq \Delta(i)$.

Finally, we briefly discuss the criterion described in the event $S_t(i)$ in (6.18), re-stating the criteria

$$\frac{P_t(i) + 1 + L_t(i) + r_t^o(i)}{N_{t-1}(i) + 1} \leq \frac{\max\{\hat{\Delta}_{t-1}(i), \hat{\Delta}_{t-1}^*\}}{20}. \tag{6.21}$$

The ratio in the left hand side of (6.21) represents the contribution of the unverified reward observations and their count to the empirical mean $\hat{\mu}_t(i)$ relative to the total number of (verified and un-verified) observations considered. This relative contribution is required to be less than $\max\{\hat{\Delta}_{t-1}(i), \hat{\Delta}_{t-1}^*\}/20$ for a new un-verified reward observation to be considered in $\hat{\mu}_t(i)$. This implies that if the number of verified observations is large, then the new unverified observations can be considered in $\hat{\mu}_t(i)$, and the error in the estimate will be small even if these unverified observations are corrupted by an adversary. Thus, the event $S_t(i)$ balances the tradeoff between the gain from utilization of information and the adversarial effects that may occur if the information is corrupted.

Order-Optimality of Secure-UCB.

The following theorems establish the upper bound on both the regret of Secure-UCB and the expected number of verifications performed.

Theorem 24. *For all T such that $T \geq c_2 \log T / \min_{i \neq i^*} \Delta^2(i)$, Secure-UCB performs $O(\log T)$ number of verification in expectation, and the expected regret of the algorithm is $O(\log T)$ irrespective of the attacker's strategy. Namely,*

$$\sum_{i \in [K]} \mathbb{E}[N_T^s(i)] \leq c_3 \left(\sum_{i \neq i^*} \log T / \Delta^2(i) \right), \tag{6.22}$$

$$R^{SUCB}(T) \leq c_4 \bigg(\sum_{i \neq i^*} \log T / \Delta(i) \bigg), \tag{6.23}$$

where c_2, c_3 and c_4 are numerical constants whose values can be found in the appendix.

Theorem 24 establishes that the regret of Secure-UCB in stochastic bandit setting is $O(\sum_{i \neq i^*} \log(T)/\Delta(i))$ *irrespective of the attacker's strategy*. This regret bound is of the same order as the regret bound of the classical UCB algorithm without attack. This implies that Secure-UCB is order optimal in terms of regret, and is robust to any adversary if it can selectively verify up to $O(\log T)$ reward observations in expectation.

We now show that the number of verifications performed by Secure-UCB is essentially order-optimal. Specifically, the following theorem establishes that for all consistent learning algorithm[3] $\mathcal{A}$ and sufficiently large T, if the algorithm $\mathcal{A}$ uses $O((\log T)^{1-\alpha})$ verifications with $0 < \alpha < 1$, then the expected regret is $\Omega((\log T)^{\beta})$ with $\beta > 1$ in the MAB setting with verification.

Theorem 25. *Let $KL(i_1, i_2)$ denote the KL divergence between the distributions of actions i_1 and i_2. For all $0 < \alpha < 1$, $1 < \beta$ and all consistent learning algorithm $\mathcal{A}$, there exists a time t^* and an attacking strategy such that for all $T \geq 2t^*$ satisfying $(\log T)^{1-\alpha} + \beta \log(4 \log T) \leq \log T$, if the total number of verifications N_T^s until round T is*

$$N_T^s < (\log T)^{1-\alpha} / \min_{i_1, i_2 \in [K]} KL(i_1, i_2), \tag{6.24}$$

then the expected regret of $\mathcal{A}$ is at least $\Omega((\log T)^{\beta})$.

Theorem 25 establishes that $\Omega(\log T)$ verifications are necessary to obtain $O(\log T)$ regret. Here, we assume that the number of verifications is bounded above *almost surely*. To

[3] A learning algorithm is consistent [106] if for all t, the action i_{t+1} (a random variable) is measurable given the history $\mathcal{F}_t = \sigma(i_1, r_1^o(i_1), i_2, r_2^o(i_2) \ldots, i_t, r_t^o(i_t))$.

make this result comparable with the result from Theorem 24 which provides an upper bound on the expected number of verifications, we consider the following bound instead.

Corollary 25.1. *For all $0 < \alpha < 1$, $1 < \beta$, all consistent learning algorithm $\mathcal{A}$ and sufficiently large T such that the requirements in Theorem 25 are satisfied, there exists an attacking strategy such that if the* expected number of verifications N_T^s *until round T is*

$$\mathbb{E}[N_T^s] < (\log T)^{1-\alpha} / \min_{i_1, i_2 \in [K]} KL(i_1, i_2), \tag{6.25}$$

then the expected regret of $\mathcal{A}$ is at least $\Omega((\log T)^{\beta})$.

This with Theorem 24 show that Secure-UCB uses order-optimal number of verification, and enjoys an order-optimal expected regret, irrespective of the attacker's strategy.

6.7 Comparison of Attacker Models

In this section we provide a more detailed comparison between the different attacker models from the (robust bandits) literature and their corresponding performance guarantees. In particular, at each round t, a *weak attacker* has to make the contamination *before* the actual action is chosen. On the other hand, a *strong attacker* can observe both the chosen actions and the corresponding rewards before making the contamination. From the perspective of contamination budget (or the amount of contamination), it can either be bounded above surely by a threshold, or that bound only holds in expectation. We refer to the former as *deterministic budget*, while we call the latter as *expected budget*. To date, the following three attacker models have been studied: (i) weak attacker with deterministic budget; (ii) strong attacker with deterministic budget; and (iii) strong attacker with expected budget.

Weak attacker with deterministic budget.

For this attacker model, [78] have proposed a robust bandit algorithm (called BARBAR) that provably achieves $O(KC + (\log T)^2)$ regret against a weak attacker with (unknown) de-

terministic budget C. They have also proved a matching regret lower bound of $\Omega(C)$. These results imply that in order to successfully attack BARBAR (i.e., to force a $\Omega(T)$ regret), a weak attacker with deterministic budget would need a contamination budget of $\Omega(T)$.

Strong attacker with deterministic budget.

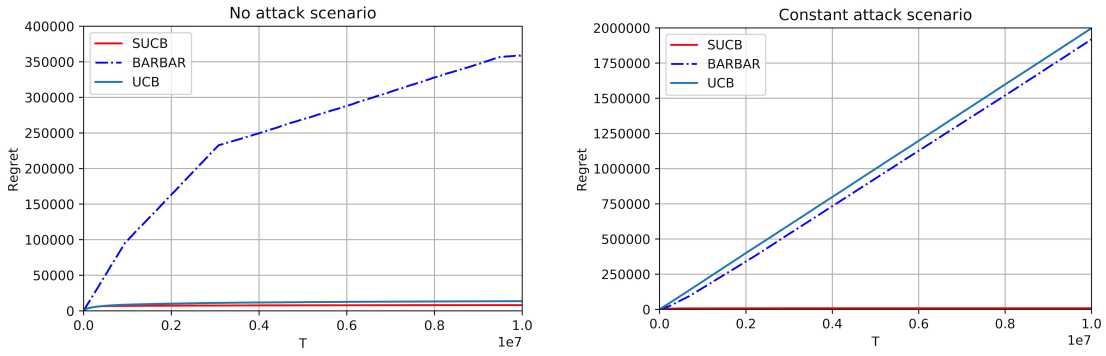
[27] have shown that there is a phased elimination based bandit algorithm that achieves $O(\sqrt{T} + C \log T)$ regret if C is known to the algorithm, and $O(\sqrt{T} + C \log T + C^2)$ if C is unknown. Note that by moving from the weaker attacker model to the stronger one, we suffer an extra loss in terms of achievable regret (i.e., from $O(C)$ to $O(C^2)$) in case of unknown C. While the authors have also proved a matching regret lower bound of $\Omega(C)$ for the known budget case, they have not provided any similar results for the case of unknown budget. Nevertheless, their results show that in order to successfully attack their algorithm, an attacker of this type would need a contamination budget of $\Omega(T)$ for the case of known contamination budget, and $\Omega(\sqrt{T})$ if that budget is unknown.

Strong attacker with expected budget.

Our Proposition 1 shows that this attacker can successfully attack any order-optimal algorithm with a $O(\log T)$ expected contamination budget (note that [135] have also proved a similar, but somewhat weaker result). We have also provided a matching lower bound on the necessary amount of expected contamination budget against UCB. It is worth noting that if the rewards are unbounded, then the attacker may use even less amount contamination (e.g., $O(\sqrt{\log T})$) to achieve a successful attack [253].

Saving bandit algorithms with verification.

The abovementioned results also indicate that if an attacker uses a contamination budget C (either deterministic or expected), the regret that any (robust) algorithm would suffer is $\Omega(C)$. A simple implication of this is that if an attacker has a budget of $\Theta(T)$ (e.g., he can contaminate all the rewards), then no algorithm can maintain a sub-linear regret if they can only rely on the observed rewards. Secure-UCB breaks this barrier of $\Omega(C)$ regret with verification. In particular,



(a) Regret versus T in the absence of attacker

(b) Regret versus T in the presence of attacker

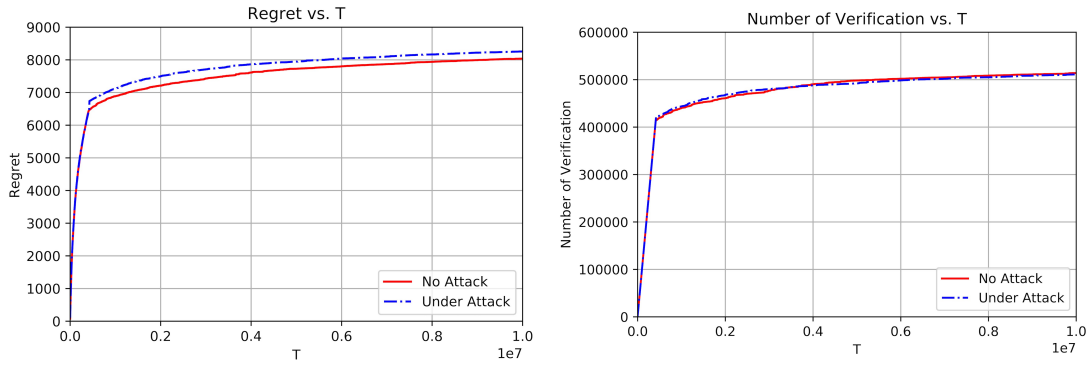
Figure 6.1. Comparison between Secure-UCB, UCB and BARBAR

it can still enjoy an order-optimal regret of $O(\log T)$ against any attacker (even when they have $\Theta(T)$ contamination budget) while only using $O(\log T)$ verifications.

6.8 Simulation Results

For numerical analysis, we consider two actions, namely $K = 2$, with Bernoulli reward, and we have $\mu_1 = 0.5$ and $\mu_2 = 0.7$. Hence, the optimal action is the second one. We evaluate the performance of Secure-UCB in two scenarios. First, the attacker is absent, namely for all t, we have $r_t^o(i_t) = r_t(i_t)$. Second, the attacker is present, and uses the attack strategy defined by (6.3), namely for all t, we have $r_t^o(i_t) = r_t(i_t)\mathbf{1}(i_t = 1)$. We also compare the performances of Secure-UCB, the classical UCB and BARBAR algorithm in [78] in these two scenarios. BARBAR algorithm has been considered here since it is a robust algorithm for stochastic bandits and an improvement over the algorithm in [139].

Figure 6.1 shows a comparison between the regret of the three algorithms, Secure-UCB, UCB and BARBAR, under the two scenarios. In Figure 6.1a, the regret of Secure-UCB and UCB is close to each other when the attacker is absent. Additionally, the regret of BARBAR algorithm is greater than both UCB and Secure-UCB in the absence of the attacker. This is inline with the theoretical results since the regret of UCB and Secure-UCB is $O(\log T)$ and the regret



(a) **(b)**

Figure 6.2. Performance of Secure-UCB

of BARBAR is $O((\log T)^2)$ in the absence of attacker [78]. In Figure 6.1b, the regret of both the algorithms UCB and BARBAR grows linearly in T in the presence of attacker. On contrary, by comparing Figures 6.1b and 6.2a, the regret of Secure-UCB is $O(\log T)$ in the presence of attacker. This is inline with our results in Theorems 1 and 24, and the results in [78].

Figure 6.2 shows a comparison between the performance of Secure-UCB in the two scenarios. Figure 6.2a shows that the regret of Secure-UCB in the presence of attacker is more than the regret in the absence of attacker. However, the regret grows $O(\log T)$ in both these scenarios. Figure 6.2b shows that the number of verifications performed in the two scenarios are close to each other. This is also inline with our theoretical result since the verification grows $O(\log T)$. Thus, the regret and the number of verifications are similar in both the presence and absence of attacker. Hence, Secure-UCB is immune to the attack.

6.9 Conclusion

This paper proposes Secure-UCB, which uses verification to mitigate a strong attacker. We show that with $O(\log T)$ expected number of verifications, Secure-UCB can recover the order optimal regret irrespective of the attacker's strength, and this number of verifications is necessary. We also prove that without verification, with $O(\log T)$ expected amount of contamination, a strong attacker can succeed against any order optimal bandit methods, and that this amount is

tight, in case of bounded rewards.

Since bounding the contamination in expectation and almost surely leads to different results (see Section 6.7), it would be interesting to study the setting where number of verifications is bounded almost surely. Another interesting extension is a *limited verification* model, where the learner can request a feedback if the observed reward is corrupted or not, however it cannot observe the true reward. Both these problems can be studied in a strong and weak attacker model.

Extending these results beyond bandits setting, we can study these attack models in episodic reinforcement learning (RL) setting, and can derive the lower bound on the amount of contamination for a successful attack on Q-learning using analogue of the *conservativeness of UCB*. Similar to Secure-UCB, algorithms can be developed in the RL setting to save it from adversarial attacks using the verification.

In adversarial bandits setup, the lower bound on the attack cost is an open question. Additionally, we can explore the feasibility of saving the adversarial bandits from data poising attacks by using the reward verification. Finally, designing the secure and optimal algorithms for adversarial setting is an interesting future direction.

MAB has been studied in the presence of switching cost, where switching between action requires additional cost [49, 175, 12]. Extending along the same direction, using our verification model, MAB can also be studied in the presence of feedback cost, where observing feedback corresponding to the selected action requires additional cost.

6.10 Acknowledgement

Chapter 6, in part, contains material as it appears in Anshuka Rangi, Long Tran-Thanh, Haifeng Xu and Massimo Franceschetti, "Saving Stochastic Bandits from Poisoning Attacks via Limited Data Verification", *under preparation*. The dissertation author was the co-primary investigator and co-author of this paper.

6.11 Appendix

6.11.1 Proof of Proposition 1

Let $N_t(i)$ be the number of times action i is chosen by the learner until time t, namely

$$N_t(i) = \sum_{n=1}^{t} \mathbf{1}(i_n = i). \tag{6.26}$$

Then, we have that

$$\sum_{t=1}^{T} \mathbf{1}(\epsilon_t(i_t) \neq 0) \leq \sum_{i \neq i_A} N_T(i). \tag{6.27}$$

Using (6.3), for all $i \in [K] \setminus i_A$ and $t \leq T$, we have that

$$\mathbb{E}[r_t^o(i)] = 0, \tag{6.28}$$

and

$$\mathbb{E}[r_t^o(i_A)] = \mu_{i_A}. \tag{6.29}$$

Since the algorithm $\mathcal{A}$ makes decision based on the $r_t^o(.)$, using (6.4), (6.28) and (6.29), we have that

$$\mathbb{E}[T\mu_{i_A} - \sum_{t=1}^{T} r_t^o(i_t)] = O\bigg(\frac{(K-1)\log^{\alpha}(T)}{\mu_{i_A}^{\beta}}\bigg). \tag{6.30}$$

Also, we have

$$\mathbb{E}[T\mu_{i_A} - \sum_{t=1}^{T} r_t^o(i_t)] \overset{(a)}{=} \mu_{i_A}\mathbb{E}[\sum_{i \neq i_A} N_T(i)], \tag{6.31}$$

where (a) follows from the fact that $\Delta(i) = \mu_{i_A}$ for the learner. This along with (6.30) implies that

$$\mathbb{E}[\sum_{i \neq i_A} N_T(i)] = O\bigg(\frac{(K-1)\log^{\alpha}(T)}{\mu_{i_A}^{\beta+1}}\bigg). \tag{6.32}$$

Now, we have

$$\mathbb{E}[N_T(i_A)] = T - \sum_{i \neq i_A} \mathbb{E}[N_T(i)], \tag{6.33}$$

which using (6.30) and (6.31), implies the attack is successful, i.e., $\mathbb{E}[\sum_{t=1}^{T} \mathbf{1}(i_t = i_A)] = \Omega(T)$. Combining (6.27) and (6.32), we have

$$\mathbb{E}[\sum_{t=1}^{T} \mathbf{1}(\epsilon_t(i_t) \neq 0)] = O\bigg(\frac{(K-1)\log^{\alpha}(T)}{\mu_{i_A}^{\beta+1}}\bigg). \tag{6.34}$$

Hence, the statement of the proposition follows.

6.11.2 Attacks Based on Gap Estimation

The attack is similar to the ACE attack in [141]. Specifically, the attacker maintains an estimate $\hat{\Delta}_t^A(i, i_A)$ of $\mu_i - \mu_{i_A}$ using the previously selected actions and their rewards, namely

$$\hat{\Delta}_t^A(i, i_A) = \hat{\mu}_t(i) + \sqrt{\frac{2\log T}{\tilde{N}_t(i)}} - \hat{\mu}_t(i_A) + \sqrt{\frac{2\log T}{\tilde{N}_t(i_A)}}, \tag{6.35}$$

where

$$\hat{\mu}_t(i) = \frac{\sum_{n=1}^{t} r_n(i)\mathbf{1}(i_n = i)}{\sum_{n=1}^{t} \mathbf{1}(i_n = i)}, \tag{6.36}$$

and $\tilde{N}_t(i) = \sum_{n=1}^{t} \mathbf{1}(i_n = i)$. In this attack, we have

$$r_t^o(i_t) = \begin{cases} \max\{0, r_t(i_t) - 2\max\{0, \hat{\Delta}_t^A(i_t, i_A)\}\} & \text{if } i_t \neq i_A, \\ r_t(i_t) & \text{if } i_t = i_A. \end{cases} \tag{6.37}$$

This implies that for all $t \leq T$, the noise added by the attacker is

$$\epsilon_t(i_t) = -2\max\{0, \hat{\Delta}_t^A(i_t, i_A)\}\mathbf{1}(i_t \neq i_A). \tag{6.38}$$

In this attack, for all action $i \neq i_A$ such that $\mu_i > \mu_{i_A}$, the attacker forces the expected observed reward to be at most $\mu_{i_A} - (\mu_i - \mu_{i_A})$, and for all action $i \neq i_A$ such that $\mu_i < \mu_{i_A}$, the attacker forces the expected observed reward to be at most μ_i. Therefore, this strategy ensures that the optimal action is i_A based on the observed rewards. The following proposition establishes the success of the attack, and provides an upper bound on the expected number of contaminations needed by the attacker.

Proposition 2. *For any stochastic bandit algorithm $\mathcal{A}$ with expected regret in the absence of attack given by*

$$R^{\mathcal{A}} = O\bigg(\sum_{i \neq i^*} \frac{\log^{\alpha}(T)}{\Delta^{\beta}(i)}\bigg), \tag{6.39}$$

where $\alpha \geq 1$ and $\beta \geq 1$; and for any sub-optimal target action $i_A \in [K] \setminus i^*$, if an attacker follows strategy (6.37)*, then it will use an expected number of attacks*

$$\mathbb{E}[\sum_{t=1}^{T} \mathbf{1}(\epsilon_t(i_t) \neq 0)] = O\bigg(\sum_{i \neq i_A} \frac{\log^{\alpha}(T)}{|\mu_{i_A} - \mu_i|^{\beta}(\min_{i' \neq i_A} |\mu_{i'} - \mu_{i_A}|)}\bigg), \tag{6.40}$$

and it will force $\mathcal{A}$ to select the action i_A at least $\Omega(T)$ times in expectation, namely $\mathbb{E}[\sum_{t=1}^{T} \mathbf{1}(i_t = i_A)] = \Omega(T)$.

Proof. We will use the following lemma.

Lemma 22. *For all $t > K$ and $i \in [K]$, we have that*

$$\mathbb{P}(\hat{\Delta}_t^A(i, i_A) \leq \mu_i - \mu_{i_A}) \leq \frac{1}{T^3}. \tag{6.41}$$

Proof. Using Theorem 26, for all $i \in [K]$, we have that

$$\mathbb{P}\bigg(\hat{\mu}_t(i) + \sqrt{\frac{2 \log T}{\bar{N}_t(i)}} \leq \mu_i\bigg) \leq 1/T^4, \tag{6.42}$$

$$\mathbb{P}\left(\hat{\mu}_t(i_A) - \sqrt{\frac{2\log T}{\bar{N}_t(i_A)}} \geq \mu_{i_A}\right) \leq 1/T^4. \tag{6.43}$$

This implies that for all $i \in [K]$ and $K < t \leq T$, we have

$$\begin{aligned}
&\mathbb{P}(\hat{\Delta}_t^A(i, i_A) \leq \mu_i - \mu_{i_A}) \\
&\leq \mathbb{P}\left(\hat{\mu}_t(i) + \sqrt{\frac{2\log T}{\bar{N}_t(i)}} \leq \mu_i\right) + \mathbb{P}\left(\hat{\mu}_t(i_A) - \sqrt{\frac{2\log T}{\bar{N}_t(i_A)}} \geq \mu_{i_A}\right), \\
&\leq 2/T^4 \leq 1/T^3.
\end{aligned} \tag{6.44}$$

The statement of the lemma follows. □

Now consider the following event

$$\mathcal{E} = \{\forall i \in [K], \forall t \leq T : \hat{\Delta}_t^A(i, i_A) \geq \mu_i - \mu_{i_A}\}. \tag{6.45}$$

Similar to (6.27), we have that

$$\sum_{t=1}^{T} \mathbf{1}(\epsilon_t(i_t) \neq 0) \leq \sum_{i \neq i_A} N_T(i). \tag{6.46}$$

Using (6.37), under event $\mathcal{E}$, for all $i \in [K] \setminus i_A$ such that $\mu_i > \mu_{i_A}$ and $t \leq T$, we have that

$$\mathbb{E}[r_t^o(i)] \leq \mu_i - 2(\mu_i - \mu_{i_A}) = \mu_{i_A} - (\mu_i - \mu_{i_A}). \tag{6.47}$$

Also, for all $i \in [K] \setminus i_A$ such that $\mu_i < \mu_{i_A}$ and $t \leq T$, we have that

$$\mathbb{E}[r_t^o(i)] \leq \mu_i. \tag{6.48}$$

Since the algorithm $\mathcal{A}$ makes decision based on the $r_t^o(.)$, under event $\mathcal{E}$, using (6.47) and (6.48),

we have that

$$\mathbb{E}[T\mu_{i_A} - \sum_{t=1}^{T} r_t^o(i_t)\big|\mathcal{E}] = O\Big(\sum_{i\neq i_A} \frac{\log^\alpha(T)}{|\mu_{i_A} - \mu_i|^\beta}\Big). \tag{6.49}$$

Also, we have

$$\mathbb{E}[T\mu_{i_A} - \sum_{t=1}^{T} r_t^o(i_t)\big|\mathcal{E}] = \sum_{i\neq i_A} |\mu_{i_A} - \mu_i|\mathbb{E}[N_T(i)\big|\mathcal{E}] \geq \min_{i\neq i_A} |\mu_i - \mu_{i_A}|\mathbb{E}[\sum_{i\neq i_A} N_T(i)\big|\mathcal{E}]. \tag{6.50}$$

Additionally, using Lemma 22, we have

$$\mathbb{P}(\bar{\mathcal{E}}) = \sum_{t=1}^{T} \frac{K-1}{T^3} \leq \frac{K-1}{T^2}. \tag{6.51}$$

Now, we have

$$\mathbb{E}[N_T(i_A)] = T - \sum_{i\neq i_A} \mathbb{E}[N_T(i)], \tag{6.52}$$

which using (6.49), (6.50) and (6.51), implies $\mathbb{E}[\sum_{t=1}^{T} \mathbf{1}(i_t = i_A)] = \Omega(T)$. Combining (6.46), (6.49), (6.50) and (6.51), we have

$$\mathbb{E}[\sum_{t=1}^{T} \mathbf{1}(\epsilon_t(i_t) \neq 0)] = O\Big(\sum_{i\neq i_A} \frac{\log^\alpha(T)}{|\mu_{i_A} - \mu_i|^\beta (\min_{i'\neq i_A} |\mu_{i'} - \mu_{i_A}|)}\Big). \tag{6.53}$$

Hence, the statement of the proposition follows.

□

6.11.3 Proof of Theorem 23

The proof crucially relies on the following "conservativeness" property of the UCB algorithm.

Lemma 23. *[Restating Lemma 21] Let t_0 be such that $t_0/(\log t_0)^2 \geq 36K^2$. For all $t \geq t_0$ and for any sequence of rewards $\{r_n^o(i)\}_{i\in[K],n\leq t}$ in $[0,1]$ in $[0,1]$ (can even be adversarial), UCB will select every action $i \in [K]$ at least $\log(t/2)$ times until round t.*

Proof. Let $N_t(i)$ be the number of times action i is selected until round t, namely

$$N_t(i) = \sum_{n=1}^{t} \mathbf{1}(i_n = i), \tag{6.54}$$

and $M_t(i)$ be the number of attacks on action i until round t by the attacker. With slight abuse of notation, we use $[k]$ to denote the set of actions that are pulled strictly less than $\log(t/2)$ times until round t.

We prove this lemma by contradiction. Suppose that there exists some time $t \geq t_0$ and $k \leq K$ actions such that for all $i \in [k]$,

$$N_t(i) < \log(t/2). \tag{6.55}$$

We now divide the time interval $[t/2, 3t/4]$ into $k \log(t/2)$ consecutive blocks of the same length. Thus, the length of each block is $t/(4k \log(t/2))$. By the pigeonhole principle, there must exist one block $[t_1, t_3]$ in which we did not select any action in $[k]$, namely

$$t_3 = t_1 + t/4k \log(t/2), \tag{6.56}$$

and for all $i \in [k]$, we have

$$N_{t_1-1}(i) = N_{t_3}(i). \tag{6.57}$$

First, we provide a lower bound on the UCB index of all actions $i \in [k]$ within the time interval (or block) $[t_1, t_3]$. For all $t_2 \in [t_1, t_3]$ and $i \in [k]$, we have

$$\hat{\mu}_{t_2}(i) + \sqrt{\frac{8 \log t_2}{N_{t_2-1}(i)}} \overset{(a)}{>} \sqrt{\frac{8 \log(t/2)}{\log(t/2)}} = 2\sqrt{2}, \forall i \in [k] \tag{6.58}$$

where (a) follows from the facts that $t_2 \geq t_1 \geq t/2$, and $N_{t_2-1}(i) \leq N_t(i) < \log(t/2)$ using (6.55).

Second, we show that using the lower bound on the UCB index in (6.58) for actions in $[k]$, no actions outside $[k]$ can be pulled by more than $8\log(3t/4)+1$ times within the interval $[t_1, t_3]$, namely for all $i \in [K] \setminus [k]$, we have

$$N_{t_3}(i) - N_{t_1-1}(i) \leq 8\log(3t/4) + 1. \tag{6.59}$$

We prove (6.59) by contradiction. Suppose (6.59) does not hold. Then, there exists an action $j \in [K] \setminus [k]$ and a time $t_2 \in [t_1, t_3]$ such that action j is selected, namely $i_{t_2} = j$, and

$$N_{t_2-1} = 8\log(3t/4) + 1. \tag{6.60}$$

Therefore, at round t_2, the UCB index of action j is

$$\hat{\mu}_{t_2}(j) + \sqrt{\frac{8\log t_2}{N_{t_2-1}(j)}} \overset{(a)}{\leq} 1 + \sqrt{\frac{8\log(3t/4)}{N_{t_2-1}(j)}} \overset{(b)}{<} 2, \tag{6.61}$$

where (a) follows from the facts that observed rewards are in the interval $[0,1]$, and $t_2 \leq t_3 \leq 3t/4$, and (b) follows from (6.60). This however is a contradiction since (6.58) shows that the UCB index of action $i \in [k]$ at time t_2 is strictly larger than 2. Therefore, action j cannot have the the largest UCB index at t_2, and thus cannot be selected. Thus, we have that (6.59) follows.

Finally, combining (6.59) and the fact that actions in the set $[k]$ are not selected in $[t_1, t_3]$, we have that

$$\sum_{i\in[K]} (N_{t_3}(i) - N_{t_1}(i)) \leq (K-k)(8\log(3t/4)+1). \tag{6.62}$$

This along with (6.56) implies that

$$(K-k)(8\log(3t/4)+1) \geq \frac{t}{4k\log(t/2)}, \tag{6.63}$$

or equivalently

$$4k(K-k) \geq \frac{t}{\log(t/2)(8\log(3t/4)+1)}. \tag{6.64}$$

We also have

$$\frac{t}{\log(t/2)(8\log(3t/4)+1)} \geq \frac{t}{\log(t)(9\log(t))} \overset{(a)}{\geq} \frac{t_0}{9(\log(t_0))^2} \overset{(b)}{\geq} 4K^2, \tag{6.65}$$

where (a) follows from the facts that $t \geq t_0$ and $t/(\log t)^2$ is an increasing function of t, and (b) follows from the assumption the lemma that $t_0/(\log(t_0))^2 \geq 36K^2$. Thus, since $k \geq 1$, we have that (6.64) and (6.65) contradict each other. Thus, the interval $[t_1, t_3]$ does not exists, which implies (6.55) does not hold.

Note that in our proof we did not make any assumption about the sequence of the rewards, except that they are bounded within $[0, 1]$. Therefore, it holds for arbitrary reward sequence. □

For the remainder of the theorem's proof, we will use the definition of t_0 from Lemma 21. For all $0 < \epsilon < 1$ and $\alpha > 0$ such that $0 < \epsilon\alpha \leq 1/2$, and for all $T > \max\{(t_0)^{\frac{1}{1-\alpha\epsilon}}, \exp(4^\alpha)\}$, we have that

$$\log T \geq 4^\alpha \overset{(a)}{\geq} (1-\alpha\epsilon)^{-\alpha/(\alpha\epsilon)}, \tag{6.66}$$

where (a) follows from the fact that using $\epsilon\alpha \leq 1/2$, we have $4 \geq (1-\alpha\epsilon)^{-1/(\alpha\epsilon)}$. Using (6.66), we have that

$$(\log T)^{1-\epsilon} \leq (1-\alpha\epsilon)\log T. \tag{6.67}$$

Let $\hat{\mu}_t(i)$ is the empirical mean of action i at time t using all the observed rewards (including the contaminated ones), namely

$$\hat{\mu}_t(i) = \frac{\sum_{n=1}^{t} r_n^o(i)\mathbf{1}(i_n = i)}{N_t(i)}, \tag{6.68}$$

where $N_t(i) = \sum_{n=1}^{t} \mathbf{1}(i_n = i)$. Let $\hat{\mu}_t^m(i)$ denote the empirical mean of action i at time t using the true reward without contamination, namely

$$\hat{\mu}_t^m(i) = \frac{\sum_{n=1}^{t} r_n(i)\mathbf{1}(i_n = i)}{N_t(i)}. \tag{6.69}$$

Also, let $c_t(i) = \sum_{n=1}^{t} |\epsilon_n(i_n)|\mathbf{1}(i_n = i)$ denote the total amount of contamination on action i until time t. Thus, we have

$$\hat{\mu}_t^m(i) - \frac{c_t(i)}{N_t(i)} \leq \hat{\mu}_t(i) \leq \hat{\mu}_t^m(i) + \frac{c_t(i)}{N_t(i)}. \tag{6.70}$$

By our hypothesis in the theorem statement, for all $t \leq T$ and $i \in [K]$, we have that

$$c_t(i) \leq (\log T)^{1-\epsilon}. \tag{6.71}$$

We now will examine the UCB index of all actions for any time $t > 2T^{1-\alpha\epsilon}$, which is at least t_0 by our choice of T. Using Lemma 23, for all $t > 2T^{1-\alpha\epsilon} \geq t_0$ and $i \in [K]$, we have that

$$N_t(i) \geq \log(t/2). \tag{6.72}$$

Using (6.67) and the fact that $t > 2T^{1-\alpha\epsilon}$, we also have

$$\log(t/2) > \log T^{1-\alpha\epsilon} = (1 - \alpha\epsilon)\log T \geq (\log T)^{1-\epsilon}. \tag{6.73}$$

Now, for all $t > 2T^{1-\alpha\epsilon}$, the UCB index of the optimal action i^* satisfies

$$\begin{aligned}
\hat{\mu}_t(i^*) + \sqrt{\frac{8\log t}{N_t(i^*)}} &\overset{(a)}{\geq} \hat{\mu}_t^m(i^*) - \frac{c_t(i^*)}{N_t(i^*)} + \sqrt{\frac{8\log t}{N_t(i^*)}} \\
&\overset{(b)}{\geq} \hat{\mu}_t^m(i^*) - \frac{(\log T)^{1-\epsilon}}{N_t(i^*)} + \sqrt{\frac{8\log t}{N_t(i^*)}} \\
&\overset{(c)}{\geq} \hat{\mu}_t^m(i^*) - \sqrt{\frac{(\log T)^{1-\epsilon}}{N_t(i^*)}} + \sqrt{\frac{8\log t}{N_t(i^*)}} \\
&\overset{(d)}{\geq} \hat{\mu}_t^m(i^*) - \sqrt{\frac{\log(t/2)}{N_t(i^*)}} + \sqrt{\frac{8\log t}{N_t(i^*)}} \\
&\geq \hat{\mu}_t^m(i^*) + (\sqrt{8}-1)\sqrt{\frac{\log t}{N_t(i^*)}},
\end{aligned} \tag{6.74}$$

where (a) follows from (6.70), (b) follows from (6.71), (c) follows from the fact that using (6.72) and (6.73), we have $(\log T)^{1-\epsilon}/N_t(i^*) \leq 1$, and (d) follows from (6.73).

Similarly, we can show that for all $t > 2T^{1-\alpha\epsilon}$, the UCB index of any sub-optimal action $i \neq i^*$ satisfies

$$\begin{aligned}
\hat{\mu}_t(i) + \sqrt{\frac{8\log t}{N_t(i)}} &\leq \hat{\mu}_t^m(i) + \frac{c_t(i)}{N_t(i)} + \sqrt{\frac{8\log t}{N_t(i)}} \\
&\leq \hat{\mu}_t^m(i) + (\sqrt{8}+1)\sqrt{\frac{\log t}{N_t(i)}}.
\end{aligned} \tag{6.75}$$

Combining (6.74) and (6.75), and using the standard analysis of the UCB algorithm for $2T^{1-\alpha\epsilon} < t \leq T$, we can show that sub-optimal action $i \neq i^*$ is pulled at most $O(\log T/\Delta_i^2)$ times after round $2T^{1-\alpha\epsilon}$. Consequently, the total number of times that sub-optimal actions are pulled is at most $2T^{1-\alpha\epsilon} + O(\log T/\Delta_i^2)$ times, and the regret is upper bound by $2T^{1-\alpha\epsilon}\max_i \Delta(i) + O(\sum_{i\neq i^*} \log T/\Delta(i))$. Since $2T^{1-\alpha\epsilon} < T$, the regret of UCB is sub-linear in $o(T)$

6.11.4 Proof of Corollary 23.1

Let $\delta = (\log T)^{-\epsilon/2}$. Using Markov inequality, and the fact that the expected amount of contamination is at most $(\log T)^{1-\epsilon}$, we have

$$\mathbb{P}(\sum_{t=1}^{T} |\epsilon_t(i_t)| \geq (\log T)^{1-\epsilon/2}) \leq \frac{(\log T)^{1-\epsilon}}{(\log T)^{1-\epsilon/2}} = \delta. \tag{6.76}$$

Also, for all $0 < \epsilon < 1$, there exists a constant $\beta > 0$ such that

$$(\log T)^{-\epsilon/2} T \leq T^{1-\beta}, \tag{6.77}$$

or equivalently

$$\beta \leq \frac{\epsilon \log\log T}{2 \log T}. \tag{6.78}$$

Using Theorem 23 and (6.76), we have that

$$\begin{aligned} R^{UCB}(T) &\leq (1-\delta) c_1 \big(T^{1-\alpha\epsilon/2} \max_i \Delta(i) + \sum_{i \neq i^*} \log T/\Delta(i)\big) + \delta T, \\ &\leq c_1 \big(T^{1-\alpha\epsilon/2} \max_i \Delta(i) + \sum_{i \neq i^*} \log T/\Delta(i)\big) + (\log T)^{-\epsilon/2} T, \\ &\overset{a}{\leq} c_1 \big(T^{1-\alpha\epsilon/2} \max_i \Delta(i) + \sum_{i \neq i^*} \log T/\Delta(i)\big) + T^{1-\beta}, \end{aligned} \tag{6.79}$$

where (a) follows from (6.77). Since $1 - \alpha\epsilon < 1$ and $1 - \beta < 1$, the statement of the theorem follows.

6.11.5 Proof of Theorem 24

Theorem 26. *Hoeffding's inequality: Let $x_1, \ldots, X_n$ be independent and identically distributed random variable, such that for all i, we have $0 \leq X_i \leq 1$ and $\mathbb{E}[X_i] = \mu$. Then,*

$$\mathbb{P}\left(\frac{\sum_{i=1}^{n} X_i}{n} - \mu \geq \sqrt{\frac{\log(1/\delta)}{2n}}\right) \leq \delta, \tag{6.80}$$

$$\mathbb{P}\left(\mu - \frac{\sum_{i=1}^{n} X_i}{n} \geq \sqrt{\frac{\log(1/\delta)}{2n}}\right) \leq \delta. \tag{6.81}$$

Lemma 24. *For all $i \in [K]$, $K < t \leq T$, we have*

$$\mathbb{P}(\hat{\Delta}_t(i) \geq \Delta(i)) \leq K/T^6. \tag{6.82}$$

Proof. Using Theorem 26, for all $i \in [K]$, we have that

$$\mathbb{P}\left(\hat{\mu}_t^s(i) - \sqrt{\frac{3\log T}{N_t^s(i)}} \geq \mu_i\right) \leq 1/T^6, \tag{6.83}$$

$$\mathbb{P}\left(\hat{\mu}_t^s(i) + \sqrt{\frac{3\log T}{N_t^s(i)}} \leq \mu_i\right) \leq 1/T^6. \tag{6.84}$$

This implies that for all $i \in [K]$ and $K < t \leq T$, we have

$$\begin{aligned}
&\mathbb{P}(\hat{\Delta}_t(i) \geq \Delta(i)) \\
&\leq \mathbb{P}\left(\hat{\mu}_t^s(i) + \sqrt{\frac{3\log T}{N_t^s(i)}} \leq \mu_i\right) + \sum_{i' \neq i} \mathbb{P}\left(\hat{\mu}_t^s(i') - \sqrt{\frac{3\log T}{N_t^s(i')}} \geq \mu_{i'}\right), \\
&\leq K/T^6
\end{aligned} \tag{6.85}$$

The statement of the lemma follows. □

Lemma 25. *For all $K < t \leq T$, we have*

$$\mathbb{P}(\hat{\Delta}_t^* > \min_{i \neq i^*} \Delta(i)) \leq 2K/T^6. \tag{6.86}$$

Proof. Let $\mu_{i^*} - \mu_{i_1} = \min_{i \neq i^*} \Delta(i)$. Also, for all $K < t \leq T$, let an event

$$\mathcal{E}(t) = \left\{ \forall i \in [K] : |\hat{\mu}_t^s(i) - \mu_i| \leq \sqrt{\frac{3 \log T}{N_t^s(i)}} \right\}. \tag{6.87}$$

If $\mathcal{E}(t)$ occurs, then we have

$$\hat{\mu}_t^s(a_t^*) - \sqrt{\frac{3 \log T}{N_t^s(a_t^*)}} \leq \mu_{i^*}. \tag{6.88}$$

Also, if $\mathcal{E}(t)$ occurs, then there exist two action i^* and i_1 such that

$$\hat{\mu}_t^s(i^*) + \sqrt{3 \log T / N_t^s(i^*)} \geq \mu_{i_1}, \tag{6.89}$$

and

$$\hat{\mu}_t^s(i_1) + \sqrt{3 \log T / N_t^s(i_1)} \geq \mu_{i_1}. \tag{6.90}$$

This implies that

$$\hat{\mu}_t^s(\tilde{a}_t) + \sqrt{\frac{3 \log T}{N_t^s(\tilde{a}_t)}} \geq \mu_{i_1}. \tag{6.91}$$

Using (6.88) and (6.91), if $\mathcal{E}(t)$ occurs, then we have

$$\hat{\Delta}_t^* = \hat{\mu}_t^s(a_t^*) - \sqrt{\frac{3 \log T}{N_t^s(a_t^*)}} - \hat{\mu}_t^s(\tilde{a}_t) - \sqrt{\frac{3 \log T}{N_t^s(\tilde{a}_t)}} \leq \mu_{i^*} - \mu_{i_1} = \min_{i \neq i^*} \Delta(i). \tag{6.92}$$

This implies that

$$\begin{aligned} \mathbb{P}(\Delta_t^* > \min_{i \neq i^*} \Delta(i)) &\leq \mathbb{P}(\bar{\mathcal{E}}(t)) \\ &\overset{(a)}{\leq} \frac{2K}{T^6}, \end{aligned} \tag{6.93}$$

where $\bar{\mathcal{E}}(t)$ denotes the complement of the event $\mathcal{E}(t)$, and (a) follows from (6.83) and (6.84).

□

Lemma 26. *For all $i \in [K]$ and $K < t \leq T$, we have that*

$$\mathbb{P}\bigg(|\hat{\mu}_t(i) - \hat{\mu}_t^s(i)| \geq \Delta(i)/20\bigg) \leq 3K/T^6. \tag{6.94}$$

Proof. Let

$$t' = \max\{n \leq t : i_t = i, r_n^o(i) \text{ was not verified, and } \mathcal{S}_t(i) \text{ occurs}\}. \tag{6.95}$$

If t' does not exist, then using the algorithm, we have $\hat{\mu}_t(i) = \hat{\mu}_t^s(i)$, and the lemma trivially follows.

We have

$$\begin{aligned}
|\hat{\mu}_t(i) - \hat{\mu}_t^s(i)| &= \left|\frac{\hat{\mu}_t^s(i)N_t^s(i) + \sum_{n=1}^{t'} r_n^o(i)\mathbf{1}(i_n = i \text{ and } \mathcal{S}_n(i))}{N_t(i)} - \hat{\mu}_t^s(i)\right| \\
&\overset{(a)}{\leq} \hat{\mu}_t^s(i)\left|1 - \frac{N_t^s(i)}{N_t(i)}\right| + \left|\frac{\sum_{n=1}^{t'} r_n^o(i)\mathbf{1}(i_n = i \text{ and } \mathcal{S}_n(i))}{N_t(i)}\right| \\
&\overset{(b)}{\leq} \frac{N_t(i) - N_t^s(i) + \sum_{n=1}^{t'} r_n^o(i)\mathbf{1}(i_n = i \text{ and } \mathcal{S}_n(i))}{N_t(i)} \\
&\overset{(c)}{\leq} \frac{N_{t'}(i) - N_{t'}^s(i) + \sum_{n=1}^{t'} r_n^o(i)\mathbf{1}(i_n = i \text{ and } \mathcal{S}_n(i))}{N_{t'}(i)} \\
&\leq \frac{\max\{\hat{\Delta}_{t-1}(i), \hat{\Delta}_{t-1}^*\}}{20},
\end{aligned} \tag{6.96}$$

where (a) follows from the fact that $|a+b| \leq |a| + |b|$, (b) follows from the fact that $\hat{\mu}_t^s(i) \leq 1$, (c) follows from the fact that $N_{t'}(i) \leq N_t(i)$, and the fact that using (6.95), we have

$$N_{t'}(i) - N_{t'}^s(i) = N_t(i) - N_t^s(i), \tag{6.97}$$

and (d) follows from the fact that $N^s_{t'}(i) = N^s_{t'-1}(i)$ and $\mathcal{S}_{t'}(i)$ occurs. This implies that

$$\begin{aligned}\mathbb{P}(|\hat{\mu}_t(i) - \hat{\mu}^s_t(i)| > \Delta(i)/20) &= \mathbb{P}(\max\{\hat{\Delta}_{t-1}(i), \hat{\Delta}^*_{t-1}\} > \Delta(i)) \\ &\leq \mathbb{P}(\hat{\Delta}_{t-1}(i) > \Delta(i)) + \mathbb{P}(\hat{\Delta}^*_{t-1} > \Delta(i)) \\ &\overset{(a)}{\leq} \frac{3K}{T^6}, \end{aligned} \tag{6.98}$$

where (a) follows from Lemma 24 and Lemma 25.

□

Lemma 27. *Let $\mathcal{T}$ be the set of rounds for which verification is not performed. Let function $f(T)$ be*

$$f(T) = \frac{1200 \log T}{\min_{i \neq i^*} \Delta^2(i)} + \sum_{i \neq i^*} \frac{10000 \log T}{9\Delta^2(i)} + K - 1. \tag{6.99}$$

Let T is sufficiently large such that $T \geq f(T)$. Then, for all $f(T) \leq t \leq T$ and for all $i \in [K]$, we have

$$\mathbb{P}\left(N^s_t(i^*) \leq \frac{1200 \log T}{\min_{i \neq i^*} \Delta^2(i)} \right) \leq \frac{10K^2}{T^5}, \tag{6.100}$$

and

$$\mathbb{P}\left(\forall i \neq i^* : N^s_t(i) \leq \frac{2500 \log T}{121\Delta^2(i)} \text{ or } N^s_t(i) \geq \frac{10000 \log T}{9\Delta^2(i)} + 1 \right) \leq \frac{10K^2}{T^5}. \tag{6.101}$$

Additionally, for all $t \in \mathcal{T}$ such that $K \leq t$, we have that

$$\mathbb{P}(i_t \neq i^*) \leq \frac{10K^2}{T^5}. \tag{6.102}$$

Proof. Let $\mathcal{T}$ be a set of rounds such that for all $t \in \mathcal{T}$, the action $i_t \in [K]$ satisfies

$$N^s_{t-1}(i_t) > \frac{1200 \log T}{\hat{\Delta}^{*2}_t}. \tag{6.103}$$

Consider the following events

$$\mathcal{E}_1(t) = \left\{\forall i \in [K] \text{ and } \forall K \leq t' \leq t : |\hat{\mu}^s_{t'}(i) - \mu_i| \leq \frac{1}{2}\sqrt{\frac{400 \log T}{N^s_{t'}(i)}}\right\}, \tag{6.104}$$

$$\mathcal{E}_2(t) = \left\{\forall i \in [K] \text{ and } \forall K \leq t' \leq t : |\hat{\mu}^s_{t'}(i) - \hat{\mu}_{t'}(i)| \leq \Delta(i)/20\right\}, \tag{6.105}$$

$$\mathcal{E}_3(t) = \{\forall K \leq t' \leq t : \hat{\Delta}^*_{t'} \leq \min_{i \neq i^*} \Delta(i)\}, \tag{6.106}$$

$$\mathcal{E}_4(t) = \{\forall i \in [K] \text{ and } \forall K \leq t' \leq t : \hat{\Delta}_{t'}(i) \leq \Delta(i)\}. \tag{6.107}$$

Now, we will show by induction that for all $i \neq i^*$ and $K \leq t \leq T$, if $\mathcal{E}_1(t)$, $\mathcal{E}_2(t)$, $\mathcal{E}_3(t)$ and $\mathcal{E}_4(t)$ occurs, then we have

$$N^s_t(i) \leq \frac{10000 \log T}{9\Delta^2(i)} + 1. \tag{6.108}$$

We have that (6.108) trivially holds for $t = K$. Also, if $\mathcal{E}_1(t)$, $\mathcal{E}_2(t)$, $\mathcal{E}_3(t)$ and $\mathcal{E}_4(t)$ occurs, then $\mathcal{E}_1(t-1)$, $\mathcal{E}_2(t-1)$, $\mathcal{E}_3(t-1)$ and $\mathcal{E}_4(t-1)$ occurs. Now, let (6.108) holds for $t-1$. If $i_t \neq i$, then (6.108) holds for t since $\log(.)$ is an increasing function. If $i_t = i$, then we have

$$\hat{\mu}_{t-1}(i) + \sqrt{\frac{400 \log T}{N^s_{t-1}(i)}} \geq \hat{\mu}_{t-1}(i^*) + \sqrt{\frac{400 \log T}{N^s_{t-1}(i^*)}}. \tag{6.109}$$

Under the events $\mathcal{E}_1(t)$, $\mathcal{E}_2(t)$, $\mathcal{E}_3(t)$ and $\mathcal{E}_4(t)$, this implies that

$$\mu_i + \frac{3}{2}\sqrt{\frac{400 \log T}{N^s_{t-1}(i)}} \geq \mu_{i^*} - \frac{2\Delta(i)}{20}, \tag{6.110}$$

or equivalently

$$N_{t-1}^s(i) \leq \frac{10000 \log T}{9\Delta^2(i)}. \tag{6.111}$$

This along with $N_t^s(i) \leq N_{t-1}^s(i) + 1$ implies (6.108).

Second, we will show that for all $f(T) \leq t \leq T$, we have

$$N_t^s(i^*) \geq \frac{1200 \log T}{\min_{i \neq i^*} \Delta^2(i)}. \tag{6.112}$$

We have that

$$\begin{aligned} N_t^s(i^*) &= t - \sum_{i \neq i^*} N_t^s(i) \\ &\overset{(a)}{\geq} f(T) - \sum_{i \neq i^*} N_t^s(i) \\ &\overset{(b)}{\geq} f(T) - \sum_{i \neq i^*} \frac{10000 \log T}{9\Delta^2(i)} - K + 1 \\ &\overset{(c)}{\geq} \frac{1200 \log T}{\min_{i \neq i^*} \Delta^2(i)}, \end{aligned} \tag{6.113}$$

where (a) follows from the fact that $f(T) \leq t$, (b) follows (6.108), and (c) follows from the definition of $f(T)$.

Third, we will show by induction that for all $i \neq i^*$ and $k \leq t \leq T$, we have

$$N_t^s(i) \geq \frac{1}{16} \min \left(\frac{40000 \log T}{121\Delta^2(i)}, \frac{4}{9}(N_t^s(i^*) - 1) \right). \tag{6.114}$$

Similar to (6.108), we only need to show that (6.114) holds if $i_t = i^*$. If $i_t = i^*$, then we have

$$\hat{\mu}_{t-1}(i^*) + \sqrt{\frac{400 \log T}{N_{t-1}^s(i^*)}} \geq \hat{\mu}_{t-1}(i) + \sqrt{\frac{400 \log T}{N_{t-1}^s(i)}}, \tag{6.115}$$

which implies that under the events $\mathcal{E}_1(t)$, $\mathcal{E}_2(t)$, $\mathcal{E}_3(t)$ and $\mathcal{E}_4(t)$,

$$\mu_{i^*} + \frac{3}{2}\sqrt{\frac{400\log T}{N^s_{t-1}(i^*)}} + 2\Delta(i)/20 \geq \mu_i + \frac{1}{2}\sqrt{\frac{400\log T}{N^s_{t-1}(i)}}. \tag{6.116}$$

Thus, we have

$$\begin{aligned} N^s_{t-1}(i) &\geq \frac{1}{4}\frac{400\log T}{\left(11\Delta(i)/10 + \frac{3}{2}\sqrt{\frac{400\log T}{N^s_{t-1}(i^*)}}\right)^2} \\ &\geq \frac{1}{4}\frac{400\log T}{\left(2\max\{11\Delta(i)/10, \frac{3}{2}\sqrt{\frac{400\log T}{N^s_{t-1}(i^*)}}\}\right)^2}. \end{aligned} \tag{6.117}$$

This along with $N^s_t(i) \leq N^s_{t-1}(i) + 1$ and $N^s_t(i^*) \leq N^s_{t-1}(i^*) + 1$ implies (6.114).

Now, combining (6.112) and (6.114), under the events $\mathcal{E}_1(t)$, $\mathcal{E}_2(t)$, $\mathcal{E}_3(t)$ and $\mathcal{E}_4(t)$, for all $f(T) \leq t \leq T$ and $i \neq i^*$, we have that

$$N^s_t(i) \geq \frac{2500\log T}{121\Delta^2(i)}. \tag{6.118}$$

Let $\mathcal{E}_1(t)$, $\mathcal{E}_2(t)$, $\mathcal{E}_3(t)$ and $\mathcal{E}_4(t)$ occurs. Then, for all $t \in \mathcal{T}$, we have

$$N^s_{t-1}(i_t) = N^s_t(i_t) > \frac{1200\log T}{\hat{\Delta}^{*2}_t} \geq \frac{1200\log T}{\min_{i\neq i^*}\Delta^2(i)}. \tag{6.119}$$

Using (6.119) and (6.108), under the events $\mathcal{E}_1(t)$, $\mathcal{E}_2(t)$, $\mathcal{E}_3(t)$ and $\mathcal{E}_4(t)$, for all $t \in \mathcal{T}$, we have that

$$i_t = i^*, \tag{6.120}$$

which implies

$$N^s_t(i^*) \geq \frac{1200\log T}{\min_{i\neq i^*}\Delta^2(i)}. \tag{6.121}$$

Also, using Hoeffding's inequality and Lemmas 24, 25 and 26, we have

$$\mathbb{P}(\bar{\mathcal{E}}_1(t)) + \mathbb{P}(\bar{\mathcal{E}}_2(t)) + \mathbb{P}(\bar{\mathcal{E}}_3(t)) + \mathbb{P}(\bar{\mathcal{E}}_4(t)) \leq \sum_{t'=K}^{t} \frac{K}{T^6} + \frac{6K^2}{T^6} + \frac{2K^2}{T^6} + \frac{K^2}{T^6}$$
$$\leq \frac{10K^2}{T^5}. \tag{6.122}$$

Combining (6.108), (6.112), (6.118), (6.120) and (6.122), the statement of the lemma follows.

□

Lemma 28. *For all $f(T) \leq t \leq T$, we have*

$$\mathbb{P}(\hat{\Delta}_t^* \leq 0.1 \min_{i \neq i^*} \Delta(i)) \leq 22K^2/T^5. \tag{6.123}$$

Proof. Let $\min_{i \neq i^*} \Delta(i) = \mu_{i^*} - \mu_{i_1}$. Now, consider the following events

$$\mathcal{E}_1(t) = \left\{ \forall i \in [K] : |\hat{\mu}_t^s(i) - \mu_i| \leq \sqrt{\frac{3 \log T}{N_t^s(i)}} \right\}, \tag{6.124}$$

$$\mathcal{E}_2(t) = \left\{ \forall i \in [K] \setminus i^* : \frac{2500 \log T}{121 \Delta^2(i)} \leq N_t^s(i) \right\}, \tag{6.125}$$

$$\mathcal{E}_3(t) = \left\{ \frac{1200 \log T}{\min_{i \neq i^*} \Delta^2(i)} \leq N_t^s(i^*) \right\}. \tag{6.126}$$

Under events $\mathcal{E}_1(t)$, $\mathcal{E}_2(t)$ and $\mathcal{E}_3(t)$, for all $i \neq i^*$ and $f(T) \leq t \leq T$, we have that

$$|\hat{\mu}_t^s(i) - \mu_i| \leq \sqrt{\frac{3 \log T}{N_t^s(i)}} < 0.4 \Delta(i), \tag{6.127}$$

and

$$|\hat{\mu}_t^s(i^*) - \mu_{i^*}| \leq \sqrt{\frac{3 \log T}{N_t^s(i^*)}} \leq \min_{i \neq i^*} 0.05 \Delta(i). \tag{6.128}$$

This implies that for all $i \in [K]$ and $t \geq t_*$, we have

$$|\hat{\mu}_t^s(i) - \mu_i| < \Delta(i)/2. \tag{6.129}$$

Using (6.129) and the definitions of a_t^* and $\tilde{a}_t$, we have that

$$a_t^* = i^*, \text{ and } \tilde{a}_t = i_1. \tag{6.130}$$

This implies that under events $\mathcal{E}_1(t)$, $\mathcal{E}_2(t^*)$ and $\mathcal{E}_3(t^*)$, we have

$$\begin{aligned}
\hat{\Delta}_t^* &\geq \hat{\mu}_t^s(a_t^*) - \sqrt{\frac{3\log T}{N_t^s(a_t^*)}} - \hat{\mu}_t^s(\tilde{a}_t) - \sqrt{\frac{3\log T}{N_t^s(\tilde{a}_t)}} \\
&\overset{(a)}{\geq} \mu_{i^*} - \mu_{i_1} - 2\sqrt{\frac{3\log T}{N_t^s(i^*)}} - 2\sqrt{\frac{3\log T}{N_t^s(i_1)}} \\
&\overset{(b)}{>} (\mu_{i^*} - \mu_{i_1})(1 - 0.8 - 0.1) \\
&\geq 0.1 \min_{i \neq i^*} \Delta(i),
\end{aligned} \tag{6.131}$$

where (a) follows from $\mathcal{E}_1(t)$, and (b) follows from (6.127) and (6.128). Thus, we have

$$\begin{aligned}
\mathbb{P}(\hat{\Delta}_t^* \leq 0.1 \min_{i \neq i^*} \Delta(i)) &\leq \mathbb{P}(\bar{\mathcal{E}}_1(t)) + \mathbb{P}(\bar{\mathcal{E}}_2(t)) + \mathbb{P}(\bar{\mathcal{E}}_3(t)) \\
&\leq \frac{2K}{T^6} + \frac{10K^2}{T^5} + \frac{10K^2}{T^5} \leq \frac{22K^2}{T^5},
\end{aligned} \tag{6.132}$$

where the last inequality follows from Hoeffding's inequality and Lemma 27.

□

Lemma 29. *For all $f(T) \leq t \leq T$, we have that*

$$\mathbb{P}\left(N_t^s(i^*) > \max\left\{ \frac{1200\log 2T}{(0.1\min_{i \neq i^*} \Delta(i))^2}, f(T) \right\} \right) \leq 31K^2/T^4. \tag{6.133}$$

Proof. Consider the following events

$$\mathcal{E}_1 = \{\forall t \in \mathcal{T} : i_t = i^*\}, \tag{6.134}$$

and

$$\mathcal{E}_2 = \{\forall t \geq f(T) : \hat{\Delta}_t^* \geq 0.1 \min_{i \neq i^*} \Delta(i)\}. \tag{6.135}$$

We will show that if $\mathcal{E}_1$ and $\mathcal{E}_2$ occurs, then for all $f(T) \leq t \leq T$, we have

$$N_t^s(i^*) \leq \max\left\{\frac{1200 \log 2T}{(0.1 \min_{i \neq i^*} \Delta(i))^2}, f(T)\right\}. \tag{6.136}$$

For all $t \in \mathcal{T}$, let $\ell(t) = \max\{f(T) < t_1 \leq t : i_{t_1} = i^* \text{ and } t_1 \notin \mathcal{T}\}$ be the latest time instance before t and after $f(T)$ where verification is performed for i^*. If $\ell(t)$ does not exists, then (6.136) follows trivially. Then, under the events $\mathcal{E}_1$ and $\mathcal{E}_2$, for all $t \in \mathcal{T}$, we have that

$$\begin{aligned} N_t^s(i^*) &= N_{\ell(t)-1}^s(i^*) + 1 \\ &\overset{(a)}{\leq} \frac{1200 \log T}{\hat{\Delta}_{\ell(t)-1}^{*2}} + 1 \\ &\overset{(b)}{\leq} \frac{1200 \log 2T}{\hat{\Delta}_{\ell(t)-1}^{*2}} \\ &\overset{(c)}{\leq} \frac{1200 \log 2T}{(0.1 \min_{i \neq i^*} \Delta(i))^2}, \end{aligned} \tag{6.137}$$

where (a) follows from the fact that verification is performed for i^* at round $\ell(t)$, (b) follows from the facts that $1 \leq 800 \log 2$ and $\hat{\Delta}_t^{*2} \leq 1$, and (c) follows from $\mathcal{E}_2$.

Now, using Lemma 27 and Lemma 28, we have

$$\mathbb{P}(\bar{\mathcal{E}}_1) + \mathbb{P}(\bar{\mathcal{E}}_2) \leq \sum_{t=K}^{T} \frac{10K^2}{T^5} + \frac{22K^2}{T^5} \leq \frac{32K^2}{T^4}. \tag{6.138}$$

□

Proof of Theorem 24

Proof. Let $t^* = \max\{t \in \mathcal{T}\}$. Then, we have

$$\begin{aligned}
&\sum_{i\in[K]} \mathbb{E}[N_T^s(i)] \\
&= \sum_{i\in[K]} \mathbb{E}[N_{t^*}^s(i)] \\
&\overset{(a)}{\leq} \max\left\{f(T), \frac{1200\log 2T}{(0.1\min_{i\neq i^*}\Delta(i))^2}\right\} + 2K + \sum_{i\neq i^*}\frac{10000\log T}{9\Delta^2(i)} + \left(\frac{10K^2}{T^5} + \frac{32K^2}{T^4}\right)t^* \\
&\leq \max\left\{f(T), \frac{1200\log 2T}{(0.1\min_{i\neq i^*}\Delta(i))^2}\right\} + 2K + \sum_{i\neq i^*}\frac{10000\log T}{9\Delta^2(i)} + \frac{42K^2}{T^3} \\
&\leq \max\left\{f(T), \frac{1200\log 2T}{(0.1\min_{i\neq i^*}\Delta(i))^2}\right\} + 2K + \sum_{i\neq i^*}\frac{10000\log T}{9\Delta^2(i)} + \frac{42}{K}, \qquad (6.139)
\end{aligned}$$

where (a) follows from Lemma 27 and Lemma 29.Using the definition of $f(T)$ in (6.99), we have that (6.22) follows, namely

$$\sum_{i\in[K]} \mathbb{E}[N_T^s(i)] \leq c_3\left(\sum_{i\neq i^*}\log T/\Delta^2(i)\right), \qquad (6.140)$$

Consider the event

$$\mathcal{E} = \{\forall t \in \mathcal{T} \text{ such that } t > K : i_t = i^*\}. \qquad (6.141)$$

Under event $\mathcal{E}$, for all $i \in [K] \setminus i^*$ and $t \leq T$, we have that

$$N_t^s(i) = N_t(i). \qquad (6.142)$$

We have

$$\begin{aligned}\mathbb{E}[N_T(i)|\mathcal{E}] &= \mathbb{E}[N_T^s(i)|\mathcal{E}] \\ &= \mathbb{E}[N_{t^*}^s(i)|\mathcal{E}] \\ &\overset{(a)}{\leq} \frac{10000\log T}{9\Delta^2(i)} + \frac{10K^2t^*}{T^5} + 1 \\ &\overset{(b)}{\leq} \frac{10000\log T}{9\Delta^2(i)} + \frac{10K^2}{T^4} + 1 \\ &\overset{(c)}{\leq} \frac{10000\log T}{9\Delta^2(i)} + \frac{10}{K^2} + 1, \end{aligned} \tag{6.143}$$

where (a) follows from Lemma 27, (b) follows from the fact that $t^* \leq T$, and (c) follows from the fact that $k \leq T$.

Also, using Lemma 27, we have

$$\mathbb{P}(\bar{\mathcal{E}}) \leq \sum_{t=K}^{T} \frac{10K^2}{T^5} \leq \frac{10K^2}{T^4}. \tag{6.144}$$

Now, combining (6.143) and (6.144), for all $i \in [K] \setminus i^*$, we have that

$$\mathbb{E}[N_T(i)] = \frac{10000\log T}{9\Delta^2(i)} + \frac{10}{K^2} + 1 + \frac{10K^2}{T^4}. \tag{6.145}$$

This implies that the regret of the algorithm is

$$\begin{aligned} R^{SUCB}(T) &= \sum_{i\neq i^*} \Delta(i)\mathbb{E}[N_T(i)] \\ &= \sum_{i\neq i^*} \left(\frac{10000\log T}{9\Delta(i)} + \frac{10\Delta(i)}{K^2} + \frac{10K^2\Delta(i)}{T^4} + \Delta(i) \right). \end{aligned} \tag{6.146}$$

□

Hence, we have that (6.23) follows.

6.11.6 Proof of Theorem 25

Proof. Consider the following specific attacker strategy that for all t, the attacker always set $r_t^o(i_t) = 0$. This implies that if verification is not performed by the algorithm at any round t, then for all $i_1, i_2 \in [K]$, we have that

$$KL(i_1, i_2) = 0. \tag{6.147}$$

Combining (6.147) and Theorem 12 in [106], there exists a constant t^* such that for all $t \geq t^*$, we have

$$\mathbb{P}(i_t \neq i^*) \geq \exp\left(-\min_{i_1,i_2\in[K]} KL(i_1, i_2)N_t^s\right), \tag{6.148}$$

where N_t^s is the total number of verifications performed by the learner until round t.[4]

Now, divide the interval $[T/2, T]$ into $2(\log T)^{1-\alpha}/\min_{i_1,i_2\in[K]} KL(i_1, i_2)$ equal sized intervals. This implies that $T\min_{i_1,i_2\in[K]} KL(i_1, i_2)/(4(\log T)^{1-\alpha})$ is the size of each interval. Using the Pigeonhole principle, we have that there exists at least $(\log T)^{1-\alpha}/\min_{i_1,i_2\in[K]} KL(i_1, i_2)$ intervals such that no verification is performed during these intervals. Let $I = [t_1, t_1 + T\min_{i_1,i_2\in[K]} KL(i_1, i_2)/(4(\log T)^{1-\alpha})]$ denote an interval where no verification is performed. Thus, for all $t \in I$, we have

$$\begin{aligned}\mathbb{P}(i_t \neq i^*) &\overset{(a)}{\geq} \exp\left(-\min_{i_1,i_2\in[K]} KL(i_1, i_2)N_{t_1}^s\right) \\ &\overset{(b)}{\geq} \exp\left(-(\log T)^{1-\alpha}\right),\end{aligned} \tag{6.149}$$

where (a) follows from (6.148) and the fact that $t \geq t_1 \geq T/2 \geq t^*$, and (b) follows from (6.24)

[4] In [106], N_t^s is the number of observed true rewards collected so far. In our model, since the algorithm only gets 0 when the reward is not verified, regardless of which action is selected. Such uninformative reward feedback will not make a difference, and the number of observed true rewards in our case is thus precisely N_t^s.

and the fact that $N_{t_1}^s \leq N_T^s$. Using (6.149), the total expected regret in the interval I is at least

$$\begin{aligned}
&\min_{i \neq i^*} \Delta(i) \sum_{t \in I} \mathbb{P}(i_t \neq i^*) \\
&\overset{(a)}{\geq} \min_{i \neq i^*} \Delta(i) \frac{T \min_{i_1, i_2 \in [K]} KL(i_1, i_2)}{4(\log T)^{1-\alpha}} \exp\left(-(\log T)^{1-\alpha}\right),
\end{aligned} \tag{6.150}$$

where (a) follows from the fact that the size of I is $T \min_{i_1, i_2 \in [K]} KL(i_1, i_2)/(4(\log T)^{1-\alpha})$. Since the number of intervals with no verification is $(\log T)^{1-\alpha}/\min_{i_1, i_2 \in [K]} KL(i_1, i_2)$, the regret of the algorithm is at least

$$\begin{aligned}
&\min_{i \neq i^*} \Delta(i) \sum_{I} \sum_{t \in I} \mathbb{P}(i_t \neq i^*) \\
&\geq \min_{i \neq i^*} \Delta(i) \frac{T}{4} \exp\left(-(\log T)^{\alpha}\right) \\
&\overset{(a)}{\geq} \min_{i \neq i^*} \Delta(i) (\log T)^{\beta},
\end{aligned} \tag{6.151}$$

where (a) follows from the fact that $(\log T)^{\alpha} + \beta \log(4 \log T) \leq \log T$. The statement of the theorem follows. □

6.11.7 Proof of Corollary 25.1

Proof. Let $\delta = (\log T)^{-\alpha/2}$. Using Markov's inequality, and the fact that the expected number of verification is at most $(\log T)^{1-\alpha}/\min_{i_1, i_2 \in [K]} KL(i_1, i_2)$, we have

$$\mathbb{P}(N_T^s \geq (\log T)^{1-\alpha/2}) \leq \frac{(\log T)^{1-\alpha}}{(\log T)^{1-\alpha/2}} = \delta. \tag{6.152}$$

Using Theorem 25 and (6.152), we have that

$$\begin{aligned}
R^{UCB}(T) &\geq c_3(1-\delta)(\log T)^{\beta} \\
&\overset{(a)}{\geq} (1 - (\log 2)^{-\alpha/2})(\log T)^{\beta} \\
&= c_4 (\log T)^{\beta},
\end{aligned} \tag{6.153}$$

where c_4 is a numerical constant, and (a) follows from the fact that $\log T$ is an increasing function of T. The statement of the corollary follows. □

Chapter 7

Attacks in Episodic Reinforcement Learning

7.1 Introduction

Learning algorithms are often used in web services [245, 2, 129], conversational AI [50], sensor networks [212], medical trials [21], and crowdsourcing systems [174]. The distributed nature of these applications makes these algorithms prone to third party attacks. For example, in web services decision making critically depends on reward collection, and this is prone to attacks that can impact observations and monitoring, delay or temper rewards, produce link failures, and generally modify or delete information through hijacking of communication links [2, 37]. Making these systems secure requires an understanding of the regime where the systems may be vulnerable, as well as designing ways to mitigate these attacks. The present paper focuses on the former aspect, namely understanding of the regime where the systems can be attacked, in an episodic Reinforcement Learning (RL) setting.

We consider poisoning attack, also referred as *man in the middle* (MITM) attack. In this attack, there are three agents: the environment, the learner (RL algorithm), and the attacker. The learner interacts with the environment for T episodes, and each episode has H steps. In episode $t \leq T$ at step $h \leq H$, the learner observes the state $s_t(h) \in \mathcal{S}$ of the environment, selects an action $a_t(h)$ among $\mathcal{A}$ choices, the environment then generates a reward $r_t(s_t(h), a_t(h))$ and changes its state based on an underlying Markov Decision Process (MDP), and attempts to communicate

it to the learner. However, an adversary acts as a "man in the middle" between the learner and the environment. It can observe and may *manipulate the action* $a_t(h)$ to $a_t^o(h) \in \mathcal{A}$ which will generate reward $r_t(s_t(h), a_t^o(h))$ corresponding to the manipulated action. The adversary may also intercept the reward $r_t(s_t(h), a_t^o(h))$ by adding contamination noise $\epsilon_{t,h}(s_t(h), a_t(h))$. In this case, the learner only observes the contaminated reward $r_t^o(s_t(h), a_t(h)) = r_t(s_t(h), a_t^o(h)) + \epsilon_{t,h}(s_t(h), a_t(h))$. We study the *amount of contamination* $\sum_{t,h} |\epsilon_{t,h}(s_t(h), a_t(h))|$ and the *number of action manipulations* $\sum_{t,h} \mathbf{1}(a_t(h) \neq a_t^o(h))$.

Reward poisoning attack is a special case of the MITM attack where $a_t^o(h) = a_t(h)$, and has been studied previously in both RL and Multi-Armed Bandits (MAB) settings [97, 171, 172]. Likewise, action manipulation attack is a special case of the MITM attack where $\epsilon_{t,h}(s_t(h), a_t(h)) = 0$, and has been previously studied for MAB setting [137]. Another variant of action manipulation attack, previously studied in RL [172], is manipulation of the transition dynamics. This can be equivalently considered as manipulating the action $a_t(h)$ to another action, not necessarily in $\mathcal{A}$. MITM attacks has also been previously considered in cyber-physical systems [112, 182]. Given that RL algorithms are increasingly used in critical applications, including cyber-physical systems [128], it is of utmost importance to investigate the security threat to RL algorithms against different forms of poisoning attacks.

We consider MITM attacks in two different settings: *unbounded* rewards and *bounded* rewards, which turns out to differ fundamentally. In unbounded reward setting, the contamination $\epsilon_{t,h}(s_t(h), a_t(h))$ is unconstrained whereas in bounded reward setting, the contaminated reward $r_t^o(s_t(h), a_t^o(h))$ is constrained to be in the interval $[0, 1]$, just like the original rewards $r_t(s_t(h), a_t^o(h))$. This constrained situation limits the attacker's contamination at every round, and turns out to be provably more difficult to attack. In each setting, we shall start as a warm-up with the so-called "white-box" attacks in which the attacker is assumed to posses full knowledge of the underlying MDP. We then show how to adapt such white-box attacks to the more realistic black-box setting in which the attacker does not know, and needs to learn, the underlying MDP as well.

Table 7.1. Comparison of the attack cost in the episodic RL and MAB setting when the attacker has no information about the learning algorithm.

Settings	Reward	Attack	Upper Bound
White-box in RL	Unbounded	Reward Manipulation	$O(\sqrt{T})$ [172]
White-box in RL	Unbounded	Dynamics Manipulation (under sufficient conditions only)	$O(\sqrt{T})$ [172]
Black-box in RL	Unbounded	Reward Manipulation	$O(\sqrt{T})$ (This work)
Black-box and White Box in RL	Bounded	Reward Manipulation	Infeasible (This work)
Black-box and White Box in RL	Bounded	Action Manipulation	Infeasible (This work)
Black-box and White Box in RL	Bounded	Reward and Action Manipulation	$O(\sqrt{T})$ (This work)

7.1.1 Contribution

We consider poisoning attacks with the objective of forcing the learner to execute a target policy π^+. More specifically, for all $h \leq H$ and $s \in \mathcal{S}$, if $\pi_h(s) \neq \pi_h^+(s)$, then the attack aims to induce values satisfying

$$V_h^{\pi}(s) < V_h^{\pi^+}(s), \tag{7.1}$$

where policy π of an agent is a collection of H functions $\{\pi_h : \mathcal{S} \to \mathcal{A}\}$, and value function $V_h^{\pi}(s)$ is the expected reward under policy π, starting from state s at step h, until the end of the episode.

Reward manipulation attack are studied in [172] for RL, and a white-box attack is proposed in unbounded reward setting. This *white-box* attack is order optimal in the amount of contamination. Extending this research agenda, we propose *black-box* attack for episodic RL, and show that the proposed attack can attack any no-regret learning algorithm in $\tilde{O}(\sqrt{T})$ amount of contamination, which is order-optimal upto logarithmic factor.

Similar to previous work such as [172], we study the feasibility of attack in bounded reward setting under the constraints that the reward manipulation, namely $\epsilon_{t,h}(s_t(h), a_t(h)) \neq 0$, and action manipulation, namely $a_t(h) \neq a_t^o(h)$, can occur only if the selected action is different

from the desired action, namely $a_t(h) \neq \pi_h^+(s_t(h))$. We show that the reward manipulation attack does not suffice, namely there exist an MDP and a target policy π^+ which cannot be attacked by the attacker, namely (7.1) cannot be achieved, by manipulating the rewards. Similarly, we show that the action manipulation attack does not suffice as well, namely there exist an MDP and a target policy π^+ which cannot be attacked by the attacker, namely (7.1) cannot be achieved, by manipulating the actions. Hence, the attacker needs a combined power of reward manipulation and action manipulation attack. Indeed, we propose an MITM attack in bounded reward setting, which requires $O(\sqrt{T})$ amount of reward contamination and $O(\sqrt{T})$ number of action manipulations to attack any no-regret learning algorithm.

An interesting conceptual message from our results is that bounded reward setting appears more difficult to attack than the unbounded reward setting. We also summarize our results in Table 7.1 by comparing them with the relevant literature.

7.1.2 Related Work

In online bandit learning, reward manipulation attack has been studied extensively in Multi-Armed Bandits [97, 135, 253], where the attacker's objective is to mislead the learner to choose a suboptimal action. Additionally, the action manipulation attack has also been studied in Multi-Armed Bandits [137] where the attacker can manipulate (or override) the action of the learner, and the number of action manipulations required by the attacker is $O(\log T)$. All these attacks are studied in a Black-box setting, where the attacker does not posses any knowledge about the underlying reward distributions.

In online RL setting, studies related to poisoning attacks have only started recently, and have primarily focused on white-box settings, where the attacker has complete knowledge of the underlying MDP models, with unbounded rewards [172, 171]. In such white-box attacks, [172] show that reward poisoning attack requires $\Theta(\sqrt{T})$ amount of contamination to attack any no-regret learning algorithm; they also show that dynamic manipulation attack can achieve the same success with similar amount of cost in unbounded reward setting under some sufficient

conditions. [242] study the feasibility of the reward poisoning attack in white box setting for Q-learning, and the attacker is constrained by the amount of contamination. In a slightly different thread, [92] analyse the degradation of the performance of Temporal difference learning and Q-learning under falsified rewards.

To our knowledge, [173] is the only work that studies poisoning attack in black-box setting for policy teaching in RL. However, they focused on the settings with L online learners, and the objective of their attacker is to force all these learners to execute a target policy π^+. They proposed an attack with $\tilde{O}(T \log L + L\sqrt{T})$ amount of contamination when L is large enough. However, our work focuses on attacking a *single* learner and thus our setting is not comparable to [173]. However, we can indeed apply our attack repeatedly to different learners to obtain an effective attack strategy for the setup of [173], which leads to an attack cost of $\tilde{O}(L\sqrt{T})$ (note however, the attack of [173] cannot work for small L, e.g., $L = 1$ as in our setup). This improves their attack cost by an additive amount $O(T \log L)$. Our more efficient attack is due to a more efficient design for the adversary to explore and learn the MDP.

Test-time adversarial attacks against reinforcement learning (RL) has also been studied. Here, however, the policy π of the RL agent is pre-trained and fixed, and the objective of the attacker is to manipulate the perceived state of the RL agent in order to induce undesired action [91, 133, 122, 22]. Such test-time attacks do not modify the the policy π, whereas training-time attacks we study in this paper aims at poisoning the learned policy directly and thus may have a longer-term bad effects. There have also been studies on reward poisoning against *Batch RL* [142, 241] where the attacker can modify the pre-collected batch data set at once. The focus of the present work is on *online* attack where the poisoning is done on the fly.

7.2 Problem Formulation

We consider the setting of episodic Markov Decision Process (MDP) $(\mathcal{S}, \mathcal{A}, H, \mathcal{P}, \mu)$, where $\mathcal{S}$ is the set of states with $|\mathcal{S}| = S$, $\mathcal{A}$ is the set of actions with $|\mathcal{A}| = A$, H is the number of

steps in each episode, $\mathcal{P}$ is the transition metric such that $\mathbb{P}(.|s,a)$ gives the transition distribution over the next state if action a is taken in the current state s, and $\mu : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is the expected reward of state action pair (s,a).

An RL agent (or learner) interacts with the MDP for T episodes, and each episode consists of H steps. In each episode of the MDP, an initial state $s_t(1)$ can be fixed or selected from any distribution. In episode t and step h, the learner observes the current state $s_t(h) \in \mathcal{S}$, selects an action $a_t(h) \in \mathcal{A}$, and incurs a noisy reward $r_{t,h}(s_t(h), a_t(h))$. Additionally, we have that $\mathbb{E}[r_{t,h}(s_t(h), a_t(h))] = \mu(s_t(h), a_t(h))$. Finally, both $\mathcal{P}$ and μ are unknown to the agent.

We consider episodic RL under MITM attacks. The attacker can manipulate the action $a_t(h)$ selected by the learner to another action $a_t^o(h) \in \mathcal{A}$. The MDP thus undergoes transition to next state based on the action $a_t^o(h)$, namely the next state is drawn from the distribution $\mathbb{P}(.|s_t(h), a_t^o(h))$. The reward observation $r_t(s_t(h), a_t^o(h))$ is generated. If $a_t^o(h) \neq a_t(h)$, then the episode t and step h is said to be *under action manipulation attack*. Hence, the *number of action manipulations* is $\sum_{t=1}^{T} \sum_{h=1}^{H} \mathbf{1}(a_t^o(h) \neq a_t(h))$.

The adversary can intercept the reward observation $r_t(s_t(h), a_t^o(h))$ and contaminate it by adding noise $\epsilon_{t,h}(s_t(h), a_t(h))$. Learner observes reward $r_t^o(s_t(h), a_t(h))$, where

$$r_{t,h}^o(s_t(h), a_t(h)) = r_{t,h}(s_t(h), a_t^o(h)) + \epsilon_{t,h}(s_t(h), a_t(h)), \tag{7.2}$$

where the contamination $\epsilon_t(s_t(h), a_t(h))$ added by the attacker can be a function of all the states visited previously, and the actions selected previously by the learner and the attacker. If $\epsilon_{t,h}(s_t(h), a_t(h)) \neq 0$, then the episode t and step h is said to be *under reward manipulation attack*. Hence, the *number of reward manipulations* is $\sum_{t=1}^{T} \sum_{h=1}^{H} \mathbf{1}(\epsilon_{t,h}(s_t(h), a_t(h)) \neq 0)$, and the *amount of contamination* is $\sum_{t=1}^{T} \sum_{h=1}^{H} |\epsilon_{t,h}(s_t(h), a_t(h))|$.

Reward poisoning attack is a special case of MITM attack, where the adversary cannot manipulate the action which implies $a_t(h) = a_t^o(h)$. *Action manipulation attack* is a special case of MITM attack, where the adversary cannot contaminate the reward observation which implies

$\epsilon_{t,h}(s_t(h), a_t(h)) = 0$.

A (deterministic) policy π of an agent is a collection of H functions $\{\pi_h : \mathcal{S} \to \mathcal{A}\}$. The value function $V_h^\pi(s)$ is the expected reward under policy π, starting from state s at step h, until the end of the episode, namely

$$V_h^\pi(s) = \mathbb{E}[\sum_{h'=h}^{H} \mu(s_{h'}, \pi_{h'}(s_{h'}))|s_h = s], \tag{7.3}$$

where $s_{h'}$ denotes the state at step h' of the episode. Likewise, the Q-value function $Q_h^\pi(s, a)$ is the expected reward under policy π, starting from state s and action a, until the end of the episode, namely

$$Q_h^\pi(s, a) = \mu(s, a) + \mathbb{E}[\sum_{h'=h+1}^{H} \mu(s_{h'}, \pi_{h'}(s_{h'}))|s_h = s, a_h = a], \tag{7.4}$$

where $a_{h'}$ denotes the action at step h' of the episode. Since S, A and H are finite, there exists an optimal policy π^* such that $V_h^{\pi^*}(s) = \sup_\pi V_h^\pi(s)$.

The regret $R^{\mathcal{A}}(T, H)$ of any algorithm $\mathcal{A}$ is the difference between the total expected true reward from the best fixed policy π^* in the hindsight, and the expected true reward over T episodes, namely

$$R^{\mathcal{A}}(T, H) = \sum_{t=1}^{T} \big(V_1^{\pi^*}(s_t(1)) - V_1^{\pi_t}(s_t(1))\big). \tag{7.5}$$

The objective of the learner is to minimize the regret $R^{\mathcal{A}}(T, H)$. In contrast, the objective of the attacker is to poison the environment with an objective of teaching/forcing the learner to execute a target policy π^+ at least $\Omega(T)$ times, more specifically for all $h \leq H$ and $s \in \mathcal{S}$, if $\pi_h(s) \neq \pi_h^+(s)$, then

$$V_h^\pi(s) < V_h^{\pi^+}(s). \tag{7.6}$$

7.3 Reward Poisoning Attacks in Unbounded Reward Setting

In this section, we focus on unbounded reward setting, namely for all $(s, a) \in \mathcal{S} \times \mathcal{A}$, we have $r_{t,h}(s, a)$ follows sub-gaussian distribution with mean $\mu(s, a)$ and standard deviation σ. We show that the attacker can achieve its objective using reward manipulation only. In other words, action manipulation is not needed. Additionally, the amount of contamination needed is $\tilde{O}(\sqrt{T})$.

7.3.1 White-Box Attacks:a Warm-up

In white-box attack setting, the attacker posses the knowledge about the expected reward and the transition dynamics of the MDP. In this section, we propose a whitebox attack which utilizes this information about the MDP, and achieves an order optimal attack cost. This attack is different from the white-box attack proposed in [172], and is adapted to the white-box setting in episodic RL.

Given the target policy π^+ and an input parameter $\epsilon > 0$, for all $s_t(h) \in \mathcal{S}$, $a_t(h) \in \mathcal{A}$ and $h \leq H$, our proposed attack strategy is

$$r_t^o(s_t(h), a_t(h)) = \begin{cases} r_t(s_t(h), a_t(h)) & \text{if } a_t(h) = \pi_h^+(s_t(h)), \\ \tilde{Q}_h^{\pi^+}(s_t(h), \pi_h^+(s_t(h))) & otherwise \\ -\mathbb{E}_{s' \sim P(s'|s_t(h), a_t(h))}[\tilde{V}_{h+1}^{\pi^+}(s')] - \epsilon \end{cases} \tag{7.7}$$

where $\tilde{Q}_h^{\pi}(s, a)$ is the expected reward in state s for action a for the above reward observation under policy π, and $\tilde{V}_h^{\pi}(s)$ is the expected reward in state s for the above reward observation under policy π. These values will not be same as the ones defined in (7.3) and (7.4) since the reward observations are manipulated. We remark that the $r_t^o(s_t(h), a_t(h))$ can be computed through a backward induction procedure starting from horizon H. At any step h in the episode, the definition of $r_t^o(s_t(h), a_t(h))$ depends linearly on the Q-values at h, which then depends

linearly on $r_t^o(s_t(h), a_t(h))$. Therefore, $r_t^o(s_t(h), a_t(h))$ at any horizon h can be computed by solving a linear system.

During the attack in (7.7), the reward observations are manipulated only if the action selected by the learner is not the same as the desired action by π^+. The above reward manipulation strategy ensures that the target policy π^+ is the optimal policy based on the observed reward observations, namely for all $h \leq H$, and $(s, a) \in \mathcal{S} \times \mathcal{A}$ such that $a \neq \pi_h^+(s)$, we have

$$\tilde{Q}_h(s, a) \leq \tilde{Q}_h(s, \pi_h^+(s)) - \epsilon. \tag{7.8}$$

This implies that the parameter ϵ in the above attack can be tuned to obtain a desired difference between the expected rewards of the optimal policy and any other policy. This requirement been studied in form of ϵ-robust policy in [172].

Following theorem provides an upper bound on the amount of contamination for an order-optimal learning algorithm.

Theorem 27. *For any learning algorithm whose regret in the absence of attack is given by*

$$R^{\mathcal{A}}(T, H) = O(\sqrt{T} H^{\alpha}), \tag{7.9}$$

with probability at least $1 - \delta$*, where* $\alpha \geq 1$ *is a numerical constant; and for any sub-optimal target policy* π^+ *and* $\epsilon > 0$*, if an attacker follows strategy* (7.7)*, then with probability at least* $1 - \delta$*, the number of reward manipulation attacks will be*

$$\sum_{t=1}^{T} \sum_{h=1}^{H} \mathbf{1}(\epsilon_{t,h}(s_t(h), a_t(h)) \neq 0) = O\big(\sqrt{T} H^{\alpha}/\epsilon\big), \tag{7.10}$$

the amount of contamination

$$\sum_{t=1}^{T} \sum_{h=1}^{H} |\epsilon_{t,h}(s_t(h), a_t(h))| = \tilde{O}\big(\sqrt{T} H^{\alpha+1} + \sqrt{T} H^{\alpha+1}/\epsilon\big), \tag{7.11}$$

$\sum_{t=1}^{T}\sum_{h=1}^{H}\mathbf{1}(a_t(h)=\pi_h^+(s_t(h)))=\Omega(T)$*, and attacker achieves its objective in* (7.6)*.*

Similar to the white-box attack in [172], Theorem 27 shows that the attacker can achieve its objective in $O(\sqrt{T})$ attack cost in episodic RL. The attack cost is of the same order as the lower bound in [172, Theorem 1].

7.3.2 Black-Box Attack: the more realistic setting

In the black box attack setting, similar to the learner, the attacker does not know the expected reward and the transition dynamics of the MDP. Extending the white-box attack in (7.7), we propose a black-box attack which learns about the MDP, and has almost the same attack cost as the white-box attack, with an additional $O(\sqrt{\log T})$ factor.

Given the target policy π^+ and the input parameter $\epsilon > 0$, our proposed attack strategy is presented in Algorithm 13. Unlike the white box attack, the attack in the black box setting evolves in two phases: initialization phase and exploitation phase. In the initialization phase, the objective of the attacker is to obtain at least one observation of the reward from the environment corresponding to every $(s,a) \in \mathcal{S}\times\mathcal{A}$ pair. This helps in initializing the estimate $\hat{\mu}(s,a)$ of the true mean $\mu(s,a)$ for each $(s,a)\in\mathcal{S}\times\mathcal{A}$ pair. In Algorithm 13, the attacker achieves this by contaminating the observed reward $r_t^o(s,a) \sim \text{Bern}(0.5)$ for each (s,a) pair, where Bern(p) denotes the Bernoulli distribution with mean p. This contamination ensures that the expected reward of each (s,a) pair selected by the the learner appears identical, and thus promotes the exploration of all the (s,a) pairs.

The initialization phase stops if the reward corresponding to each (s,a) pair is observed at least once, namely $N(s,a)\geq 1, \forall (s,a)\in\mathcal{S}\times\mathcal{A}$ in Algorithm 13, where $N(s,a)$ denotes the number of times the (s,a) pair is selected. The attacker then turns to the exploitation phase, and utilizes its estimate of $\hat{\mu}(s,a)$, $\hat{\mu}^{UCB}(s,a)$ and $\hat{\mu}^{LCB}(s,a)$, defined in (7.12), (7.13) and (7.14) respectively, to contaminate the reward observations. These estimates are initialized in the initialization phase, and are updated in each episode $t\leq T$ and at each step $h\leq H$. The parameters $\hat{\mu}^{UCB}(s,a)$, and $\hat{\mu}^{LCB}(s,a)$ are Upper Confidence Bound (UCB) and Lower

Algorithm 13. Black box attack strategy

For all $(s,a,h) \in \mathcal{S} \times \mathcal{A} \times [H]$, $\hat{\mu}(s,a) = 0$, $\hat{\mu}^{UCB}(s,a) = 0$, $\hat{\mu}^{LCB}(s,a) = 0$ and $N(s,a) = 0$.
for episode $t \leq T$ **do**
  Observe the initial state $s_t(1)$
  for $h \leq H$ **do**
    Observe the selected action $a_t(h)$, reward $r(s_t(h), a_t(h))$ and next state $s_t(h+1)$.
    Update $N(s_t(h), a_t(h)) = N(s_t(h), a_t(h)) + 1$

$$\hat{\mu}(s_t(h), a_t(h)) = \frac{\hat{\mu}(s_t(h), a_t(h))(N(s_t(h), a_t(h)) - 1) + r(s_t(h), a_t(h))}{N(s_t(h), a_t(h))} \tag{7.12}$$

$$\hat{\mu}^{UCB}(s_t(h), a_t(h)) = \hat{\mu}(s_t(h), a_t(h)) + \sigma\sqrt{4\log(2THSA)/N(s_t(h), a_t(h))}, \tag{7.13}$$

$$\hat{\mu}^{LCB}(s_t(h), a_t(h)) = \hat{\mu}(s_t(h), a_t(h)) - \sigma\sqrt{4\log(2THSA)/N(s_t(h), a_t(h))}, \tag{7.14}$$

    if $\exists (s,a) \in \mathcal{S} \times \mathcal{A}$, such that $N(s,a) = 0$ **then**
      Contaminate the reward observation such that $r_t^o(s_t(h), a_t(h)) = \text{Bern}(1/2)$.
    else
      if $a_t(h) = \pi_h^+(s_t(h))$ **then**
        Do not contaminate, namely $r_t^o(s_t(h), a_t(h)) = r_t(s_t(h), a_t(h))$.
      else
        Contaminate the reward observation such that

$$r_t^o(s_t(h), a_t(h)) = \hat{\mu}^{LCB}(s_t(h), \pi_h^+(s_t(h))) + (H-h) \min_{s,a \in \mathcal{S} \times \mathcal{A}} \hat{\mu}^{LCB}(s,a) - (H-h) \max_{s,a \in \mathcal{S} \times \mathcal{A}} \hat{\mu}^{UCB}(s,a) - \epsilon. \tag{7.15}$$

      end if
    end if
  end for
end for

Confidence Bound (LCB) of $\mu(s,a)$. Therefore, as we will show, with high probability we have

$$\hat{\mu}^{LCB}(s,a) \leq \mu(s,a) \leq \hat{\mu}^{UCB}(s,a). \tag{7.16}$$

In this phase, the reward observations are contaminated only if the action selected by the learner is not the same as the action desired by the target policy, namely $a_t(h) \neq \pi_h^+(s_t(h))$. In this scenario, the reward observation $r_t^o(s_t(h), a_t(h))$ is defined in (7.15). This reward contamination is motivated from the white box attack (7.7). Comparing (7.15) and (7.7), we show that with

high probability

$$\hat{\mu}^{LCB}(s_t(h), \pi_h^+(s_t(h))) + (H-h) \min_{s,a \in \mathcal{S} \times \mathcal{A}} \hat{\mu}^{LCB}(s,a) \leq \tilde{Q}_h^{\pi^+}(s_t(h), \pi_h^+(s_t(h))). \quad (7.17)$$

Additionally, we have that with high probability

$$(H-h) \max_{s,a \in \mathcal{S} \times \mathcal{A}} \hat{\mu}^{UCB}(s,a) \geq \mathbb{E}_{s' \sim \mathbb{P}(s'|s_t(h), a_t(h))}[\tilde{V}_{h+1}^{\pi^+}(s')]. \quad (7.18)$$

Combining (7.17) and (7.18), we have that with high probability, the rewards contamination in the black box attack, given by (7.15), is greater than the reward contamination in the white box attack, given by (7.7). This would imply that the proposed black box attack is successful since the white-box attack was successful.

In the following theorem, similar to the white box attack, our proposed black box attack has almost the same amount of contamination, upto logarithmic factor, as white box attack.

Theorem 28. *Consider any learning algorithm $\mathcal{A}$ such that its regret in the absence of attack is*

$$R^{\mathcal{A}}(T,H) = O(\sqrt{T}H^{\alpha}), \quad (7.19)$$

with probability at least $1-\delta$ for any $T \geq t_0$, where $\alpha \geq 1$ is a numerical constant; and for any sub-optimal target policy π^+, $\epsilon > 0$ and $T \geq t_0^2$, if an attacker follows strategy in Algorithm 13, then with probability at least $1-\delta-2/(HSAT)$, the number of reward manipulation will be

$$\sum_{t=1}^{T} \sum_{h=1}^{H} \mathbf{1}(\epsilon_{t,h}(s_t(h), a_t(h)) \neq 0) = O\big(\sqrt{T}H^{\alpha}/\epsilon\big), \quad (7.20)$$

the amount of contamination is

$$\sum_{t=1}^{T} \sum_{h=1}^{H} |\epsilon_{t,h}(s_t(h), a_t(h)| = O\big(\sqrt{T}H^{\alpha+1}(\epsilon + \sqrt{\log(HTSA)})/\epsilon\big), \quad (7.21)$$

$\sum_{t=1}^{T}\sum_{h=1}^{H}\mathbf{1}(a_t(h)=\pi_h^+(s_t(h)))=\Omega(T)$*, and the attacker achieves its objective in* (7.6)*.*

Theorem 28 shows that the attacker can achieve its objective in $\tilde{O}(\sqrt{T})$ attack cost in episodic RL. Comparing Theorem 27 and Theorem 28, we have that an additional multiplicative factor in the amount of contamination (7.21) is incurred by the black box attack in comparison to the white box attack. However, this factor is small, more precisely $O(\sqrt{\log T})$.

Remark 1. *Reward poisoning attack has also been studied in black-box setting recently in RL by [173]. They consider attacking L online learners whereas we only has one learner. The attack objective of [173] is to force all these learning algorithms to execute the target policy π^+. We now highlight the key differences between our attack strategy and the strategy of [173]. Their attacker explores until the parameters of the MDP are estimated with precision, and that exploration phase last $O(T\log L)$ rounds, which is linear in T. On contrary, our exploration phase, which is the initialization phase, is completed within $O(\sqrt{T})$ episodes since it requires only a single observation for each (s,a) pair, and does not wait for the precise estimate of the parameters associated with the MDP. Our attack strategy compensate for this lack of precision by adding a negative bias $O(\sqrt{\log T})$ to the reward observation. Thus, our attack strategy saves the cost of learning in exploration phase. Additionally, unlike [173], our attack strategy does not attempt to learn the transition dynamics, which reduces the number of learning parameters. All these together makes our attack more effective. Indeed, the attack cost of the proposed attack in [173] is $\tilde{O}(T\log L+L\sqrt{T})$. However, if our attack is applied to L learners, the attack cost is $\tilde{O}(L\sqrt{T})$, which is better by an additive factor $O(T\log L)$. This is a significant save when $T\gg L$ (i.e., $L=o(T)$).*

7.4 Attacks in Bounded Reward Setting

7.4.1 Insufficiency of (Only) Reward or Action Manipulation

In this section, we focus on bounded reward setting, namely for all $(s,a)\in\mathcal{S}\times\mathcal{A}$, we have $r_{t,h}(s,a)\in[0,1]$ with mean $\mu(s,a)\in(0,1]$. We first show that there exist MDPs and

target policies π^+ such that for any reward manipulation attack and action manipulation attack, the objective of the attacker, namely (7.6), cannot be achieved.

In the reward manipulation attack, the attacker is subject to following constraints

$$r_t^o(s_t(h), a_t(h)) = r_t(s_t(h), a_t(h)) \text{ if } a_t(h) = \pi_h^+(s_t(h)), \text{ and } r_t^o(s_t(h), a_t(h)) \in [0, 1], \quad (7.22)$$

or equivalently,

$$\begin{aligned} &\epsilon_t(s_t(h), a_t(h)) = 0 \text{ if } a_t(h) = \pi_h^+(s_t(h)), \\ &\text{and } \epsilon_t(s_t(h), a_t(h)) \in [-r_t(s_t(h), a_t(h)), 1 - r_t(s_t(h), a_t(h))]. \end{aligned} \quad (7.23)$$

Thus, the reward observation can be manipulated only if the selected action is not the same as the target action of the policy, namely $a_t(h) \neq \pi_h^+(s_t(h))$. This constraint is crucial for obtaining sub-linear attack cost. The key idea is that if the objective in (7.6) is achieved by contaminating the reward of action $\pi_h^+(s_t(h))$, then the learner would execute the policy π^+ with high probability, namely $\Omega(T)$ times, and the total contamination may grow linearly with T. This constraint (or strategy of not contamination $\pi_h^+(s_t(h))$) is also applied in the previous literature in RL [172] and MAB [97, 143], where the reward manipulation are performed on non-desirable actions.

Similar to reward manipulation attack, in the action manipulation attack, the attacker is subject to following constraints

$$a_t^o(h) = a_t(h) \text{ if } a_t(h) = \pi_h^+(s_t(h)). \quad (7.24)$$

Thus, the action can be manipulated only if the selected action is not the same as the target action of the policy.

The following theorem establishes that only the reward manipulation or only the action manipulation cannot always guarantee successful attacks in bounded reward setting.

Theorem 29. *In bounded reward setting, we have*

1. *There exists an MDP and a target policy π^+ such that any reward manipulation attack satisfying* (7.22) *cannot be successful, namely achieve the objective in* (7.6)*.*

2. *There exists an MDP and a target policy π^+ such that any action manipulation attack satisfying* (7.24) *cannot be successful, namely achieve the objective in* (7.6)*.*

7.4.2 Efficient Attack by Combining Reward & Action Manipulation

We now show that the attacker can achieve its objective by combining the strength of both reward manipulation and action manipulation attacks. Additionally, the attack cost, namely sum of the amount of contamination and the number of action manipulation, is $\tilde{O}(\sqrt{T})$.

Here we start directly with the black-box attack here. For curious readers, we provide a discussion about the less realistic yet simpler white-box attack in Appendix 7.7.4. Given the target policy π^+, for all $s_t(h) \in \mathcal{S}$, $a_t(h) \in \mathcal{A}$ and $h \leq H$, the attack strategy is

$$a_t^o(h) = \begin{cases} a_t(h) & \text{if } a_t(h) = \pi_h^+(s), \\ \pi_h^+(s) & \text{if } a_t(h) \neq \pi_h^+(s), \end{cases} \tag{7.25}$$

and

$$r_t^o(s_t(h), a_t(h)) = \begin{cases} r_t(s_t(h), a_t(h)) & \text{if } a_t(h) = \pi_h^+(s), \\ 0 & \text{if } a_t(h) \neq \pi_h^+(s). \end{cases} \tag{7.26}$$

In this attack, the adversary manipulates both the action and the reward observation if $a_t(h) \neq \pi_h^+(s)$. The adversary manipulates the action to $\pi_h^+(s)$ to control the transition dynamics, and at the same time manipulates the reward observation to zero so that the action $a_t(h)$ appears to be sub-optimal in comparison to the action $\pi_h^+(s)$.

The bounded reward setting is difficult to attack since it cannot be always attacked by the reward manipulation or action manipulation alone. However, once the strength of reward

manipulation and action manipulation are available together, a simple attack in (7.25) and (7.26) can achieve the objective in (7.6), and does not require learning the parameters of the MDP.

The following theorem establishes that the attack cost is $O(\sqrt{T})$, which is sub-linear in T.

Theorem 30. *For any learning algorithm $\mathcal{A}$ such that for all $T \geq t_0$, the regret in the absence of attack is*

$$R^{\mathcal{A}}(T, H) = O(\sqrt{T}H^{\alpha}), \tag{7.27}$$

with probability at least $1-\delta$, where $\alpha \geq 1$ is a numerical constant; and for any sub-optimal target policy π^{+}, if an attacker follows strategy in (7.25) *and* (7.26)*, then with probability at least $1-\delta$, the number of reward manipulation attacks will be*

$$\sum_{t=1}^{T}\sum_{h=1}^{H}\mathbf{1}(\epsilon_{t,h}(s_t(h), a_t(h)) \neq 0) = O\big(\sqrt{T}H^{\alpha}/\min_{h,s}\mu(s, \pi_h^{+}(s))\big), \tag{7.28}$$

the amount of contamination is

$$\sum_{t=1}^{T}\sum_{h=1}^{H}|\epsilon_{t,h}(s_t(h), a_t(h)| = O\big(\sqrt{T}H^{\alpha}/\min_{h,s}\mu(s, \pi_h^{+}(s))\big), \tag{7.29}$$

the number of action manipulation attacks is

$$\sum_{t=1}^{T}\sum_{h=1}^{H}\mathbf{1}(a_t^{o}(h) \neq a_t(h)) = O\big(\sqrt{T}H^{\alpha}/\min_{h,s}\mu(s, \pi_h^{+}(s))\big), \tag{7.30}$$

$\sum_{t=1}^{T}\sum_{h=1}^{H}\mathbf{1}(a_t(h) = \pi_h^{+}(s_t(h))) = \Omega(T)$*, and the attacker achieves its objective in* (7.6)*.*

Comparing Theorem 30 and our white-box attack in Appendix 7.7.4, we can conclude that the attack cost of both black-box and white-box attack in this setting is $O(\sqrt{T})$. Unlike the unbounded reward setting, in bounded reward setting, the attack cost of the white-box and black box attack only vary by a numerical constant. This is due to the fact that amount of contamination

in each round is bounded above by unity.

7.5 Conclusion and Future Directions

This paper tries to understand poisoning attacks in reinforcement learning. Towards that end, we propose a reward manipulation attack for unbounded reward setting which successfully fool any no-regret RL algorithm to pull a target policy with $\tilde{O}(\sqrt{T})$ attack cost. Extending the study to bounded reward setting, we show that the adversary cannot achieve its objective using either reward manipulation or action manipulation attack even in white-box setting, where the information about the MDP is assumed to be known. Hence, to contaminate a no-regret RL algorithm, the adversary needs to combine the power of reward manipulation and action manipulation. Indeed, we show that an attack that uses both reward manipulation and action manipulation can achieve adversary's objective with $O(\sqrt{T})$ attack cost.

We studied the in-feasibility of the attack under the constraint that the adversary can attack only if $a_t(h) \neq \pi_h^+(s)$. Since this is a common constraint in the literature [172] and some efficient attacks in MAB also satisfy this constraint [97, 143], it would be interesting to establish theoretically that this constraint is also necessary for designing an attack with sub-linear cost. The study of infeasibility or feasibility of attack can be extended to dynamic manipulation attacks in RL. On the other hand, our results reveals the vulnerability of no-regret learning algorithms in RL. We hope this could spur more research on designing more robust algorithms for RL settings through ideas such as limited reward verification and corruption robustness [139, 78, 27].

7.6 Acknowledgement

Chapter 7, in part, contains material as it appears in Anshuka Rangi, Haifeng Xu, Long Tran-Thanh and Massimo Franceschetti, "Poisoning Attacks in Reinforcement Learning", *under preparation*. The dissertation author was the co-primary investigator and co-author of this paper.

7.7 Appendix

7.7.1 Proof of Theorem 27

Proof. First, we will show that the optimal policy under the reward manipulation attack in (7.7) is π^+, namely for all $\pi \neq \pi^+$, $h \leq H$ and $s \in \mathcal{S}$, we have

$$\tilde{V}_h^{\pi^+}(s) > \tilde{V}_h^{\pi}(s). \tag{7.31}$$

We will show this by induction. We will show that (7.31) holds for $h = H$. Then, we will show that (7.31) holds for $h < H$ if it holds for $h + 1$. At $h = H$, for all π, using (7.7), we have that

$$\tilde{Q}_H^{\pi}(s,a) = \begin{cases} \mu(s,a) & \text{if } a = \pi_H^+(s), \\ \mu(s,\pi_H^+(s)) - \epsilon & \text{if } otherwise. \end{cases} \tag{7.32}$$

This implies that for $h = H$, we have that (7.31) holds, and for all $(s,a) \in \mathcal{S} \times \mathcal{A}$ such that $a \neq \pi_H^+(s)$, we have

$$\tilde{Q}_H^{\pi}(s,a) = \tilde{Q}_H^{\pi^+}(s,\pi_H^+(s)) - \epsilon. \tag{7.33}$$

Now, consider any $h < H$. Let (7.31) holds for $h + 1$. Using (7.7), for all π, we have that

$$\tilde{Q}_h^{\pi}(s,a) = \begin{cases} \mu(s,a) + \mathbb{E}_{s' \sim P(s'|s,a)}[\tilde{V}_{h+1}^{\pi}(s')] & \text{if } a = \pi_h^+(s), \\ \tilde{Q}_h^{\pi^+}(s,\pi_h^+(s)) - \mathbb{E}_{s' \sim P(s'|s,a)}[\tilde{V}_{h+1}^{\pi^+}(s')] + \mathbb{E}_{s' \sim P(s'|s,a)}[\tilde{V}_{h+1}^{\pi}(s')] - \epsilon & \text{if } otherwise. \end{cases} \tag{7.34}$$

Since (7.31) holds for $h + 1$, we have that for $a = \pi_h^+(s)$

$$\tilde{Q}_h^{\pi}(s,a) < \mu(s,a) + \mathbb{E}_{s' \sim P(s'|s,a)}[\tilde{V}_{h+1}^{\pi^+}(s')] = \tilde{Q}_h^{\pi^+}(s,a). \tag{7.35}$$

Additionally, for $a \neq \pi_h^+(s)$, we have

$$\begin{aligned}\tilde{Q}_h^{\pi}(s,a) &= \tilde{Q}_h^{\pi^+}(s,\pi_h^+(s)) - \mathbb{E}_{s'\sim P(s'|s,a)}[\tilde{V}_{h+1}^{\pi^+}(s')] + \mathbb{E}_{s'\sim P(s'|s,a)}[\tilde{V}_{h+1}^{\pi}(s')] - \epsilon, \\ &= \tilde{Q}_h^{\pi^+}(s,\pi_h^+(s)) + \mathbb{E}_{s'\sim P(s'|s,a)}[\tilde{V}_{h+1}^{\pi}(s') - \tilde{V}_{h+1}^{\pi^+}(s')] - \epsilon, \\ &\overset{(a)}{<} \tilde{Q}_h^{\pi^+}(s,\pi_h^+(s)) - \epsilon, \end{aligned} \tag{7.36}$$

where (a) follows from the fact that (7.31) holds for $h+1$. Hence, the first step of the proof follows.

Let $\Delta(a) = \min_{s,h,\pi} \tilde{Q}_h^{\pi^+}(s,\pi_h^+(s)) - \tilde{Q}_h^{\pi}(s,a)$. Using (7.36), for $a \neq \pi_h^+(s)$, we have that

$$\Delta(a) \geq \epsilon. \tag{7.37}$$

Now, using (7.37), we have that

$$\begin{aligned}\sum_{t=1}^{T}\sum_{h=1}^{H} \epsilon \mathbf{1}(a_t(h) \neq \pi_h^+(s_t(h))) &\leq R^{\mathcal{A}}(T,H), \\ &= O(\sqrt{T}H^{\alpha}), \end{aligned} \tag{7.38}$$

with probability $1-\delta$, where the last inequality follows from (7.9). This along with (7.7) implies that (7.10) follows since the contamination happens only if $a_t(h) \neq \pi_h^+(s)$.

Additionally, for all $h \leq H$ and $(s_t(h), a_t(h)) \in \mathcal{S} \times \mathcal{A}$ such that $a_t(h) \neq \pi_h^+(s)$, we have that the amount of contamination is

$$\begin{aligned}&\left|\tilde{Q}_h^{\pi^+}(s_t(h),\pi_h^+(s_t(h))) - \mathbb{E}_{s'\sim P(s'|s_t(h),a_t(h))}[\tilde{V}_{h+1}^{\pi^+}(s')] - \epsilon - \mu(s_t(h),a_t(h))\right| \\ &\leq \left|\tilde{Q}_h^{\pi^+}(s_t(h),\pi_h^+(s_t(h))) - \mathbb{E}_{s'\sim P(s'|s_t(h),a_t(h))}[\tilde{V}_{h+1}^{\pi^+}(s')]\right| + \epsilon + \max_{s,a}\left|\mu(s,a)\right| \\ &\leq (H+1)\max_{s,a}\left|\mu(s,a)\right| + \epsilon. \end{aligned} \tag{7.39}$$

This along with (7.38) implies that with probability $1-\delta$,

$$\sum_{t=1}^{T}\sum_{h=1}^{H}|\epsilon_{t,h}(s_t(h),a_t(h))| = O\bigg(((H+1)\max_{s,a}\big|\mu(s,a)\big|+\epsilon)\frac{\sqrt{T}H^{\alpha}}{\epsilon}\bigg). \tag{7.40}$$

Hence, we have that (7.11) follows. Finally, we have that with probability $1-\delta$

$$\begin{aligned}\sum_{t=1}^{T}\sum_{h=1}^{H}\mathbf{1}(a_t(h)=\pi_h^+(s_t(h))) &= TH-\sum_{t=1}^{T}\sum_{h=1}^{H}\mathbf{1}(a_t(h)\neq\pi_h^+(s_t(h)))\\ &= \Omega(T),\end{aligned} \tag{7.41}$$

where the last equality follows from (7.38). Hence, the statement of the theorem follows.

□

7.7.2 Proof of Theorem 28

The following lemma will be used in our proof.

Lemma 30. *For all $(s,a)\in\mathcal{S}\times\mathcal{A}$, $T\geq 1$ and $H\geq 1$, we have that*

$$\mathbb{P}\bigg(|\hat{\mu}(s,a)-\mu(s,a)|>\sigma\sqrt{4\log(2THSA)/N_{t,h}(s,a)}\bigg)\leq\frac{1}{(THSA)^2}. \tag{7.42}$$

Proof. The above lemma following using concentration inequality for sub-gaussian random variable. □

Proof of Theorem 28. Let

$$T^*=\max\{t\leq T:\exists(s,a)\in S\times A,\text{ we have } N_{t,h}(s,a)=0\}, \tag{7.43}$$

where $N_{t,h}(s,a)$ is the number of times action a is selected in state s until the step h in episode t.

Firstly, we show that with high probability, we have $T^*\leq\sqrt{T}$. Then, the analysis of the amount of contamination and number of attacks is divided into two parts: computing these quantities before the episode T^* and computing these quantities after the episode T^*.

First, we will show that with probability $1-\delta$,

$$T^* \leq t_0 \leq \sqrt{T}. \tag{7.44}$$

Using (7.19), we have that with probability $1-\delta$,

$$\text{SubOpt}(t_0, \epsilon, \delta) = O(\sqrt{TH}/\epsilon), \tag{7.45}$$

where $\text{SubOpt}(t_0, \epsilon, \delta)$ is the number of times suboptimal actions contributing at least ϵ to the regret are selected. This along with [173, Lemma 5.1] implies that until round t_0, with probability $1-\delta/(SA)$, we have

$$N_{t_0,H}(s,a) \geq \frac{c_2 \log^2(\delta/4SA\beta)}{\log(8SA/\delta) + 1.34\log(\delta/4SA\beta)} > 0, \tag{7.46}$$

where c_2 is a positive numerical constant, and $\beta \leq \delta/4SA$. Since $N_{t_0,H}(s,a)$ are all integers, using the union bound, with probability $1-\delta$, we have for all $(s,a) \in \mathcal{S} \times \mathcal{A}$,

$$N_{t_0,H}(s,a) \geq 1. \tag{7.47}$$

Thus, we have that (7.44) follows using the fact that $T \geq t_0^2$.

Until round T^*, using the fact that reward observation are σ^2-subgaussian random variable, we have that

$$\mathbb{P}\left(|r_t(s,a) - \mu(s,a)| > \sigma\sqrt{4\log(2HSAT)}\right) \leq 1/(HSAT)^2. \tag{7.48}$$

Let event

$$\mathcal{E} = \{\forall t \leq T^*, h \leq H : |r_t(s_t(h), a_t(h)) - \mu(s_t(h), a_t(h))| \leq \sigma\sqrt{4\log(2HSAT)}\} \tag{7.49}$$

If $\mathcal{E}$ is true, then we have that the amount of contamination is at most

$$T^*H(1+\max_{s,a}|\mu(s,a)|+\sigma\sqrt{4\log(2HSAT)}). \tag{7.50}$$

Now, using (7.48), we have that

$$\mathbb{P}(\bar{\mathcal{E}}) \leq 1/HSAT, \tag{7.51}$$

where $\bar{\mathcal{E}}$ denotes the complement of event $\mathcal{E}$. This along with (7.44) implies that the amount of contamination until T^* is at most

$$\sum_{t=1}^{T^*}\sum_{h=1}^{H}|\epsilon_{t,h}(s_t(h),a_t(h)|=O(\sqrt{T}H(1+\max_{s,a}|\mu(s,a)|+\sqrt{4\log(2HSAT)})), \tag{7.52}$$

with probability $1-\delta-1/(HSAT)$. Likewise, the number of attacks until T^* is

$$\sum_{t=1}^{T^*}\sum_{h=1}^{H}\mathbf{1}(\epsilon_{t,h}(s_t(h),a_t(h)\neq 0)\leq\sqrt{T}H, \tag{7.53}$$

with probability $1-\delta$.

After round T^*, let the event

$$\mathcal{E}_2=\{\forall T^*\leq t\leq T, h\leq H, s\in\mathcal{S}, a\in\mathcal{A}:\hat{\mu}_{t,h}^{LCB}(s,a)\leq\mu(s,a)\leq\hat{\mu}_{t,h}^{UCB}(s,a)\}, \tag{7.54}$$

where $\hat{\mu}_{t,h}^{LCB}(s,a)$ and $\hat{\mu}_{t,h}^{UCB}(s,a)$ is the estimate $\hat{\mu}^{LCB}(s,a)$ and $\hat{\mu}^{UCB}(s,a)$ in Algorithm 13 in episode t at step h.

Using Lemma 30, we have that

$$\mathbb{P}\big(\bar{\mathcal{E}}_2\big)\leq 1/(HSAT). \tag{7.55}$$

If event $\mathcal{E}_2$ occurs, then for all $T^* < t \leq T$ and $h \leq H$, we have that

$$\begin{aligned}
&\hat{\mu}_{t,h}^{LCB}(s_t(h), \pi_h^+(s_t(h))) + (H-h) \min_{s,a \in \mathcal{S} \times \mathcal{A}} \hat{\mu}_{t,h}^{LCB}(s,a) \\
&\leq \mu(s_t(h), \pi_h^+(s_t(h))) + (H-h) \min_{s,a \in S \times A} \mu(s,a), \\
&\leq \tilde{Q}_h^{\pi^+}(s_t(h), \pi_h^+(s_t(h))).
\end{aligned} \tag{7.56}$$

Also, if event $\mathcal{E}_2$ occurs, then for all $T^* < t \leq T$ and $h \leq H$, we have that

$$\begin{aligned}
(H-h) \max_{s,a \in \mathcal{S} \times \mathcal{A}} \hat{\mu}_{t,h}^{UCB}(s,a) &\geq (H-h) \max_{s,a \in S \times A} \mu(s,a) \\
&\geq \mathbb{E}_{s' \sim \mathbb{P}(s'|s_t(h), a_t(h))}[\tilde{V}_{h+1}^{\pi^+}(s')].
\end{aligned} \tag{7.57}$$

Now, combining (7.56) and (7.57), under event $\mathcal{E}_2$, for all $T^* < t \leq T$ and $h \leq H$, we have that

$$\begin{aligned}
&r_t^o(s_t(h), a_t(h)) \\
&= \hat{\mu}_{t,h}^{LCB}(s_t(h), \pi_h^+(s_t(h))) + (H-h) \min_{s,a \in S \times A} \hat{\mu}_{t,h}^{LCB}(s,a) - (H-h) \max_{s,a \in S \times A} \hat{\mu}_{t,h}^{UCB}(s,a) - \epsilon \\
&\leq \tilde{Q}_h^{\pi^+}(s_t(h), \pi_h^+(s_t(h))) - \mathbb{E}_{s' \sim \mathbb{P}(s'|s_t(h), a_t(h))}[\tilde{V}_{h+1}^{\pi^+}(s')] - \epsilon.
\end{aligned} \tag{7.58}$$

Now, similar to (7.31), using (7.58), we can show that under event $\mathcal{E}_2$, we have

$$V_h^{\pi^+}(s) = \sup_{\pi} V_h^{\pi}(s). \tag{7.59}$$

Additionally, similar to (7.37), we also have that under event $\mathcal{E}_2$, we have

$$\Delta(a) = \min_{s,h,\pi} \tilde{Q}_h^{\pi^+}(s, \pi_h^+(s)) - \tilde{Q}_h^{\pi}(s,a) \geq \epsilon. \tag{7.60}$$

Now, using (7.60), under event $\mathcal{E}_2$, we have that with probability $1-\delta$,

$$\sum_{t=T^*}^{T}\sum_{h=1}^{H}\epsilon\mathbf{1}(a_t(h)\neq\pi_h^+(s_t(h)))\leq R^{\mathcal{A}}(T,H)=O(\sqrt{T}H^{\alpha}). \tag{7.61}$$

Thus, combining (7.55) and (7.61), we have that with probability $1-\delta-1/(THSA)$,

$$\sum_{t=1}^{T^*}\sum_{h=1}^{H}\mathbf{1}(\epsilon_{t,h}(s_t(h),a_t(h)\neq 0)\leq O(\sqrt{TH}/\epsilon). \tag{7.62}$$

The amount of contamination after T^* is

$$\begin{aligned}
&|r_t^o(s_t(h),a_t(h))-r_t(s_t(h),a_t(h))|\\
&\overset{(a)}{\leq}|r_t^o(s_t(h),a_t(h))-\mu(s_t(h),a_t(h))|+\sigma\sqrt{4\log(2HSAT)}\\
&\leq|\mu(s_t(h),a_t(h))|+|r_t^o(s_t(h),a_t(h))|+\sigma\sqrt{4\log(2HSAT)},\\
&\overset{(b)}{\leq}|\mu(s_t(h),a_t(h))|+(2H+1)\max_{s,a}|\mu(s,a)|+\epsilon+(4H+3)\sigma\sqrt{4\log(2HSAT)},\\
&\leq(2H+2)\max_{s,a}|\mu(s,a)|+\epsilon+(4H+3)\sigma\sqrt{4\log(2HSAT)},
\end{aligned} \tag{7.63}$$

where (a) follows from (7.48), and (b) follows under event $\mathcal{E}_2$. This implies that the total amount of contamination following T^* is

$$\sum_{t=1}^{T^*}\sum_{h=1}^{H}|\epsilon_{t,h}(s_t(h),a_t(h)|=O\bigg(\sqrt{T}H^{\alpha}(H+\epsilon+H\sigma\sqrt{\log(HTSA)})/\epsilon\bigg), \tag{7.64}$$

with probability $1-\delta-2/(HSAT)$.

Combining (7.53) and (7.62), we have that (7.20) follows. Also, combining (7.52) and (7.64), we have that (7.21) follows. Finally, we have that with probability $1-\delta$

$$\begin{aligned}
\sum_{t=1}^{T}\sum_{h=1}^{H}\mathbf{1}(a_t(h)=\pi_h^+(s_t(h)))&=TH-\sum_{t=1}^{T}\sum_{h=1}^{H}\mathbf{1}(a_t(h)\neq\pi_h^+(s_t(h)))\\
&=\Omega(T),
\end{aligned} \tag{7.65}$$

where the last equality follows from (7.61). The statement of the theorem follows. □

7.7.3 Proof of Theorem 29

Insufficiency of (only) reward manipulation

In this example, $\mathcal{S} = \{s_1, s_2\}$ and $\mathcal{A} = \{a_1, a_2\}$. The transition dynamics is

$$\mathbb{P}(s_1|s_1, a_1) = 1, \mathbb{P}(s_2|s_1, a_1) = 0, \mathbb{P}(s_1|s_1, a_2) = 0, \mathbb{P}(s_2|s_1, a_2) = 1, \tag{7.66}$$

$$\mathbb{P}(s_1|s_2, a_1) = 0, \mathbb{P}(s_2|s_2, a_1) = 1, \mathbb{P}(s_1|s_2, a_2) = 1, \mathbb{P}(s_2|s_2, a_2) = 0. \tag{7.67}$$

Also, we have

$$\mu(s_1, a_1) = \epsilon_1 = 0.25, \mu(s_1, a_2) = 1, \mu(s_2, a_1) = \epsilon_2 = .6, \mu(s_2, a_2) = 1 \tag{7.68}$$

Let $H = 2$. The target policy π^+ for the attacker is

$$\forall h \leq H: \quad \pi_h^+(s_1) = a_1 \text{ and } \pi_h^+(s_2) = a_1. \tag{7.69}$$

Similar to [172], the attacker is subject to following constraints

$$r_t^o(s_t(h), a_t(h)) = r_t(s_t(h), a_t(h)) \text{ if } a_t(h) = \pi_h^+(s_t(h)), \text{ and } r_t^o(s_t(h), a_t(h)) \in [0, 1], \tag{7.70}$$

or equivalently,

$$\begin{aligned} &\epsilon_t(s_t(h), a_t(h)) = 0 \text{ if } a_t(h) = \pi_h^+(s_t(h)), \\ &\text{and } \epsilon_t(s_t(h), a_t(h)) \in [-r_t(s_t(h), a_t(h)), 1 - r_t(s_t(h), a_t(h))]. \end{aligned} \tag{7.71}$$

The objective of the attacker is that for all $\pi \neq \pi^+$, $h \leq H$ and $s \in \mathcal{S}$,

$$V_h^{\pi}(s) < V_h^{\pi^+}(s). \tag{7.72}$$

Let policy $\tilde{\pi}$ be

$$\begin{aligned} h = 1: \quad & \tilde{\pi}_h(s_1) = a_2 \text{ and } \tilde{\pi}_h(s_2) = a_2, \\ h = 2: \quad & \tilde{\pi}_h(s_1) = a_1 \text{ and } \tilde{\pi}_h(s_2) = a_1. \end{aligned} \tag{7.73}$$

At $h = H - 1 = 1$, for all reward manipulation attack satisfying (7.70), we will show that

$$V_h^{\tilde{\pi}}(s_1) > V_h^{\pi^+}(s_1). \tag{7.74}$$

At $h = H - 1 = 1$, we have

$$V_h^{\pi^+}(s_1) = \mu(s_1, a_1) + \mu(s_1, a_1) = 2\epsilon_1. \tag{7.75}$$

Additionally, we have

$$\begin{aligned} V_h^{\tilde{\pi}}(s_1) &\overset{(a)}{=} r_t^o(s_1, a_2) + \epsilon_2, \\ &\overset{(b)}{\geq} \epsilon_2, \end{aligned} \tag{7.76}$$

where (a) follows from the facts that $\tilde{\pi}_{H-1}(s_1) = a_2 \neq \pi_H^+(s_1)$, which implies that the attacker can manipulate this observation, the next state at step H is s_2, and $\tilde{\pi}_H(s_2) = a_1 = \pi_H^+(s_2)$, which implies that the attacker can not manipulate this observation, and (b) follows from the fact that $r_t^o(s_1, a_2) \in [0, 1]$ using (7.70).

Now, since $\epsilon_1 = 0.25$ and $\epsilon_2 = 0.6$, comparing (7.75) and (7.76), we have that

$$V_{H-1}^{\tilde{\pi}}(s_1) > V_{H-1}^{\pi^+}(s_1). \tag{7.77}$$

This implies that for any reward manipulation attack in bounded setting satisfying (7.70), there exists a policy $\pi \neq \pi^+$, a step $h \leq H$ and a state $s_1 \in \mathcal{S}$ such that

$$V_h^{\pi}(s) > V_h^{\pi^+}(s). \tag{7.78}$$

Insufficiency of (only) Action Manipulation

The MDP construction and target policy π^+ are the same as the one in Theorem 29. Let $H = 2$. We also consider policy $\tilde{\pi}$ as follows

$$\begin{aligned} h = 1: &\qquad \tilde{\pi}_h(s_1) = a_2 \text{ and } \tilde{\pi}_h(s_2) = a_2, \\ h = 2: &\qquad \tilde{\pi}_h(s_1) = a_1 \text{ and } \tilde{\pi}_h(s_2) = a_1. \end{aligned} \tag{7.79}$$

At $h = H - 1 = 1$, for all reward manipulation attack satisfying (7.70), we will show that

$$V_h^{\tilde{\pi}}(s_1) > V_h^{\pi^+}(s_1). \tag{7.80}$$

At $h = H - 1 = 1$, we have

$$V_h^{\pi^+}(s_1) = \mu(s_1, a_1) + \mu(s_1, a_1) = 2\epsilon_1. \tag{7.81}$$

Additionally, we have

$$\begin{aligned} V_h^{\tilde{\pi}}(s_1) &\overset{(a)}{\geq} \min\{\mu(s_1, a_1) + \mu(s_1, a_1), \mu(s_1, a_2) + \mu(s_2, a_1)\} \\ &\overset{(b)}{\geq} \min\{2\epsilon_1, 1 + \epsilon_2\} \end{aligned} \tag{7.82}$$

where (a) follows from the facts that $\tilde{\pi}_{H-1}(s_1) = a_2 \neq \pi_H^+(s_1)$, which implies that the attacker can manipulate the action, namely $a_t(h) = a_1$ or $a_t(h) = a_2$, and if $a_t(h) = a_1$ (or $a_t(h) = a_2$), then $V_h^{\tilde{\pi}}(s_1)$ is $2\mu(s_1, a_1)$ (or $\mu(s_1, a_2) + \mu(s_2, a_1)$), and (b) follows from (7.68).

Now, since $\epsilon_1 = 0.25$ and $\epsilon_2 = 0.6$, comparing (7.81) and (7.82), we have that

$$V_{H-1}^{\tilde{\pi}}(s_1) \geq V_{H-1}^{\pi^+}(s_1). \tag{7.83}$$

This implies that for any action manipulation attack in bounded setting satisfying (7.70), there exists a policy $\pi \neq \pi^+$, a step $h \leq H$ and a state $s_1 \in \mathcal{S}$ such that

$$V_h^{\pi}(s) \geq V_h^{\pi^+}(s). \tag{7.84}$$

7.7.4 White-box Attack in Bounded Reward Setting

The attack strategy is

$$a_t^o(h) = \begin{cases} a_t(h) & \text{if } a_t(h) = \pi_h^+(s), \\ \pi_h^+(s) & \text{if } a_t(h) \neq \pi_h^+(s), \end{cases} \tag{7.85}$$

and

$$r_t^o(s_t(h), a_t(h)) = \begin{cases} r_t(s_t(h), a_t(h)) & \text{if } a_t(h) = \pi_h^+(s_t(h)), \\ \tilde{Q}_h^{\pi^+}(s_t(h), \pi_h^+(s_t(h))) & otherwise \\ -E_{s' \sim \mathbb{P}(s'|s_t(h), \pi_h^+(s_t(h)))}[\bar{V}_{h+1}^{\pi^+}(s')] - \epsilon \end{cases} \tag{7.86}$$

where $\bar{Q}_h^{\pi}(s, a)$ is the expected reward in state s for action a introduced by the above reward and action manipulation under policy π, and $\bar{V}_h^{\pi}(s)$ is the expected reward in state s for the above reward and action manipulation under policy π.

Theorem 31. *For any learning algorithm $\mathcal{A}$ such thatt for all $T \geq t_0$, the regret in the absence of attack is*

$$R^{\mathcal{A}}(T, H) = O(\sqrt{T} H^{\alpha}), \tag{7.87}$$

with probability at least $1-\delta$*, where* $\alpha \geq 1$ *is a numerical constant; and for any sub-optimal target policy* π^+ *and* $0 < \epsilon \leq \min_{h \leq H, s \in \mathcal{S}} \mu(s, \pi_h^+(s))$*, if an attacker follows strategy in* (7.85) *and* (7.86)*, then with probability at least* $1-\delta$*, the number of reward manipulation attacks will be*

$$\sum_{t=1}^{T} \sum_{h=1}^{H} \mathbf{1}(\epsilon_{t,h}(s_t(h), a_t(h)) \neq 0) = O\big(\sqrt{T} H^{\alpha}/\epsilon\big), \tag{7.88}$$

the amount of contamination is

$$\sum_{t=1}^{T} \sum_{h=1}^{H} |\epsilon_{t,h}(s_t(h), a_t(h)| = O\big(\sqrt{T} H^{\alpha}/\epsilon\big), \tag{7.89}$$

the number of action manipulation attacks is

$$\sum_{t=1}^{T} \sum_{h=1}^{H} \mathbf{1}(a_t^o(h) \neq a_t(h)) = O\big(\sqrt{T} H^{\alpha}/\epsilon\big), \tag{7.90}$$

and $\sum_{t=1}^{T} \sum_{h=1}^{H} \mathbf{1}(a_t(h) = \pi_h^+(s_t(h))) = \Omega(T)$.

Proof. First, we will show that the optimal policy under action and reward manipulation attack in (7.85) and (7.86) is π^+, namely for all $\pi \neq \pi^+$, $h \leq H$ and $s \in \mathcal{S}$, we have

$$\bar{V}_h^{\pi^+}(s) > \bar{V}_h^{\pi}(s). \tag{7.91}$$

We will show this by induction. We will that that (7.91) holds for $h = H$. Then, we will show that (7.91) holds for $h < H$ if it holds for $h+1$. At $h = H$, for all π, using (7.85) and (7.86), we have that

$$\bar{Q}_H^{\pi}(s,a) = \begin{cases} \mu(s,a) & \text{if } a = \pi_H^+(s), \\ \mu(s, \pi_H^+(s)) - \epsilon & \text{if } otherwise, \end{cases} \tag{7.92}$$

since episode terminates at step H. This implies that for $h = H$, we have that (7.91) holds, and

for all $(s,a) \in \mathcal{S} \times \mathcal{A}$ such that $a \neq \pi_H^+(s)$, we have

$$\bar{Q}_H^{\pi}(s,a) = \bar{Q}_H^{\pi^+}(s, \pi_H^+(s)) - \epsilon. \tag{7.93}$$

Now, consider any $h < H$. Let (7.91) holds for $h+1$. Using (7.85) and (7.86), for all π, we have that

$$\bar{Q}_h^{\pi}(s,a) = \begin{cases} \mu(s,a) + \mathbb{E}_{s' \sim P(s'|s,a)}[\bar{V}_{h+1}^{\pi}(s')] & \text{if } a = \pi_h^+(s), \\ \bar{Q}_h^{\pi^+}(s, \pi_h^+(s)) - \mathbb{E}_{s' \sim P(s'|s,\pi_h^+(s))}[\bar{V}_{h+1}^{\pi^+}(s')] & otherwise \\ + \mathbb{E}_{s' \sim P(s'|s,\pi_h^+(s))}[\bar{V}_{h+1}^{\pi}(s')] - \epsilon & \end{cases} \tag{7.94}$$

Since (7.91) holds for $h+1$, we have that for $a = \pi_h^+(s)$,

$$\bar{Q}_h^{\pi}(s,a) < \mu(s,a) + \mathbb{E}_{s' \sim P(s'|s,a)}[\bar{V}_{h+1}^{\pi^+}(s')] = \bar{Q}_h^{\pi^+}(s,a). \tag{7.95}$$

Additionally, for $a \neq \pi_h^+(s)$, we have

$$\begin{aligned} \bar{Q}_h^{\pi}(s,a) &= \bar{Q}_h^{\pi^+}(s, \pi_h^+(s)) - \mathbb{E}_{s' \sim P(s'|s,\pi_h^+(s))}[\bar{V}_{h+1}^{\pi^+}(s')] + \mathbb{E}_{s' \sim P(s'|s,\pi_h^+(s))}[\bar{V}_{h+1}^{\pi}(s')] - \epsilon, \\ &= \bar{Q}_h^{\pi^+}(s, \pi_h^+(s)) + \mathbb{E}_{s' \sim P(s'|s,\pi_h^+(s))}[\bar{V}_{h+1}^{\pi}(s') - \bar{V}_{h+1}^{\pi^+}(s')] - \epsilon, \\ &\overset{(a)}{<} \bar{Q}_h^{\pi^+}(s, \pi_h^+(s)) - \epsilon, \end{aligned} \tag{7.96}$$

where (a) follows from the fact that (7.91) holds for $h+1$. Hence, the first step of the proof follows.

Additionally, the attack satisfies the constraint that $r_t^o(s, a_t(h)) \in [0,1]$. For $a_t(h) \neq$

$\pi_h^+(s_t(h))$, we have

$$\begin{aligned} r_t^o(s_t(h), a_t(h)) &= \bar{Q}_h^{\pi^+}(s_t(h), \pi_h^+(s_t(h))) - \mathbb{E}_{s' \sim P(s'|s_t(h), \pi_h^+(s_t(h)))}[\bar{V}_{h+1}^{\pi^+}(s')] - \epsilon, \\ &\overset{(a)}{=} \mu(s_t(h), \pi_h^+(s_t(h))) - \epsilon, \\ &\overset{(b)}{\geq} 0, \end{aligned} \tag{7.97}$$

where (a) follows from the fact that

$$\bar{Q}_h^{\pi^+}(s_t(h), \pi_h^+(s_t(h))) = \mu(s_t(h), \pi_h^+(s_t(h))) + E_{s' \sim P(s'|a_t(h), \pi_h^+(s_t(h)))}[\bar{V}_{h+1}^{\pi^+}(s')], \tag{7.98}$$

and (b) follows from the fact that $0 < \epsilon \leq \min_{h \leq H, s \in \mathcal{S}} \mu(s, \pi_h^+(s))$. Additionally, we have

$$r_t^o(s_t(h), a_t(h)) = \mu(s_t(h), a_t(h)) - \epsilon \leq 1, \tag{7.99}$$

since $\mu(s_t(h), a_t(h)) \in (0, 1]$.

Let $\Delta(a) = \min_{s,h,\pi} \bar{Q}_h^{\pi^+}(s, \pi_h^+(s)) - \bar{Q}_h^{\pi}(s, a)$. Using (7.96), we have that

$$\Delta(a) \geq \epsilon. \tag{7.100}$$

Now, using (7.100), we have that

$$\begin{aligned} \sum_{t=1}^{T} \sum_{h=1}^{H} \epsilon \mathbf{1}(a_t(h) \neq \pi_h^+(s_t(h))) &\leq R^{\mathcal{A}}(T, H), \\ &= O(\sqrt{T} H^{\alpha}), \end{aligned} \tag{7.101}$$

with probability $1 - \delta$, where the last inequality follows from (7.87). This along with (7.85) and (7.86) implies that (7.88) and (7.90) follows since the contamination happens only if $a_t(h) \neq \pi_h^+(s)$.

Additionally, for all $h \leq H$ and $(s_t(h), a_t(h)) \in \mathcal{S} \times \mathcal{A}$ such that $a_t(h) \neq \pi_h^+(s)$, we

have that the amount of contamination is at most one, which implies (7.89) follows.

Finally, we have that with probability $1-\delta$

$$\sum_{t=1}^{T}\sum_{h=1}^{H}\mathbf{1}(a_t(h)=\pi_h^+(s_t(h))) = TH - \sum_{t=1}^{T}\sum_{h=1}^{H}\mathbf{1}(a_t(h)\neq\pi_h^+(s_t(h))) \\ = \Omega(T), \tag{7.102}$$

where the last equality follows from (7.101). Hence, the statement of the theorem follows. □

7.7.5 Proof of Theorem 30

Proof. First, we will show that the optimal policy under action and reward manipulation attack in (7.85) and (7.86) is π^+, namely for all $\pi \neq \pi^+$, $h \leq H$ and $s \in \mathcal{S}$, we have

$$\bar{V}_h^{\pi^+}(s) > \bar{V}_h^{\pi}(s). \tag{7.103}$$

We will show this by induction. We will that that (7.103) holds for $h = H$. Then, we will show that (7.103) holds for $h < H$ if it holds for $h+1$. At $h = H$, for all π, using (7.25) and (7.26), we have that

$$\bar{Q}_H^{\pi}(s,a) = \begin{cases} \mu(s,a) & \text{if } a = \pi_H^+(s), \\ 0 & otherwise, \end{cases} \tag{7.104}$$

since episode terminates at step H. This implies that for $h = H$, we have that (7.103) holds, and for all $(s,a) \in \mathcal{S} \times \mathcal{A}$ such that $a \neq \pi_H^+(s)$, we have

$$\bar{Q}_H^{\pi}(s,a) = 0. \tag{7.105}$$

Now, consider any $h < H$. Let (7.103) holds for $h+1$. Using (7.25) and (7.26), for all π, we have that

$$\bar{Q}_h^{\pi}(s,a) = \begin{cases} \mu(s,a) + \mathbb{E}_{s'\sim P(s'|s,a)}[\bar{V}_{h+1}^{\pi}(s')] & \text{if } a = \pi_h^+(s), \\ 0 + \mathbb{E}_{s'\sim P(s'|s,\pi_h^+(s))}[\bar{V}_{h+1}^{\pi}(s')] & \text{if } otherwise. \end{cases} \tag{7.106}$$

Since (7.103) holds for $h+1$, we have that for $a = \pi_h^+(s)$,

$$\bar{Q}_h^{\pi}(s,a) < \mu(s,a) + \mathbb{E}_{s' \sim P(s'|s,a)}[\bar{V}_{h+1}^{\pi^+}(s')] = \bar{Q}_h^{\pi^+}(s,a). \tag{7.107}$$

Additionally, for $a \neq \pi_h^+(s)$, we have

$$\begin{aligned}\bar{Q}_h^{\pi}(s,a) &= \mathbb{E}_{s' \sim P(s'|s,\pi_h^+(s))}[\bar{V}_{h+1}^{\pi}(s')], \\ &\overset{(a)}{<} \mathbb{E}_{s' \sim P(s'|s,\pi_h^+(s))}[\bar{V}_{h+1}^{\pi^+}(s')], \\ &\overset{(b)}{<} \bar{Q}_h^{\pi^+}(s,\pi_h^+(s)),\end{aligned} \tag{7.108}$$

where (a) follows from the fact that (7.103) holds for $h+1$, and (b) follows from the definition of $\bar{Q}_h^{\pi^+}(s,\pi_h^+(s))$. Hence, the first step of the proof follows.

Additionally, the attack satisfies the constraint that $r_t^o(s,a_t(h)) \in [0,1]$.

Let $\Delta(a) = \min_{s,h,\pi} \bar{Q}_h^{\pi^+}(s,\pi_h^+(s)) - \bar{Q}_h^{\pi}(s,a)$. Using the fact that $r_t^o(s_t(h),a_t(h)) = 0$ if $a_t(h) \neq \pi_h^+(s_t(h))$, we have that

$$\Delta(a) \geq \min_{h,s} \mu(s,\pi_h^+(s)). \tag{7.109}$$

Now, using (7.109), we have that

$$\begin{aligned}\sum_{t=1}^{T}\sum_{h=1}^{H} \min_{h,s} \mu(s,\pi_h^+(s)) \mathbf{1}(a_t(h) \neq \pi_h^+(s_t(h))) &\leq R^{\mathcal{A}}(T,H), \\ &= O(\sqrt{T}H^{\alpha}),\end{aligned} \tag{7.110}$$

with probability $1-\delta$, where the last inequality follows from (7.27). This along with (7.25) and (7.26) implies that (7.28) and (7.30) follows since the contamination happens only if $a_t(h) \neq \pi_h^+(s)$.

Additionally, for all $h \leq H$ and $(s_t(h),a_t(h)) \in \mathcal{S} \times \mathcal{A}$ such that $a_t(h) \neq \pi_h^+(s)$, we have that the amount of contamination is at most one, which implies (7.29) follows.

Finally, we have that with probability $1 - \delta$

$$\sum_{t=1}^{T} \sum_{h=1}^{H} \mathbf{1}(a_t(h) = \pi_h^+(s_t(h))) = TH - \sum_{t=1}^{T} \sum_{h=1}^{H} \mathbf{1}(a_t(h) \neq \pi_h^+(s_t(h)))$$
$$= \Omega(T), \tag{7.111}$$

where the last equality follows from (7.110). Hence, the statement of the theorem follows. □

Chapter 8

Learning-based attacks in Cyber-Physical Systems

8.1 Introduction

Attacks directed to Cyber-Physical Systems (CPS) can have catastrophic consequences ranging from hampering the economy through financial scams, to possible losses of human lives through hijacking autonomous vehicles and drones, see [167, 201, 86]. In this framework, two important problems arise: understanding of the regime where the system can be attacked, and designing ways to mitigate these attacks and render the system secure, see [142, 243, 97, 240, 43, 217, 64, 146, 111, 113, 183]. Techniques developed to secure CPS include watermarking, moving target and baiting, and typically require either a loss of performance, or additional resources available at the controller, see [192, 153, 98, 67].

In this paper, we focus on the former aspect of the problem, namely understanding the regime under which the system can be attacked. We focus on linear plants and on an important and widely used class of attacks based on the "man-in-the-middle" (MITM) technique. In this case, the attacker takes over the physical plant's control and feedback signals, and acts as a malicious controller for the plant and fictitious plant for the controller. By doing so, it overrides the control signals with malicious inputs aimed at destroying the plant; and it overrides the feedback signals to the controller, trying to mimic the safe and legitimate operation of the system. In learning based MITM attack, we assume that the attacker has full access to both sensor and

control signals, but the plant dynamics are unknown to the attacker. Thus, the attacker needs to learn about the plant in order to being able to generate the fictitious signals to the controller that allow the attacker to remain undetected for the time needed to cause harm. On the other hand, the controller has perfect (or nearly perfect) knowledge of the system dynamics and is actively looking out for an anomalous behaviour in the feedback signals from the plant. This assumed information pattern is justified, since the controller is typically tuned in much longer than the attacker, and has knowledge of the system dynamics to a far greater precision than the attacker. Following the detection of the attacker, the controller can shut the plant down, or switch to a "safe" mode where the system is secured using additional resources, and the attacker is prevented from causing additional "harm" to the plant, see [51, 226, 208, 82].

We consider a learning-based MITM attack that evolves in two phases: *exploration and exploitation*. In the exploration phase, the attacker observes the plant state and control inputs, and learns the plant dynamics. In the exploitation phase, the attacker hijacks the plant, and utilizes the learned estimate to feed the fictitious feedback signals to the controller. During this phase, the attacker may also refine its estimate by continuing to learn. Within this context, our results are as follows: first, we provide a lower bound on the expected ϵ-deception time, namely the time required by the controller to make a decision regarding the presence of an attacker with confidence at least $1 - \epsilon \log(1/\epsilon)$. This bound is expressed in terms of the parameters of the attacker's learning algorithm and the controller's strategy. Second, we show that there exists a learning-based attack and a detection strategy such that a matching upper bound on the expected ϵ-deception time is obtained. We then show that for a wide range of learning algorithms, if the expected ϵ-deception time is at least of duration D, then the duration of the exploration phase of the attacker must be at least $\Omega(D/\log(1/\epsilon))$, as $\epsilon \to 0$. We establish that this bound is also order-optimal since there exists a learning algorithm such that if the duration of the exploration phase is $O(D/\log(1/\epsilon))$ as $\epsilon \to 0$, then the expected ϵ-deception time is at least D. Finally, we show that if the controller wants to detect the attacker in at most D duration with confidence at least $1 - \epsilon \log(1/\epsilon)$, then the expected energy expenditure on the control signal must be at least

of order $\Omega(D/\log(1/\epsilon))$, as $\epsilon \to 0$.

8.2 Related Work

There is a wide range of recent research on learning-based control for linear systems [47, 191, 23, 66, 110, 220, 95, 44, 61, 127, 35]. In these works, learning algorithms are proposed to design controllers in the presence of uncertainty. In contrast, in our setting we assume that the controller has full knowledge of the system dynamics, while the attacker may take advantage of these algorithms. Thus, our focus is not on the optimal control design given the available data, but rather on the trade-offs between the attacker's learning capability, the controller's detection strategy, and the control cost.

The MITM attack has been extensively studied in control systems for two special cases, namely, the replay attack and the statistical duplicate attack. The detection of replay attacks has been studied in [152, 153, 151], and ways to mitigate these attacks have been studied in [249]. Likewise, the ways to detect and mitigate statistical duplicate attacks has been studied in [192, 204, 170]. These works do not consider learning-enabled attackers, and analyze the performance of the controller for only a specific detection strategy. In contrast, we investigate learning-enabled attacks, and present trade-offs between the attacker's learning capability through observations, the controller detection strategy, and the control cost. Learning based attacks have been recently considered in [111, 113, 251]. In [111, 113], a variance based detection strategy has been investigated to present bounds on the probabilities of detection (or false alarm) of the attacker. In [251], an optimization-based controller is proposed that has the additional capability of injecting noise to interfere with the learning process of the attacker. Here, we consider a wider class of learning-based attacks and detection strategies, and provide tight trade-offs for these attacks.

Multiple variants of MITM attacks are studied in Reinforcement Learning (RL). In [172], the work studies the MITM attacks under the assumption that the attacker has perfect knowledge

of the underlying MDP. The results are further extended to the setting where attacker has no knowledge of the underlying MDP [173]. This is analogues to studying learning based attacks in RL where the attacker eavesdrops on the actions performed by the learner and manipulates the feedback from the environment. In [242], the work studies the feasibility of MITM attack under the constraint on the amount of contamination introduced by the attacker in the feedback signal. The relationship between the problem of designing optimal MITM attack in RL and the problem of designing optimal control is discussed in [250]. Finally, the learning based MITM attacks are also an active area of research in the Multi-Armed Bandits (MAB), see [97, 141, 27, 183]. In the same spirit of our work, these works study the feasibility of the attacks, and provide bounds on the amount of contamination needed by the attacker to achieve its objective. However, these works do not consider the possibility of the detection of the attacker. In this work, we focus on understanding the regime where the system can be attacked without the detection of the attacker.

8.3 Problem Setup

We consider the networked control system depicted in Figure 8.1 and Figure 8.2, where the plant dynamics are described by a discrete-time and linear time-invariant (LTI) system, namely at time $k \in \mathbb{N}$, we have

$$X_{k+1} = AX_k + U_k + W_k, \tag{8.1}$$

where X_k, U_k, W_k are vectors of dimension $M \times 1$ representing the plant state, control input, and plant disturbance respectively, and A is a matrix of dimension $M \times M$, representing the open-loop gain of the plant. At time k, the controller observes the feedback signal Y_k and generates a control signal U_k as a function of $Y_{1:k} = \{Y_1, \ldots, Y_k\}$. The initial state X_0 is known to both the controller and the attacker, and is independent of the disturbance sequence $\{W_k\}_{k=1}^{\infty}$, where W_k is i.i.d. Gaussian noise $\mathcal{N}(0, \sigma^2 I_M)$ with PDF known to both the parties, and I_M is the identity matrix of dimension $M \times M$. Our results can also be extended to the scenario where the PDF of

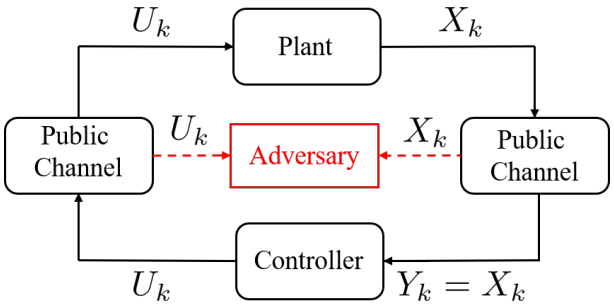


Figure 8.1. Exploration Phase.

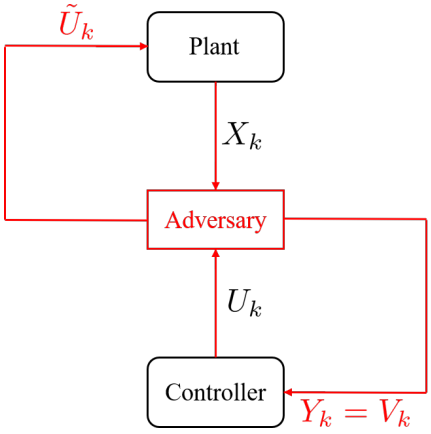


Figure 8.2. Exploitation Phase.

the noise known to the attacker is different from the actual PDF of the noise (or PDF known to the controller). With a slight loss of generality, we assume that $U_0 = W_0 = 0$ for analysis.

The controller attempts to detect the presence of the attacker based on the observations $Y_{1:k}$. When the controller detects an attack, it shuts the system down and prevents the attacker from causing further "damage" to the plant. The controller is aware of the plant dynamics in (8.1), and knows the gain A. This is justified because one can assume that the controller is tuned to the plant for a long duration and thus has knowledge of A to a great precision. On the other hand, the attacker only knows the form of the state evolution equation (8.1), but does not know the gain matrix A.

8.4 Learning based Attacks

We consider learning based attacks that evolve in two phases.

Phase 1: Exploration. Let L be the duration of the exploration phase. For all $k \leq L$, as illustrated in Figure 8.1, the attacker passively eavesdrops on the control input U_k and the plant state $Y_k = X_k$ with the objective of learning the open loop gain of the plant. We let $\hat{A}_k$ be the attacker's estimate of A at time step k. The duration L can be considered as the cost incurred by the attacker, since its actions are limited to eavesdropping during this phase.

Phase 2: Exploitation. The exploration phase is followed by the exploitation phase. For all $k \geq L+1$, as illustrated in Figure 8.2, the attacker hijacks the system and feeds a malicious control signal $\tilde{U}_k$ to the plant in order to destroy the plant. Additionally, the attacker may continue to learn about A, and utilizes its estimate $\hat{A}_k$ to design a fictitious feedback signal $Y_k = V_k$ in Figure 8.2 to deceive the controller, namely

$$V_{k+1} = \hat{A}_k V_k + U_k + \tilde{W}_k, \tag{8.2}$$

where for all $k \geq L+1$, $\tilde{W}_k$ are i.i.d. with $f_{\tilde{W}} = f_W = \mathcal{N}(0, \sigma^2 I_M)$. Let R denote an attack strategy whose feedback signal satisfies (8.2). Thus, for all $L > 0$, our class of learning based attacks is

$$\mathcal{A}(L) = \{R : \text{for all } k \leq L, Y_k = X_k \text{ and for all } k \geq L+1, Y_k = V_k\}. \tag{8.3}$$

Note that in the class $\mathcal{A}(L)$, the learning of A may or may not continue during the exploitation phase. Additionally, the attacker may use different learning algorithms in the two phases.

If the attacker learns A perfectly, i.e. $\hat{A}_k = A$, then (8.2) will perfectly mimic the plant behavior, making it impossible for the controller to detect the attacker. Otherwise, the controller can attempt to detect the presence of the attacker by testing for statistical deviations from the typical behavior in (8.1). The following example illustrates this point.

Example 1. *Let $R^* \in \mathcal{A}(L)$ be an attack whose learning is only limited to the exploration phase, namely $\hat{A}_k = \hat{A}_L$ for all $k \geq L+1$. Also, let $\|\cdot\|_{op}$ be the operator norm induced by the*

Euclidean norm $\|\cdot\|_2$ when applied to a matrix. In the exploration phase there is no interference from the attacker and for all $k \leq L$, the observation $Y_k = X_k$ satisfies

$$Y_{k+1} - AY_k - U_k = W_k \sim \text{ i.i.d. } f_W. \tag{8.4}$$

In the exploitation phase, for all $k \geq L+1$, the controller observation $Y_k = V_k$ satisfies

$$V_{k+1} - AV_k - U_k = V_{k+1} - AV_k + \hat{A}_L V_k - \hat{A}_L V_k - U_k = \tilde{W}_k + \left(\hat{A}_L - A\right) V_k, \tag{8.5}$$

where (8.5) *follows from* (8.4)*. Since $\tilde{W}_k$ and W_k have the same distribution and $\|Ax\|_2 \leq \|A\|_{op}\|x\|_2$ holds, the controller can test the statistical deviation of* (8.4) *from* (8.5)*. In this case, the detection of the attack is controlled by two factors: the estimation error $\|\hat{A}_L - A\|_{op}$ and the fictitious signal V_k.*

At the controller's side, the detection becomes easier when the error $\|\hat{A}_L - A\|_{op}$ increases. Thus, at the attacker's side it is desirable to reduce the error $\|\hat{A}_L - A\|_{op}$. This can be done by increasing the duration L, and incurring an additional learning cost.

The detection is also easier if the energy of the fictitious signal V_k is large. Since V_k is a function of the control signal U_{k-1}, it follows that the energy spent by the controller can help in the detection of the attacker.

We then conclude that the probability of successful detection (or the time required to detect the attacker with a given confidence) should reveal a trade-off between the duration L of the exploration phase (or the estimation error $\|\hat{A}_L - A\|_{op}$), and the energy of the fictitious signal (or of the control signal). In this paper we quantify both upper bound and lower bound on this trade-off.

8.4.1 Performance Measures

Definition 3. *The decision time τ is the time at which the controller makes a decision regarding the presence or absence of the attacker.*

Definition 4. *The probability of deception is the probability of the attacker deceiving the controller and remaining undetected at the decision time τ, namely $P_{\text{Dec}}^{\tau} \triangleq \mathbb{P}(\hat{\Theta}_\tau = 0 | \Theta_\tau = 1)$, where $\hat{\Theta}_\tau$ denotes the decision of the controller at the decision time τ, and the hijack indicator Θ_k at time k is*

$$\Theta_k \triangleq \begin{cases} 0, & \forall j \leq k : \ Y_j = X_j \, ; \\ 1, & \textit{otherwise}. \end{cases} \tag{8.6}$$

Likewise, the probability of false alarm is the probability of detecting the attacker when it is not present at the decision time τ, namely $P_{\text{FA}}^{\tau} \triangleq \mathbb{P}(\hat{\Theta}_\tau = 1 | \Theta_\tau = 0)$.

In the class $\mathcal{A}(L)$ in (8.3), for all $k \leq L$, we have that $\Theta_k = 0$ (exploration phase); and for all $k \geq L + 1$, we have $\Theta_k = 1$ (exploitation phase).

Definition 5. *For all attacks in the class $\mathcal{A}(L)$ and $0 < \epsilon < 1$, the ϵ-deception time $T(\epsilon)$ is the time required by the controller to make a decision, with $P_{\text{Dec}}^{\tau} \leq \epsilon \log(1/\epsilon)$, where $\tau = L + T(\epsilon) + 1$.*

Thus, $T(\epsilon)$ is the largest possible duration during which the attacker can deceive the controller, and remain undetected with confidence at least $1 - \epsilon \log(1/\epsilon)$, namely for all $L + 1 \leq k \leq T(\epsilon) + L$, we have

$$\mathbb{P}(\hat{\Theta}_k = \Theta_k | \Theta_k = 1) = \mathbb{P}(\hat{\Theta}_k = 1 | \Theta_k = 1) < 1 - \epsilon \log(1/\epsilon). \tag{8.7}$$

Definition 6. *For all $n > L$, the expected deception cost of the attacker until time n is defined as*

$$C(n) \triangleq \frac{1}{n} \mathbb{E}\Big[\sum_{k=L+1}^{n} \frac{V_{k-1}^T (\hat{A}_{k-1} - A)^T (\hat{A}_{k-1} - A) V_{k-1}}{2\sigma^2} \Big]. \tag{8.8}$$

8.4.2 Main results

We start with defining a non-divergent learning algorithm.

Definition 7. *A learning algorithm is non-divergent if its estimation error is non-increasing in the duration of the learning, namely for all $k_2 > k_1$, we have $\|\hat{A}_{k_2} - A\|_{op} \leq \|\hat{A}_{k_1} - A\|_{op}$.*

We introduce the following notation. Let $p_0(y_{1:\tau})$ be the conditional probability of $y_{1:\tau}$ given the attacker did not hijack the system, namely $\Theta_1 = \ldots \Theta_L = \Theta_{L+1} = \ldots \Theta_\tau = 0$, where $y_{1:\tau}$ denotes the realization of the random variables $Y_1, \ldots, Y_\tau$. Likewise, let $p_1(y_{1:\tau})$ be the conditional probability of $y_{1:\tau}$ given the attacker has hijacked the system, namely $\Theta_1 = \ldots = \Theta_L = 0$ and $\Theta_{L+1} = \ldots \Theta_\tau = 1$. The following proposition characterises the KL divergence $D(p_1(Y_{1:\tau})||p_0(Y_{1:\tau}))$ between $p_1(Y_{1:\tau})$ and $p_0(Y_{1:\tau})$, and is useful to derive our main results.

Proposition 3. *For all attacks in the class $\mathcal{A}(L)$ and $n > L$, the cumulative KL divergence is*

$$D(p_1(Y_{1:n})||p_0(Y_{1:n})) = nC(n). \tag{8.9}$$

The KL divergence between the distributions p_0 and p_1 is characterized by $C(n)$, and is the key quantity to establish both the lower bound and the upper bound on $T(\epsilon)$. If the PDF of the noise known to the attacker is different from the actual PDF of the noise (or the PDF known to the controller), Proposition 3 can be modified to include this discrepancy, and an additional non-negative term would be added to $C(n)$. The bounds on $T(\epsilon)$ will follow along the same lines.

The following theorem presents a lower bound on $\mathbb{E}[T(\epsilon)]$ that holds for any detection strategy. The bound is expressed in terms of $C(n)$, which depends on the attacker's learning algorithm, the fictitious signal and the control signal in (8.2).

Theorem 32. *For all attacks in $\mathcal{A}(L)$ and $\tau > L$, if*

$$P^\tau_{\text{Dec}} = O(|\epsilon \log \epsilon|) \text{ and } P^\tau_{FA} = O(|\epsilon \log \epsilon|), \text{ as } \epsilon \to 0, \tag{8.10}$$

then the deception time $T(\epsilon) = \tau - L - 1$ *is*

$$\mathbb{E}[T(\epsilon)] \geq \frac{\log(1/\epsilon)}{C(n_0)} + o(\log(1/\epsilon)) \quad \textit{as } \epsilon \to 0, \tag{8.11}$$

where $n_0 = \max\{n > L : nC(n) < \log(1/\epsilon)\}$.

It follows that for any detection strategy with probability of error $O(|\epsilon \log \epsilon|)$, the expected ϵ-deception time is at least $\Omega(\log(1/\epsilon)/C(n_0))$. The next theorem establishes that the lower bound in Theorem 32 is tight.

Theorem 33. *There exists an attack in* $\mathcal{A}(L)$ *and a detection strategy such that at* $\tau > L$*, we have*

$$P^{\tau}_{\text{Dec}} = O(\epsilon) \textit{ and } P^{\tau}_{FA} = O(\epsilon), \textit{ as } \epsilon \to 0, \tag{8.12}$$

and the deception time $T(\epsilon) = \tau - L - 1$ *is*

$$\mathbb{E}[T(\epsilon)] \leq \frac{\log(1/\epsilon)}{C(n_0 + 1)} + o(\log(1/\epsilon)), \quad \textit{as } \epsilon \to 0. \tag{8.13}$$

In Theorems 32 and 33, as $\epsilon \to 0$, we have that $C(n_0) \to C(n_0 + 1)$, and $|\epsilon| \leq |\epsilon \log \epsilon|$. Thus, the lower bound and the upper bound in Theorems 32 and 33 are tight. It turns out that the attack achieving the upper bound on $\mathbb{E}[T(\epsilon)]$ in Theorem 33 learns about A in the exploration phase only, and focuses on destabilizing the system in the exploitation phase. The corresponding detection strategy is a classic sequential probability ratio test ([222]), which computes the ratio of the posterior probability of the two hypotheses, namely the attacker is present or absent, and makes a decision when this ratio crosses the threshold $\log(1/\epsilon)$. While this strategy has been previously studied under the assumption that the samples $y_{1:n}$ are identically and independently distributed (i.i.d) ([45, 178, 177, 179]), here we extend the analysis to the samples dependent on both the control input and the state of the feedback signal at the controller.

We point out that to extend these results to non-linear systems, a key step would be finding an analogue of Proposition 3 in a non-linear setting. This proposition relates the KL divergence to the expected deception cost $C(n)$, which is a function of the fictitious signal and the error in the estimation of A. For non-linear systems, an equivalent relationship needs to be derived between the KL divergence, the fictitious signal and the error in the estimation of non-linear system dynamics. The proof of the Theorems 32 and 33 can then be obtained using a similar argument, given an analogue of Proposition 3 for non-linear systems.

Next, we derive some useful implications of Theorems 32 and 33. For simplicity of presentation, in the following we restrict the class of learning algorithms in the exploration phase, although results can also be extended to more general settings.

Definition 8. *A learning algorithm is said to be convergent if there exists an $\alpha \geq 1$ such that for all $\eta > 0$ and time step $k > 0$, we have*

$$\mathbb{P}(\|\hat{A}_k - A\|_{op} > \eta) \leq \frac{c}{(\eta^2 k)^\alpha}. \tag{8.14}$$

It follows that any convergent learning algorithm provides an unbiased estimate of A as the learning time $k \to \infty$, and the operator norm of the estimation error converges to the interval $[-\eta, +\eta]$ at rate $O(1/(\eta^2 k)^\alpha)$. There are many convergent learning algorithms. For example, for scalar systems and measurable control policy, the Least Squares (LS) algorithm in [184] satisfies

$$\mathbb{P}(|\hat{A}_k - a| > \eta) \leq \frac{2}{(1+\eta^2)^{k/2}}. \tag{8.15}$$

For the vector case sufficiently large learning time k, if the control input is $U_k = -\bar{K}X_k$ and $A - \bar{K}$ is a marginally stable matrix, then the LS algorithm in [203] satisfies

$$\mathbb{P}(\|\hat{A}_k - A\|_{op} > \eta) \leq \frac{c_1}{e^{\eta^2 k}}, \tag{8.16}$$

where $c_1 > 0$ is a constant.

The following theorem provides a lower bound on the duration of the exploration phase for the attacker to achieve a given expected ϵ-deception time.

Theorem 34. *For all $0 < \delta < 1$ and $D > 0$, and all attacks in $\mathcal{A}(L)$ using a convergent learning algorithm in the exploration phase and a non-divergent learning algorithm in the exploitation phase, if $\mathbb{E}[T(\epsilon)] \geq D + o(1)$ as $\epsilon \to 0$, then with probability at least $1 - \delta$ the following asymptotic inequality holds*

$$L \geq \frac{D\tilde{C}(n_0)}{\log(1/\epsilon)} \left(\frac{c}{\delta}\right)^{1/\alpha} + o\left(\frac{1}{\log(1/\epsilon)}\right), \text{ as } \epsilon \to 0, \tag{8.17}$$

where $\tilde{C}(n) = \mathbb{E}\big[\sum_{k=L+1}^{n} V_{k-1}^T V_{k-1}\big]/(2\sigma^2 n)$.

The following theorem establishes that the lower bound on L in Theorem 34 is order optimal, and a matching order-optimal bound on L holds for the LS algorithm in [203].

Theorem 35. *For all $0 < \delta < 1$ and $D > 0$, using the LS algorithm in [203] in the exploration phase only, and assuming the control input is $U_k = -\bar{K}X_k$, where $A - \bar{K}$ is a marginally stable matrix, if*

$$L = D\tilde{C}(n_0) \log(c_1/\delta)/\log(1/\epsilon) + o(1/\log(1/\epsilon)) \text{ as } \epsilon \to 0, \tag{8.18}$$

then, with probability at least $1 - \delta$ we have

$$\mathbb{E}[T(\epsilon)] \geq D + o(1), \text{ as } \epsilon \to 0. \tag{8.19}$$

The choice of the control policy can play a crucial role in the reduction of the deception time. However, this can occur at the expense of the energy used to construct the control signal U_k. The following theorem provides a lower bound on the amount of energy that the controller needs to spend to achieve a desired expected ϵ-deception time.

Theorem 36. *For all $D > 0$, and for all attacks in $\mathcal{A}(L)$ using a non-divergent learning algorithm in the exploitation phase, if $\mathbb{E}[T(\epsilon)] \leq D + o(1)$ as $\epsilon \to 0$, and for all $k > L$, the control policy satisfies*

$$\mathbb{E}[V_k^T \hat{A}_k^T \hat{A}_k V_k] + \sigma^2 + 2\mathbb{E}[V_k^T \hat{A}_k^T U_k] \leq 0, \tag{8.20}$$

then the expected energy of the control signal is

$$R(n_0) \geq \frac{2\sigma^2 \log(1/\epsilon)}{\|\hat{A}_L - A\|_{op}^2 D} + o(\log(1/\epsilon)), \text{ as } \epsilon \to 0, \tag{8.21}$$

where $R(n_0) \triangleq \mathbb{E}\big[\sum_{k=L}^{n_0-1} U_{k-1}^T U_{k-1}\big]/n_0$.

Theorem 36 shows that the expected energy of the control signal until a time between $L \leq k \leq n_0$ is inversely proportional to the upper bound D on the deception time. Since L is unknown to the controller, it follows that the controller should maintain a high level of expected signal energy $\mathbb{E}[U_k^2]$ at every time instance k to ensure a small deception time.

8.5 Simulations

In this section, we provide two numerical examples. Although our theoretical results are valid for a large class of learning algorithms and any detection strategy chosen by the controller, we validate them here using LS algorithm and a covariance detector.

First we start with an example for scalar system, where we use the empirical performance of a variance-test to illustrate our results. Specifically, at a decision time τ, the controller tests the empirical variance for unexpected behaviour over a detection window $[0, \tau]$, using a confidence interval $2\gamma > 0$ around the expected variance. More precisely, at decision time τ, $\hat{\Theta}_\tau = 0$ if

$$\frac{1}{\tau} \sum_{k=0}^{\tau} [Y_{k+1} - aY_k - U_k]^2 \in (\text{Var}[W] - \gamma, \text{Var}[W] + \gamma), \tag{8.22}$$

otherwise $\hat{\Theta}_\tau = 1$. In this case, since the system disturbances are i.i.d. Gaussian $\mathcal{N}(0, \sigma^2)$, using Chebyshev's inequality, we have

$$P_{\text{FA}}^{\tau} \leq \frac{\text{Var}[W^2]}{\gamma^2 T} = \frac{3\sigma^4}{\gamma^2 T}. \tag{8.23}$$

In our simulations, the attacker learns in the exploration phase only, and uses the LS learning algorithm. At the end of the exploration phase, we have

$$\hat{A}_L = \frac{\sum_{k=1}^{L-1}(X_{k+1} - U_k)X_k}{\sum_{k=1}^{L-1} X_k^2}. \tag{8.24}$$

Our simulation parameters are the following: $\gamma = 0.1$, decision time $\tau = 800$, $A = 1.1$, and $\{W_k\}$ are i.i.d. Gaussian $\mathcal{N}(0, 1)$. Using (8.23), the false-alarm rate is negligible for these parameters.

Figure 8.3 compares the attacker's success rate as a function of the duration L of the exploration phase for three different control policies $U_k = -AY_k + \Gamma_k$ such that for all k, I) $\Gamma_k = 0$, II) Γ_k are i.i.d. Gaussian $\mathcal{N}(0, 9)$, III) Γ_k are i.i.d. Gaussian $\mathcal{N}(0, 16)$. As illustrated in Figure 8.3, the attacker's success rate increases as the duration of exploration phase increases. This is because the attacker's estimation error $|\hat{A}_L - A|$ reduces as L increases, which makes it difficult for the controller to detect the attacker. This is in accordance with the theoretical findings in Theorem 34. Also, for a fixed L, the attacker's success rate decreases as the input control energy increases. The increase in the control energy increases the energy of the fictitious signal which increases the probability of detection, and is in accordance with Theorem 36.

Next, we provide an example of vector system, and analyze the empirical performance of the covariance test against the learning-based attack. In vector systems, the error matrix is

$$\Delta \triangleq \Sigma - \frac{1}{\tau}\sum_{k=1}^{\tau} \left[Y_{k+1} - AY_k - U_k\right]\left[Y_{k+1} - AY_k - U_k\right]^\top, \tag{8.25}$$

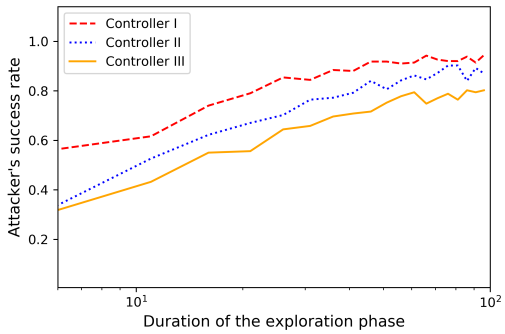


Figure 8.3. Attacker's success rate versus L.

Similar to (8.22), at decision time τ, we have $\hat{\Theta}_\tau = 0$ if $\|\Delta\|_{op} \leq \gamma$, and $\hat{\Theta}_\tau = 1$, otherwise. Similar to the scalar system, the attacker learns in the exploration phase only, and uses the LS learning algorithm, which implies that

$$\hat{A}_L = \begin{cases} 0_{n\times n}, & \det(G_{L-1}) = 0; \\ \sum_{k=1}^{L-1} (X_{k+1} - U_k) X_k^\top G_{L-1}^{-1}, & \text{otherwise}, \end{cases} \tag{8.26}$$

where $G_\tau \triangleq \sum_{k=1}^{\tau} X_k X_k^\top$. Our simulation parameters are the following: $\gamma = 0.1$, $A = [1\ \ 2\ ;\ 3\ \ 4]$, $\Sigma = [1\ \ 0\ ;\ 0\ \ 1]$, and $U_k = -0.9AY_k$.

Figure 8.4 compares the attacker's success rate, as a function of sizes of detection window τ for different duration L of the exploration phase. The false-alarm rate decreases to zero as the duration of the τ detection window tends to infinity, similarly to the argument for scalar systems. Thus, as the size of the detection window grows, the success rate of learning-based attacks increases. Finally, as as seen in Figure 8.4, as the duration of the exploration phase L increases, the attacker's success rate increases, since the attacker improves its estimate of A as L increases. This is in line with the theoretical findings in Theorem 34.

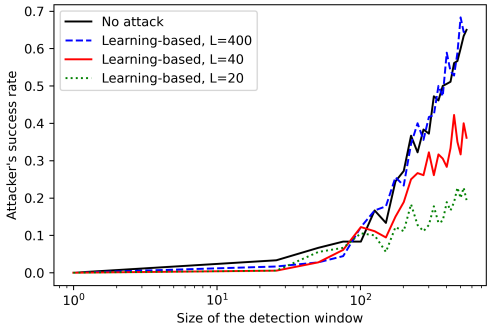


Figure 8.4. Attacker's success rate versus τ.

8.6 Conclusions and Future Directions

We have presented tight lower and upper bounds on the expected deception time for learning based MITM attacks, as the probability of correct detection tends to one. Additionally, we provided an order-optimal characterization of the length of the attacker's exploration phase and computed a lower bound on the control cost. In the future, we plan to study online phase learning based attacks, where the attacker can choose to switch between exploration and exploitation phases dynamically. We also plan to study methods to mitigate these attacks and render the system secure. The extension of our results to partially-observable linear vector systems where the input (actuation) gain is unknown, and the characterization of securable and unsecurable subspaces, similar to [193], is another possible research direction. Further extensions to nonlinear systems are also of interest.

8.7 Acknowledgement

Chapter 8, in full, is a reprint of the material as it appears in Anshuka Rangi, Mohammad Javad Khojasteh and Massimo Franceschetti, "Learning-based attacks in Cyber-Physical Systems: Exploration, Detection, and Control Cost trade-offs", *Learning for Dynamics and Control*, June 2021. The dissertation author was the co-primary investigator and co-author of this paper.

8.8 Appendix

8.8.1 Proof of Proposition 3

Proof. Since the attacker does not intervene before $k \leq L$, we have that for all $k \leq L$,

$$D(p_1(Y_{1:k})||p_0(Y_{1:k})) = 0. \tag{8.27}$$

Thus, for all $k > L$, using the chain rule, we have

$$D(p_1(Y_{1:n})||p_0(Y_{1:n})) = \sum_{k=L+1}^{n} D(p_1(Y_k|Y_{1:k-1})||p_0(Y_k|Y_{1:k-1})). \tag{8.28}$$

Also, if $\Theta_k = 1$, then for all $k > L$, we have

$$Y_k|(Y_{k-1}, U_{k-1}, \hat{A}_{k-1}) \sim \mathcal{N}(\hat{A}_{k-1}Y_{k-1} + U_{k-1}, \sigma^2 I_M), \tag{8.29}$$

since $Y_k = V_k$ for all $k > L$. Similarly, if $\Theta_k = 0$, then for all $k > L$, we have

$$Y_k|(Y_{k-1}, U_{k-1}, \hat{A}_{k-1}) \sim \mathcal{N}(AY_{k-1} + U_{k-1}, \sigma^2 I_M). \tag{8.30}$$

The result now follows by using the fact that for all $k > L$, we have $Y_k = V_k$.

We continue by noticing that the control input U_k lies in sigma field of past observations, namely U_k is measurable with respect to sigma field generated by $Y_{1:k-1}$. Thus, combining (8.28), (8.29) and (8.30), for all $k > L$, we have that

$$D(p_1(Y_k|Y_{1:k-1})||p_0(Y_k|Y_{1:k-1})) = \mathbb{E}\left[\frac{Y_{k-1}^T(\hat{A}_{k-1} - A)^T(\hat{A}_{k-1} - A)Y_{k-1}}{2\sigma^2}\right]. \tag{8.31}$$

Using (8.28) and (8.31), for all $n > L$, we have

$$D(p_1(Y_{1:n})||p_0(Y_{1:n})) = \mathbb{E}\Big[\sum_{k=L+1}^{n} \frac{Y_{k-1}^T(\hat{A}_{k-1} - A)^T(\hat{A}_{k-1} - A)Y_{k-1}}{2\sigma^2}\Big]. \tag{8.32}$$

□

8.8.2 Proof of the Theorem 32

Proof. The proof of the theorem consists of two parts. First, for all attacks in the class $\mathcal{A}(L)$ and $0 < c < 1$, we show that if the probability of error is small, namely $\mathbb{P}(\hat{\Theta}_\tau \neq \Theta_\tau) = O(|\epsilon \log \epsilon|)$, then the log-likelihood ratio $\log(p_1(y_{1:\tau})/p_0(y_{1:\tau}))$ should be greater than $(1-c)\log(1/\epsilon)$ with high probability as $\epsilon \to 0$, namely

$$\log \frac{p_1(y_{1:\tau})}{p_0(y_{1:\tau})} \geq (1-c)\log(1/\epsilon) \tag{8.33}$$

must hold with high probability, as $\epsilon \to 0$. Second, we show that there exists $0 < \bar{c} < 1$ such that for all $0 < c \leq \bar{c}$ and $T(\epsilon) < (1-c)\log(1/\epsilon)/C(n_0)$, it is unlikely that the inequality in (8.33) is satisfied.

Using (8.10), for all $k \geq L+1$, we have that both type I and type II errors of the hypothesis test $\Theta_k = 1$ vs. $\Theta_k = 0$ are $O(|\epsilon \log \epsilon|)$. Thus, using [45, Lemma 4], for all $0 < c < 1$, we have

$$\mathbb{P}\Big(S^\tau \leq -(1-c)\log \epsilon\Big) = O(-\epsilon^c \log \epsilon), \tag{8.34}$$

where

$$S^n = \log \frac{p_1(y_{1:n})}{p_0(y_{1:n})} = \sum_{k=1}^{n} \log\left(\frac{p_1(y_k|y_{1:k-1})}{p_0(y_k|y_{1:k-1})}\right). \tag{8.35}$$

Therefore, as $\epsilon \to 0$, the probability in (8.34) tends to 0, which concludes the first part of the proof.

Now, we show that for all $0 < c < 1$, we have

$$\lim_{n' \to \infty} \mathbb{P}\left(\max_{1 \le k \le n'} S^k \ge (D(p_1(y_{1:n'})||p_0(y_{1:n'})) + n'c) \right) = 0, \tag{8.36}$$

where $D(p_1(y_{1:n'})||p_0(y_{1:n'}))$ denotes the KL divergence between the distributions p_1 and p_0 of $Y_{1:n'}$. We have

$$\begin{aligned} S^n &= \sum_{k=1}^{n} \left(\log\left(\frac{p_1(y_k|y_{1:k-1})}{p_0(y_k|y_{1:k-1})}\right) - D(p_1(Y_k|Y_{1:k-1})||p_0(Y_k|Y_{1:k-1})) \right) \\ &\quad + \sum_{k=1}^{n} D(p_1(Y_k|Y_{1:k-1})||p_0(Y_k|Y_{1:k-1})) \\ &= M_1^n + M_2^n, \end{aligned} \tag{8.37}$$

where

$$M_1^n = \sum_{k=1}^{n} \left(\log\left(\frac{p_1(y_k|y_{1:k-1})}{p_0(y_k|y_{1:k-1})}\right) - D(p_1(Y_k|Y_{1:k-1})||p_0(Y_k|Y_{1:k-1})) \right), \tag{8.38}$$

is a martingale with mean 0 with respect to filtration $\mathcal{F}_k = \sigma(Y_{1:k-1})$, and

$$\begin{aligned} M_2^n &= \sum_{k=1}^{n} D(p_1(Y_k|Y_{1:k-1})||p_0(Y_k|Y_{1:k-1})), \\ &\overset{(a)}{=} D(p_1(Y_{1:n})||p_0(Y_{1:n})), \end{aligned} \tag{8.39}$$

where (a) follows from the chain rule of KL-Divergence. Now, if the event in (8.36) occurs for a fixed n_1, namely

$$M_1^{n_1} + M_2^{n_1} \ge D(p_1(Y_{1:n_1})||p_0(Y_{1:n_1})) + n_1 c, \tag{8.40}$$

then it implies that $M_1^{n_1} \ge n_1 c$. Since $Y_k|Y_{1:k-1}$ has a normal distribution using (8.29) and (8.30),

there exists a constant $b > 0$ such that the probability in (8.36) simplifies as

$$\mathbb{P}\left(\max_{1\leq k\leq n'} S^k \geq (D(p_1(y_{1:n'})||p_0(y_{1:n'})) + n'c)\right) \leq \mathbb{P}(\max_{1\leq k\leq n'} M_1^k \geq n'c) \overset{(a)}{\leq} b/n'c^2, \tag{8.41}$$

where (a) follows from the Doob-Kolmogorov extension of Chebyshev's inequality in [56], and the fact that M_1^k is a martingale with 0 mean. Hence, we have that (8.36) follows.

Now, we have

$$n_0 C(n_0) < \log(1/\epsilon). \tag{8.42}$$

Therefore, there exists $0 < \bar{c} < 1$ such that

$$n_0 C(n_0) + n_0\bar{c} = (1 - \bar{c})\log(1/\epsilon). \tag{8.43}$$

Now, using Proposition 3, for all $0 < c \leq \bar{c}$, we have

$$\begin{aligned}\mathbb{P}(N \leq n_0) &\leq \mathbb{P}\Big(N \leq n_0 \text{ and } S^N \geq n_0(C(n_0) + c)\Big) \\ &\quad + \mathbb{P}\Big(S^N \leq n_0(C(n_0) + c)\Big) \\ &\leq \mathbb{P}\Big(\max_{1\leq k\leq n_0} S^k \geq n_0(C(n_0) + c)\Big) \\ &\quad + \mathbb{P}\Big(S^N \leq n_0(C(n_0) + c)\Big),\end{aligned} \tag{8.44}$$

and the first and the second terms at the right-hand side of (8.44) approach zero by (8.36) and (8.34), respectively. □

8.8.3 Proof of the Theorem 33

Proof. In $\mathcal{A}(L)$, consider an attack R^* such that for all $k > L$, we have $\hat{A}_k = \hat{A}_L$. For all $k > L$, if $\Theta_k = 1$, then we have

$$Y_k|Y_{1:k-1} \sim \mathcal{N}(\hat{A}_L Y_{k-1} + U_{k-1}, \sigma^2 I_M). \tag{8.45}$$

Similarly, if $\Theta_k = 0$, then

$$Y_k|Y_{1:k-1} \sim \mathcal{N}(AY_{k-1} + U_{k-1}, \sigma^2 I_M). \tag{8.46}$$

Consider a the following detection strategy, also known as Sequential Probability Ratio Test (SPRT), at the controller as follows. At time n, if

$$\sum_{k=1}^{n} \log\left(\frac{p_1(y_k|y_{1:k-1})}{p_0(y_k|y_{1:k-1})}\right) \geq \log(1/\epsilon), \tag{8.47}$$

then $\hat{\Theta}_n = 1$, and if

$$\sum_{k=1}^{n} \log\left(\frac{p_0(y_k|y_{1:k-1})}{p_1(y_k|y_{1:k-1})}\right) \geq \log(1/\epsilon), \tag{8.48}$$

then $\hat{\Theta}_n = 0$. Otherwise, n is not a decision time and the test continues.

We will show that for the attack R^* and the detection strategy SPRT, the statement of the theorem holds.

For SPRT, the probability of error, both P^{τ}_{Dec} and P^{τ}_{FA}, is at most $O(\epsilon)$, and the proof is along the same direction as [179, Theorem 1]. Now, let us prove the bound on $T(\epsilon)$. Given the system is under attack, let the decision time τ of SPRT be

$$T = \min\left\{n : \sum_{k=1}^{n} \log\left(\frac{p_1(y_k|y_{1:k-1})}{p_0(y_k|y_{1:k-1})}\right) \geq \log(1/\epsilon)\right\}. \tag{8.49}$$

Using [45, Lemma 2], for system under attack $\mathcal{A}(L)$ and for all $c > 0$, there exist a $b > 0$ such

that

$$\mathbb{P}\Bigg(\sum_{k=1}^{n}\log\left(\frac{p_1(y_k|y_{1:k-1})}{p_0(y_k|y_{1:k-1})}\right) < (D(p_1(Y_{1:n})||p_0(Y_{1:n})) - nc)\Bigg) \leq e^{-bn}. \tag{8.50}$$

Using the definition of n_0, for all $\bar{n} > n_0$ we have

$$\log(1/\epsilon) \leq \bar{n}C(\bar{n}) = D(p_1(Y_{1:\bar{n}})||p_0(Y_{1:\bar{n}})), \tag{8.51}$$

where the equality follows from Proposition 3. Using (8.50) and (8.51), For all $c > 0$ and $n \geq (1+c)(n_0+1)\log(1/\epsilon)/D(p_1(Y_{1:n_0+1})||p_0(Y_{1:n_0+1}))$, we have

$$\mathbb{P}\Bigg(\sum_{k=1}^{n}\log\left(\frac{p_1(y_k|y_{1:k-1})}{p_0(y_k|y_{1:k-1})}\right) < \log(1/\epsilon)\Bigg) \leq e^{-bn}. \tag{8.52}$$

Then, using Proposition 3, the statement of the theorem follows. □

8.8.4 Proof of Theorem 34

Proof. If the learning algorithm in the exploration phase is a convergent algorithm, the learning algorithm in the exploitation phase is a non-divergent algorithm, then for all $0 < \delta < 1$, we have

$$\begin{aligned} C(n_0) &\overset{(a)}{\leq} \|\hat{A}_L - A\|_{op}^2 \frac{1}{n_0}\mathbb{E}\Bigg[\sum_{k=L+1}^{n_0}\frac{V_{k-1}^T V_{k-1}}{2\sigma^2}\Bigg], \\ &\overset{(b)}{\leq} \left(\frac{c^{1/\alpha}}{L\delta^{1/\alpha}}\right)\tilde{C}(n_0), \end{aligned} \tag{8.53}$$

with probability at least $1-\delta$, where (a) follows from the fact that

$$||Ax||_2 \leq ||A||_{op}||x||_2, \tag{8.54}$$

and the learning algorithm in the exploitation phase is non-divergent, and (b) follows from Definition 8 of convergent algorithms. Thus, we have

$$\frac{\log(1/\epsilon)}{C(n_0)} \geq \frac{\log(1/\epsilon)}{\tilde{C}(n_0)}\left(\frac{L\delta^{1/\alpha}}{c^{1/\alpha}}\right), \tag{8.55}$$

with probability at least $1-\delta$. Using Theorem 32 and (8.55), if

$$(1+o(1))\frac{\log(1/\epsilon)}{\tilde{C}(n_0)}\left(\frac{L\delta^{1/\alpha}}{c^{1/\alpha}}\right) > D(1+o(1)), \text{ as } \epsilon \to 0, \tag{8.56}$$

then $\mathbb{E}[T(\epsilon)] > D + o(1)$ as $\epsilon \to 0$. This along with (8.55) implies that

$$L \geq \frac{(1+o(1))D\tilde{C}(n_0)}{\log(1/\epsilon)}\frac{c^{1/\alpha}}{\delta^{1/\alpha}}, \text{ as } \epsilon \to 0, \tag{8.57}$$

with probability at least $1-\delta$. □

8.8.5 Proof of Theorem 35

Proof. Consider the LS learning algorithm in [203] which satisfies

$$\mathbb{P}(\|\hat{A}_k - A\|_{op} > \eta) \leq \frac{c_1}{e^{\eta^2 k}}, \tag{8.58}$$

For $\eta = \sqrt{\log(c_1/\delta)/L}$, similar to (8.53), we have that

$$C(n_0) \leq \frac{\log(c_1/\delta)}{L}\tilde{C}(n_0), \tag{8.59}$$

with probability at least $1-\delta$. Thus, we have

$$\frac{\log(1/\epsilon)}{C(n_0)} \geq \frac{\log(1/\epsilon)}{\tilde{C}(n_0)}\frac{L}{\log(c_1/\delta)}, \tag{8.60}$$

with probability at least $1-\delta$. Thus, for $L = (1+o(1))D\tilde{C}(n_0)\log(c_1/\delta)/\log(1/\epsilon)$ as $\epsilon \to 0$, using Theorem 32, we have that

$$\mathbb{E}[T(\epsilon)] \geq \frac{(1+o(1))\log(1/\epsilon)}{C(n_0)} \geq D(1+o(1)) = D + o(1), \text{ as } \epsilon \to 0, \tag{8.61}$$

with probability at least $1-\delta$. The statement of the theorem follows. □

8.8.6 Proof of Theorem 36

Proof. Since $\tilde{W}_k$ is independent of U_k and V_k and $\mathbb{E}[\tilde{W}_k] = 0$, we have

$$\mathbb{E}[V_{k+1}^T V_{k+1}] - \mathbb{E}[U_k^T U_k] = \mathbb{E}[V_k^T \hat{A}_k^T \hat{A}_k V_k] + \sigma^2 + 2\mathbb{E}[V_k^T \hat{A}_k^T U_k]. \tag{8.62}$$

Using (8.20), we have

$$\mathbb{E}[V_{k+1}^T V_{k+1}] \leq \mathbb{E}[U_k^T U_k], \tag{8.63}$$

which implies

$$\begin{aligned} C(n_0) &\overset{(a)}{\leq} \frac{\|\hat{A}_L - A\|_{op}^2}{n_0} \mathbb{E}\left[\sum_{k=L+1}^{n_0} \frac{V_{k-1}^T V_{k-1}}{2\sigma^2}\right] \\ &\overset{(b)}{\leq} \frac{\|\hat{A}_L - A\|_{op}^2}{n_0} \mathbb{E}\left[\sum_{k=L}^{n_0-1} \frac{U_{k-1}^T U_{k-1}}{2\sigma^2}\right], \end{aligned} \tag{8.64}$$

where (a) follows from the fact that $||Ax||_2 \leq ||A||_{op}||x||_2$, and (b) follows from (8.63). Since $\mathbb{E}[T(\epsilon)] \leq D + o(1)$ as $\epsilon \to 0$, using Theorem 32 and (8.64), we have that

$$D + o(1) \geq \frac{(1+o(1))2\sigma^2 \log(1/\epsilon)}{\|\hat{A}_L - A\|_{op}^2 R(n_0)}, \text{ as } \epsilon \to 0. \tag{8.65}$$

Hence, the statement of the theorem follows. □

Chapter 9

Non-Stochastic Information Theory

9.1 Introduction

This paper introduces elements of a non-stochastic information theory that parallels Shannon's probabilistic theory of information, but that provides strict deterministic guarantees for every codeword transmission. When Shannon laid the mathematical foundations of communication theory he embraced a probabilistic approach [199]. A tangible consequence of this choice is that in today's communication systems performance is guaranteed in an average sense, or with high probability. Occasional violations from a specification are permitted, and cannot be avoided. This approach is well suited for consumer-oriented digital communication devices, where the occasional loss of data packets is not critical, and made Shannon's theory the golden standard to describe communication limits, and to construct codes that achieve these limits. The probabilistic approach, however, has also prevented Shannon's theory to be relevant in systems where occasional decoding errors can result in catastrophic failures; or in adversarial settings, where the behavior of the channel may be unknown and cannot be described by a probability distribution. The basic consideration that is the leitmotiv of this paper is that the probabilistic framework is not a fundamental component of Shannon's theory, and that the path laid by Shannon's work can be extended to embrace a non-stochastic setting.

The idea of adopting a non-stochastic approach in information theory is not new. A few years after introducing the notion of capacity of a communication system [199], Shannon

introduced the zero-error capacity [198]. While the first notion corresponds to the largest rate of communication such that the probability of decoding error *tends to zero*, the second corresponds to the largest rate of communication such that the probability of decoding error *equals zero*. Both definitions of capacity satisfy coding theorems: Shannon's channel coding theorem states that the capacity is the supremum of the mutual information between the input and the output of the channel [199]. Nair introduced a non-stochastic mutual information functional and established an analogous coding theorem for the zero-error capacity in a non-stochastic setting [158]. While Shannon's theorem leads to a single letter expression, Nair's result is multi-letter, involving the non-stochastic information between codeword blocks of n symbols. The zero-error capacity can also be formulated as a graph-theoretic property and the absence of a single-letter expression for general graphs is well known [198, 187]. Extensions of Nair's nonstochastic approach to characterize the zero-error capacity in the presence of feedback from the receiver to the transmitter using nonstochastic directed mutual information have also been considered [157].

A parallel non-stochastic approach is due to Kolmogorov who, motivated by Shannon's results, introduced the notions of ϵ-entropy and ϵ-capacity in the context of functional spaces [209]. He defined the ϵ-entropy as the logarithm base two of the *covering number* of the space, namely the logarithm of the minimum number of balls of radius ϵ that can cover the space. Determining this number is analogous to designing a *source codebook* such that the distance between any signal in the space and a codeword is at most ϵ. In this way, any transmitted signal can be represented by a codeword point with at most ϵ-distortion. Notions related to the ϵ-entropy are the Hartley entropy [80] and the Rényi differential (0th-order) entropy [186]. They arise for random variables with known range but unknown distribution, and are defined by taking the logarithm of the cardinality (for discrete variables), or Lebesgue measure (for continuous variables) of their range. Thus, their definition does not require any statistical structure. Using these entropies, non-stochastic measures of mutual information have been constructed [200, 117]. Unfortunately, the absence of coding theorems makes the operational significance of these definitions lacking.

Rather than using mutual information and entropy, Kolmogorov gave an operational definition of the ϵ-capacity as the logarithm base two of the *packing number* of the space, namely the logarithm of the maximum number of balls of radius ϵ that can be placed in the space without overlap. Determining this number is analogous to designing a *channel codebook* such that the distance between any two codewords is at least 2ϵ. In this way, any transmitted codeword that is subject to a perturbation of at most ϵ can be recovered at the receiver without error. It follows that the ϵ-capacity corresponds to the zero-error capacity of an additive channel having arbitrary, bounded noise of support at most $[0, \epsilon]$. Lim and Franceschetti extended this concept introducing the (ϵ, δ) capacity [132], defined as the logarithm base two of the largest number of balls of radius ϵ that can be placed in the space with average codeword overlap of at most δ. In this setting, δ measures the amount of error that can be tolerated when designing a codebook in a non-stochastic setting. Neither the Kolmogorov capacity, nor its (ϵ, δ) generalization have a corresponding information-theoretic characterization in terms of mutual information and an associated coding theorem. This is offered in the present paper. Some possible applications of non-stochastic approaches arising in the context of estimation, control, security, communication over non-linear optical channels, and robustness of neural networks are described in [188, 189, 230, 28, 228, 218, 65]; and some are also discussed in the context of the presented theory at the end of the paper.

The rest of the paper is organized as follows. Section II provides a summary of our contributions; Section III introduces the mathematical framework of non-stochastic uncertain variables that is used throughout the paper. Section IV introduces the concept of non-stochastic mutual information. Section V gives an operational definition of capacity of a communication channel and relates it to the mutual information. Section VI extends results to more general channel models; and section VII concentrates on the special case of stationary, memoryless, uncertain channels. Sufficient conditions are obtained to obtain single-letter expressions for this case. Section VIII considers some examples of channels and computes the corresponding capacity. Finally, Section IX discusses some possible application of the developed theory, and

Section X draws conclusions and discusses future directions. A subset of the results has been presented in [176].

9.2 Contributions

We introduce a notion of δ-mutual information between non-stochastic, uncertain variables. In contrast to Nair's definition [158], which only allows to measure information with *full confidence*, our definition considers the information revealed by one variable regarding the other *with a given level of confidence*. We then introduce a notion of (ϵ, δ)-capacity, defined as the logarithm base two of the largest number of balls of radius ϵ that can be placed in the space such that the overlap between any two balls is at most a ratio of δ and the total number of balls. In contrast to the definition of Lim and Franceschetti [132], which requires the average overlap among all the balls to be at most δ, our definition requires to bound the overlap between any pair of balls. For $\delta = 0$, our capacity definition reduces to the Kolmogorov ϵ-capacity, or equivalently to the zero-error capacity of an additive, bounded noise channel, and our mutual information definition reduces to Nair's one [158]. We establish a channel coding theorem in this non-stochastic setting, showing that the largest mutual information, with confidence at least $(1 - \delta)$, between a transmitted codeword and its received version corrupted with noise at most ϵ, is the (ϵ, δ)-capacity. We then extend this result to more general non-stochastic channels, where the noise is expressed in terms of a set-valued map $N(\cdot)$ associating each transmitted codeword to a noise region in the received codeword space, that is not necessarily a ball of radius ϵ.

Next, we consider the class of non-stochastic, memoryless, stationary uncertain channels. In this case, the noise $N(\cdot)$ experienced by a codeword of n symbols factorizes into n identical terms describing the noise experienced by each codeword symbol. This is the non-stochastic analogous of a discrete memoryless channel (DMC), where the current output symbol depends only on the current input symbol and not on any of the previous input symbols, and the noise distribution is constant across symbol transmissions. It differs from Kolmogorov's ϵ-noise

channel, where the noise experienced by one symbol affects the noise experienced by other symbols. In Kolmogorov's setting, the noise occurs within a ball of radius ϵ. It follows that for any realization where the noise along one dimension (*viz.* symbol) is close to ϵ, the noise experienced by all other symbols lying in the remaining dimensions must be close to zero. In contrast, for non-stochastic, memoryless, stationary channels, the noise experienced by any transmitted symbol is described by a single, non-stochastic set-value map from the transmitted alphabet to the received symbol space. We provide coding theorems in this setting in terms of the δ-mutual information rate between received and transmitted codewords. Finally, we provide sufficient conditions for the factorization of the mutual information and to obtain a single-letter expression for the non-stochastic capacity of stationary, memoryless, uncertain channels. We provide examples in which these conditions are satisfied and compute the corresponding capacity, and we conclude with a discussion of some possible applications of the presented theory.

9.3 Uncertain variables

We start by reviewing the mathematical framework used in [158] to describe non-stochastic uncertain variables (UVs). An UV X is a mapping from a sample space Ω to a set $\mathscr{X}$, i.e. for all $\omega \in \Omega$, we have $x = X(\omega) \in \mathscr{X}$. Given an UV X, the marginal range of X is

$$[\![X]\!] = \{X(\omega) : \omega \in \Omega\}. \tag{9.1}$$

The joint range of two UVs X and Y is

$$[\![X, Y]\!] = \{(X(\omega), Y(\omega)) : \omega \in \Omega\}. \tag{9.2}$$

Given an UV Y, the conditional range of X given $Y = y$ is

$$[\![X|y]\!] = \{X(\omega) : Y(\omega) = y, \omega \in \Omega\}, \tag{9.3}$$

and the conditional range of X given Y is

$$[\![X|Y]\!] = \{[\![X|y]\!] : y \in [\![Y]\!]\}. \tag{9.4}$$

Thus, $[\![X|Y]\!]$ denotes the uncertainty in X given the realization of Y and $[\![X, Y]\!]$ represents the total joint uncertainty of X and Y, namely

$$[\![X, Y]\!] = \cup_{y \in [\![Y]\!]} [\![X|y]\!] \times \{y\}. \tag{9.5}$$

Finally, two UVs X and Y are independent if for all $x \in [\![X]\!]$

$$[\![Y|x]\!] = [\![Y]\!], \tag{9.6}$$

which also implies that for all $y \in [\![Y]\!]$

$$[\![X|y]\!] = [\![X]\!]. \tag{9.7}$$

9.4 δ-Mutual information

9.4.1 Uncertainty function

We now introduce a class of functions that are used to express the amount of uncertainty in determining one UV given another. In our setting, an uncertainty function associates a positive number to a given set, which expresses the "massiveness" or "size" of that set.

Definition 9. *Given any set $\mathscr{X}$, $m_{\mathscr{X}} : \mathscr{S} \subseteq \mathscr{X} \to \mathbb{R}_0^+$ is an uncertainty function if it is finite and strongly transitive, namely:*
For all $\mathscr{S} \subseteq \mathscr{X}, \mathscr{S} \neq \emptyset$, we have

$$0 < m_{\mathscr{X}}(\mathscr{S}) < \infty, \; m_{\mathscr{X}}(\emptyset) = 0. \tag{9.8}$$

For all $\mathscr{S}_1, \mathscr{S}_2 \subseteq \mathscr{X}$, we have

$$\max\{m_{\mathscr{X}}(\mathscr{S}_1), m_{\mathscr{X}}(\mathscr{S}_2)\} \leq m_{\mathscr{X}}(\mathscr{S}_1 \cup \mathscr{S}_2). \tag{9.9}$$

In the case $\mathscr{X}$ is measurable, an uncertainty function can easily be constructed using a measure. In the case $\mathscr{X}$ is a bounded (not necessarily measurable) metric space and the input set $\mathscr{S}$ contains at least two points, an example of uncertainty function is the diameter.

9.4.2 Association and dissociation between UVs

We now introduce notions of association and dissociation between UVs. In the following definitions, we let $m_{\mathscr{X}}(.)$ and $m_{\mathscr{Y}}(.)$ be uncertainity functions defined over sets $\mathscr{X}$ and $\mathscr{Y}$ corresponding to UVs X and Y. We use the notation $\mathscr{A} \succ \delta$ to indicate that for all $a \in \mathscr{A}$ we have $a > \delta$. Similarly, we use $\mathscr{A} \preceq \delta$ to indicate that for all $a \in \mathscr{A}$ we have $a \leq \delta$. For $\mathscr{A} = \emptyset$, we assume $\mathscr{A} \preceq \delta$ is always satisfied, while $\mathscr{A} \succ \delta$ is not. Whenever we consider $i \neq j$, we also assume that $y_i \neq y_j$ and $x_i \neq x_j$.

Definition 10. *The sets of association for UVs X and Y are*

$$\mathscr{A}(X;Y) = \left\{ \frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!]|)}{m_{\mathscr{X}}([\![X]\!])} : y_1, y_2 \in [\![Y]\!] \right\} \setminus \{0\}, \tag{9.10}$$

$$\mathscr{A}(Y;X) = \left\{ \frac{m_{\mathscr{Y}}([\![Y|x_1]\!] \cap [\![Y|x_2]\!]|)}{m_{\mathscr{Y}}([\![Y]\!])} : \ x_1, x_2 \in [\![X]\!] \right\} \setminus \{0\}. \tag{9.11}$$

Definition 11. *For any $\delta_1, \delta_2 \in [0,1)$, UVs X and Y are disassociated at levels (δ_1, δ_2) if the following inequalities hold:*

$$\mathscr{A}(X;Y) \succ \delta_1, \tag{9.12}$$

$$\mathscr{A}(Y;X) \succ \delta_2, \tag{9.13}$$

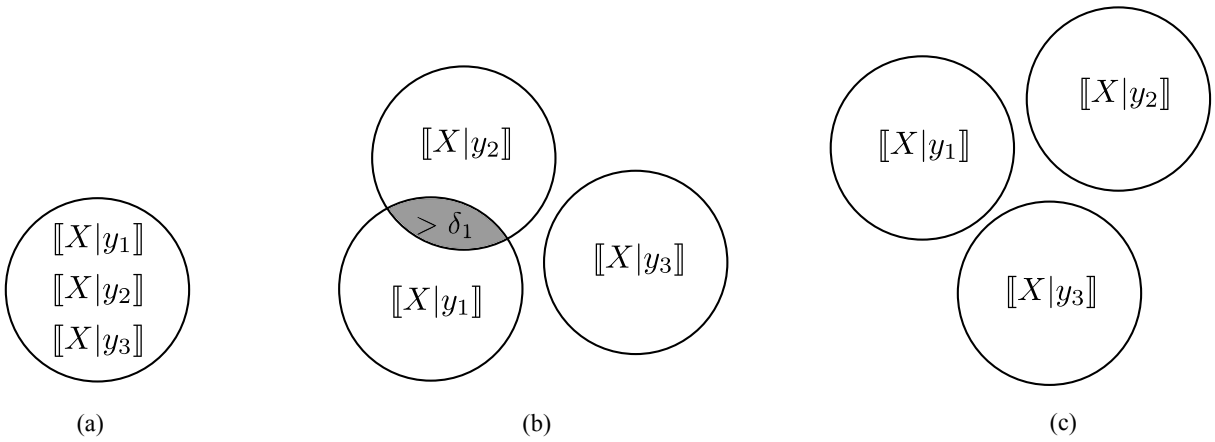


Figure 9.1. Illustration of disassociation between UVs. Case (a): variables are maximally disassociated and all conditional ranges completely overlap. Case (b): variables are disassociated at some levels (δ_1, δ_2), and there is some overlap between at least two conditional ranges. Case (c): variables are not disassociated at any levels, and there is no overlap between the conditional ranges.

and this case we write $(X, Y) \overset{d}{\leftrightarrow} (\delta_1, \delta_2)$.

Having UVs X and Y be disassociated at levels (δ_1, δ_2) indicates that at least two conditional ranges $[\![X|y_1]\!]$ and $[\![X|y_2]\!]$ have nonzero overlap, and that given any two conditional ranges, either they do not overlap or the uncertainty associated to their overlap is greater than a δ_1 fraction of the total uncertainty associated to $[\![X]\!]$; and that the same holds for conditional ranges $[\![Y|x_1]\!]$ and $[\![Y|x_2]\!]$ and level δ_2. The levels of disassociation can be viewed as lower bounds on the amount of residual uncertainty in each variable when the other is known. If X and Y are independent, then all the conditional ranges completely overlap, $\mathscr{A}(X;Y)$ and $\mathscr{A}(Y;X)$ contain only the element one, and the variables are maximally disassociated (see Figure 9.1a). In this case, knowledge of Y does not reduce the uncertainty of X, and vice versa. On the other hand, when the uncertainty associated to any of the non-zero intersections of the conditional ranges decreases, but remains positive, then X and Y become less disassociated, in the sense that knowledge of Y can reduce the residual uncertainty of X, and vice versa (see Figure 9.1b). When the intersection between every pair of conditional ranges becomes empty, the variables cease being disassociated (see Figure 9.1c).

An analogous definition of association is given to provide upper bounds on the residual uncertainty of one uncertain variable when the other is known.

Definition 12. *For any $\delta_1, \delta_2 \in [0,1]$, we say that UVs X and Y are associated at levels (δ_1, δ_2) if the following inequalities hold:*

$$\mathscr{A}(X;Y) \preceq \delta_1, \tag{9.14}$$

$$\mathscr{A}(Y;X) \preceq \delta_2, \tag{9.15}$$

and in this case we write $(X,Y) \overset{a}{\leftrightarrow} (\delta_1, \delta_2)$.

The following lemma provides necessary and sufficient conditions for association at given levels to hold. These conditions are stated for all points in the marginal ranges $[\![Y]\!]$ and $[\![X]\!]$. They show that in the case of association one can also include in the definition the conditional ranges that have zero intersection. This is not the case for disassociation.

Lemma 31. *For any $\delta_1, \delta_2 \in [0,1]$, $(X,Y) \overset{a}{\leftrightarrow} (\delta_1, \delta_2)$ if and only if for all $y_1, y_2 \in [\![Y]\!]$, we have*

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!]|)}{m_{\mathscr{X}}([\![X]\!])} \leq \delta_1, \tag{9.16}$$

and for all $x_1, x_2 \in [\![X]\!]$, we have

$$\frac{m_{\mathscr{Y}}([\![Y|x_1]\!] \cap [\![Y|x_2]\!]|)}{m_{\mathscr{Y}}([\![Y]\!])} \leq \delta_2. \tag{9.17}$$

Proof. The proof is given in Appendix 9.12.1. □

An immediate, yet important consequence of our definitions is that both association and disassociation at given levels (δ_1, δ_2) cannot hold simultaneously. We also have that, given any two UVs, one can always choose δ_1 and δ_2 to be large enough such that they are associated at levels (δ_1, δ_2). In contrast, as the smallest value in the sets $\mathcal{A}(X;Y)$ and $\mathcal{A}(Y;X)$ tends to

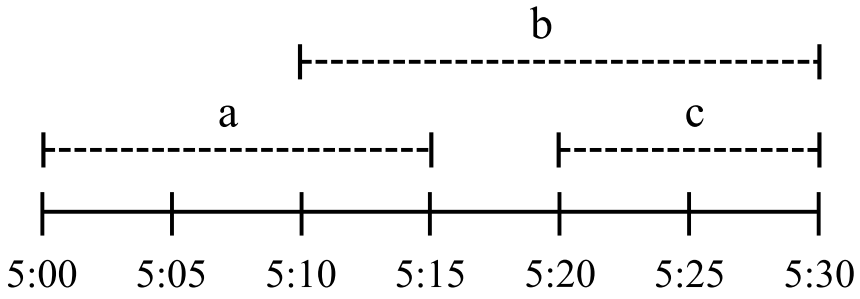


Figure 9.2. Illustration of the possible time intervals for the walkers on the path.

zero, the variables eventually cease being disassociated. Finally, it is possible that two uncertain variables are neither associated nor disassociated at given levels (δ_1, δ_2).

Example 2. *Consider three individuals a, b and c going for a walk along a path. Assume they take at most 15, 20 and 10 minutes to finish their walk, respectively. Assume a starts walking at time 5:00, b starts walking at 5:10 and c starts walking at 5:20. Figure 9.2 shows the possible time intervals for the walkers on the path. Let an uncertain variable W represent the set of walkers that are present on the path at any time, and an uncertain variable T represent the time at which any walker on the path finishes its walk. Then, we have the marginal ranges*

$$[\![W]\!] = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}\}, \tag{9.18}$$

$$[\![T]\!] = [\textit{5:00}, \textit{5:30}]. \tag{9.19}$$

We also have the conditional ranges

$$[\![T|\{a\}]\!] = [\textit{5:00}, \textit{5:15}], \tag{9.20}$$

$$[\![T|\{b\}]\!] = [\textit{5:10}, \textit{5:30}], \tag{9.21}$$

$$[\![T|\{c\}]\!] = [\textit{5:20}, \textit{5:30}], \tag{9.22}$$

$$[\![T|\{a,b\}]\!] = [\textit{5:10}, \textit{5:15}], \tag{9.23}$$

$$[\![T|\{b,c\}]\!] = [\textit{5:20}, \textit{5:30}]. \tag{9.24}$$

For all $t \in [\textit{5:00}, \textit{5:10})$*, we have*

$$[\![W|t]\!] = \{\{a\}\}, \tag{9.25}$$

for all $t \in [\textit{5:10}, \textit{5:15}]$*, we have*

$$[\![W|t]\!] = \{\{a,b\}, \{a\}, \{b\}\}, \tag{9.26}$$

for all $t \in (\textit{5:15}, \textit{5:20})$*, we have*

$$[\![W|t]\!] = \{\{b\}\}, \tag{9.27}$$

and for all $t \in [\textit{5:20}, \textit{5:30}]$*, we have*

$$[\![W|t]\!] = \{\{b,c\}, \{b\}, \{c\}\}. \tag{9.28}$$

Now, let the uncertainty function of a time set $\mathscr{S}$ *be*

$$m_{\mathscr{T}}(\mathscr{S}) = \begin{cases} \mathcal{L}(\mathscr{S}) + 10 \textit{ if } \mathscr{S} \neq \emptyset, \\ 0 \textit{ otherwise }, \end{cases} \tag{9.29}$$

where $\mathcal{L}(\cdot)$ *is the Lebesgue measure. Let the uncertainty function* $m_{\mathscr{W}}(.)$ *associated to a set of*

individuals be the cardinality of the set. Then, the sets of association are

$$\mathscr{A}(W;T) = \{1/5, 3/5\}, \tag{9.30}$$

$$\mathscr{A}(T;W) = \{3/8, 1/2\}. \tag{9.31}$$

It follows that that for all $\delta_1 < 1/5$ *and* $\delta_2 < 3/8$*, we have*

$$(W,T) \overset{d}{\leftrightarrow} (\delta_1, \delta_2), \tag{9.32}$$

and the residual uncertainty in W *given* T *is at least* δ_1 *fraction of the total uncertainty in* W*, while the residual uncertainty in* T *given* W *is at least* δ_2 *fraction of the total uncertainty in* T*. On the other hand, for all* $\delta_1 \geq 3/5$ *and* $\delta_2 \geq 1/2$ *we have*

$$(W,T) \overset{a}{\leftrightarrow} (\delta_1, \delta_2), \tag{9.33}$$

and the residual uncertainty in W *given* T *is at most* δ_1 *fraction of the total uncertainty in* W*, while the residual uncertainty in* T *given* W *is at most* δ_2 *fraction of the total uncertainty in* T*.*

Finally, if $1/5 \leq \delta_1 < 3/5$ *or* $3/8 \leq \delta_2 < 1/2$*, then* W *and* T *are neither associated nor disassociated.*

9.4.3 δ-mutual information

We now introduce the mutual information between uncertain variables in terms of some structural properties of covering sets. Intuitively, for any $\delta \in [0,1]$ the δ-mutual information, expressed in bits, represents the most refined knowledge that one uncertain variable provides about the other, at a given level of confidence $(1-\delta)$. We express this idea by considering the quantization of the range of uncertainty of one variable, induced by the knowledge of the other. Such quantization ensures that the variable can be identified with uncertainty at most δ. The

notions of association and disassociation introduced above are used to ensure that the mutual information is well defined, namely it can be positive, and enjoys a certain symmetric property.

Definition 13. *δ-Connectedness and δ-isolation.*

- *For any $\delta \in [0,1]$, points $x_1, x_2 \in [\![X]\!]$ are δ-connected via $[\![X|Y]\!]$, and are denoted by $x_1 \overset{\delta}{\rightsquigarrow} x_2$, if there exists a finite sequence $\{[\![X|y_i]\!]\}_{i=1}^{N}$ of conditional sets such that $x_1 \in [\![X|y_1]\!]$, $x_2 \in [\![X|y_N]\!]$ and for all $1 < i \leq N$, we have*

$$\frac{m_{\mathscr{X}}([\![X|y_i]\!] \cap [\![X|y_{i-1}]\!])}{m_{\mathscr{X}}([\![X]\!])} > \delta. \tag{9.34}$$

  If $x_1 \overset{\delta}{\rightsquigarrow} x_2$ and $N = 1$, then we say that x_1 and x_2 are singly δ-connected via $[\![X|Y]\!]$, i.e. there exists a y such that $x_1, x_2 \in [\![X|y]\!]$.

- *A set $\mathscr{S} \subseteq [\![X]\!]$ is (singly) δ-connected via $[\![X|Y]\!]$ if every pair of points in the set is (singly) δ-connected via $[\![X|Y]\!]$.*

- *Two sets $\mathscr{S}_1, \mathscr{S}_2 \subseteq [\![X]\!]$ are δ-isolated via $[\![X|Y]\!]$ if no point in $\mathscr{S}_1$ is δ-connected to any point in $\mathscr{S}_2$.*

Definition 14. *δ-overlap family.*
For any $\delta \in [0,1]$, a $[\![X|Y]\!]$ δ-overlap family of $[\![X]\!]$, denoted by $[\![X|Y]\!]^*_\delta$, is a largest family of distinct sets covering $[\![X]\!]$ such that:

1. *Each set in the family is δ-connected and contains at least one singly δ-connected set of the form $[\![X|y]\!]$.*

2. *The measure of overlap between any two distinct sets in the family is at most $\delta m_{\mathscr{X}}([\![X]\!])$.*

3. *For every singly δ-connected set, there exist a set in the family containing it.*

The first property of the δ-overlap family ensures that points in the same set of the family *cannot* be distinguished with confidence at least $(1-\delta)$, while also ensuring that each set cannot

be arbitrarily small. The second and third properties ensure that points that are not covered by the same set of the family *can* be distinguished with confidence at least $(1-\delta)$. It follows that the cardinality of the covering family represents the most refined knowledge at a given level of confidence $(1-\delta)$ that we can have about X, given the knowledge of Y. This also corresponds to the most refined quantization of the set $[\![X]\!]$ induced by Y. This interpretation is analogous to the one in [158], extending the concept of overlap partition introduced there to a δ-overlap family in this work. The stage is now set to introduce the δ-mutual information in terms of the δ-overlap family.

Definition 15. *The δ-mutual information provided by Y about X is*

$$I_\delta(X;Y) = \log_2 |[\![X|Y]\!]^*_\delta| \text{ bits}, \tag{9.35}$$

if a $[\![X|Y]\!]$ δ-overlap family of $[\![X]\!]$ exists, otherwise it is zero.

We now show that when variables are associated at level (δ, δ_2), then there exists a δ-overlap family, so that the mutual information is well defined.

Theorem 37. *If $(X,Y) \overset{a}{\leftrightarrow} (\delta, \delta_2)$, then there exists a δ-overlap family $[\![X|Y]\!]^*_\delta$.*

Proof. We show that

$$[\![X|Y]\!] = \{[\![X|y]\!] : y \in [\![Y]\!]\} \tag{9.36}$$

is a δ-overlap family. First, note that $[\![X|Y]\!]$ is a cover of $[\![X]\!]$, since $[\![X]\!] = \cup_{y\in[\![Y]\!]}[\![X|y]\!]$. Second, each set in the family $[\![X|Y]\!]$ is singly δ-connected via $[\![X|Y]\!]$, since trivially any two points $x_1, x_2 \in [\![X|y]\!]$ are singly δ-connected via the same set. It follows that Property 1 of Definition 14 holds.

Now, since $(X,Y) \overset{a}{\leftrightarrow} (\delta, \delta_2)$, then by Lemma 31 for all $y_1, y_2 \in [\![Y]\!]$ we have

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!])}{m_{\mathscr{X}}([\![X]\!])} \leq \delta, \tag{9.37}$$

which shows that Property 2 of Definition 14 holds. Finally, it is also easy to see that Property 3 of Definition 14 holds, since $[\![X|Y]\!]$ contains all sets $[\![X|y]\!]$. □

Next, we show that a δ-overlap family also exists when variables are disassociated at level (δ, δ_2). In this case, we also characterize the mutual information in terms of a partition of $[\![X]\!]$.

Definition 16. *δ-isolated partition.*
A $[\![X|Y]\!]$ δ-isolated partition of $[\![X]\!]$, denoted by $[\![X|Y]\!]_\delta$, is a partition of $[\![X]\!]$ such that any two sets in the partition are δ-isolated via $[\![X|Y]\!]$.

Theorem 38. *If $(X, Y) \overset{d}{\leftrightarrow} (\delta, \delta_2)$, then we have:*

1. *There exists a unique δ-overlap family $[\![X|Y]\!]^*_\delta$.*
2. *The δ-overlap family is the δ-isolated partition of largest cardinality, namely for any $[\![X|Y]\!]_\delta$, we have*

$$|[\![X|Y]\!]_\delta| \leq |[\![X|Y]\!]^*_\delta|, \tag{9.38}$$

where the equality holds if and only if $[\![X|Y]\!]_\delta = [\![X|Y]\!]^*_\delta$.

Proof. First, we show the existence of a δ-overlap family. For all $x \in [\![X]\!]$, let $\mathscr{C}(x)$ be the set of points that are δ-connected to x via $[\![X|Y]\!]$, namely

$$\mathscr{C}(x) = \{x_1 \in [\![X]\!] : x \overset{\delta}{\leftrightsquigarrow} x_1\}. \tag{9.39}$$

Then, we let

$$\mathcal{C} = \{\mathscr{C}(x) : x \in [\![X]\!]\}, \tag{9.40}$$

and show that this is a δ-overlap family. First, note that since $[\![X]\!] = \cup_{\mathscr{S} \in \mathcal{C}} \mathscr{S}$, we have that $\mathcal{C}$ is a cover of $[\![X]\!]$. Second, for all $\mathscr{C}(x) \in \mathcal{C}$ there exists a $y \in [\![Y]\!]$ such that $x \in [\![X|y]\!]$, and since any two points $x_1, x_2 \in [\![X|y]\!]$ are singly δ-connected via $[\![X|Y]\!]$, we have that $[\![X|y]\!] \subseteq \mathscr{C}(x)$.

It follows that every set in the family $\mathcal{C}$ contains at least one singly δ-connected set. For all $x_1, x_2 \in \mathscr{C}(x)$, we also have $x_1 \overset{\delta}{\leftrightsquigarrow} x$ and $x \overset{\delta}{\leftrightsquigarrow} x_2$. Since $(X, Y) \overset{d}{\leftrightarrow} (\delta, \delta_2)$, by Lemma 34 in Appendix 9.12.4 this implies $x_1 \overset{\delta}{\leftrightsquigarrow} x_2$. It follows that every set in the family $\mathcal{C}$ is δ-connected and contains at least one singly δ-connected set, and we conclude that Property 1 of Definition 14 is satisfied.

We now claim that for all $x_1, x_2 \in [\![X]\!]$, if

$$\mathscr{C}(x_1) \neq \mathscr{C}(x_2), \tag{9.41}$$

then

$$m_{\mathscr{X}}(\mathscr{C}(x_1) \cap \mathscr{C}(x_2)) = 0. \tag{9.42}$$

This can be proven by contradiction. Let $\mathscr{C}(x_1) \neq \mathscr{C}(x_2)$ and assume that $m_{\mathscr{X}}(\mathscr{C}(x_1) \cap \mathscr{C}(x_2)) \neq 0$. By (9.8) this implies that $\mathscr{C}(x_1) \cap \mathscr{C}(x_2) \neq \emptyset$. We can then pick $z \in \mathscr{C}(x_1) \cap \mathscr{C}(x_2)$, such that we have $z \overset{\delta}{\leftrightsquigarrow} x_1$ and $z \overset{\delta}{\leftrightsquigarrow} x_2$. Since $(X, Y) \overset{d}{\leftrightarrow} (\delta, \delta_2)$, by Lemma 34 in Appendix 9.12.4 this also implies $x_1 \overset{\delta}{\leftrightsquigarrow} x_2$, and therefore $\mathscr{C}(x_1) = \mathscr{C}(x_2)$, which is a contradiction. It follows that if $\mathscr{C}(x_1) \neq \mathscr{C}(x_2)$, then we must have $m_{\mathscr{X}}(\mathscr{C}(x_1) \cap \mathscr{C}(x_2)) = 0$, and therefore

$$\frac{m_{\mathscr{X}}(\mathscr{C}(x_1) \cap \mathscr{C}(x_2))}{m_{\mathscr{X}}([\![X]\!])} = 0 \leq \delta. \tag{9.43}$$

We conclude that Property 2 of Definition 14 is satisfied.

Finally, we have that for any singly δ-connected set $[\![X|y]\!]$, there exist an $x \in [\![X]\!]$ such that $x \in [\![X|y]\!]$, which by (9.39) implies $[\![X|y]\!] \subseteq \mathscr{C}(x)$. Namely, for every singly δ-connected set, there exist a set in the family containing it. We can then conclude that $\mathcal{C}$ satisfies all the properties of a δ-overlap family.

Next, we show that $\mathcal{C}$ is a unique δ-overlap family. By contradiction, consider another δ-overlap family $\mathcal{D}$. For all $x \in [\![X]\!]$, let $\mathscr{D}(x)$ denote a set in $\mathcal{D}$ containing x. Then, using the

definition of $\mathscr{C}(x)$ and the fact that $\mathscr{D}(x)$ is δ-connected, it follows that

$$\mathscr{D}(x) \subseteq \mathscr{C}(x). \tag{9.44}$$

Next, we show that for all $x \in [\![X]\!]$, we also have

$$\mathscr{C}(x) \subseteq \mathscr{D}(x), \tag{9.45}$$

from which we conclude that $\mathcal{D} = \mathcal{C}$.

The proof of (9.45) is also obtained by contradiction. Assume there exists a point $\tilde{x} \in \mathscr{C}(x) \setminus \mathscr{D}(x)$. Since both x and $\tilde{x}$ are contained in $\mathscr{C}(x)$, we have $\tilde{x} \overset{\delta}{\leftrightsquigarrow} x$. Let x^* be a point in a singly-connected set that is contained in $\mathscr{D}(x)$, namely $x^* \in [\![X|y^*]\!] \subseteq \mathscr{D}(x)$. Since both x and x^* are in $\mathscr{D}(x)$, we have that $x \overset{\delta}{\leftrightsquigarrow} x^*$. Since $(X, Y) \overset{d}{\leftrightarrow} (\delta, \delta_2)$, we can apply Lemma 34 in Appendix 9.12.4 to conclude that $\tilde{x} \overset{\delta}{\leftrightsquigarrow} x^*$. It follows that there exists a sequence of conditional ranges $\{[\![X|y_i]\!]\}_{i=1}^N$ such that $\tilde{x} \in [\![X|y_1]\!]$ and $x^* \in [\![X|y_N]\!]$, which satisfies (9.34). Since x^* is in both $[\![X|y_N]\!]$ and $[\![X|y^*]\!]$, we have $[\![X|y_N]\!] \cap [\![X|y^*]\!] \neq \emptyset$ and since $(X, Y) \overset{d}{\leftrightarrow} (\delta, \delta_2)$, we have

$$\frac{m_{\mathscr{X}}([\![X|y_N]\!] \cap [\![X|y^*]\!])}{m_{\mathscr{X}}([\![X]\!])} > \delta. \tag{9.46}$$

Without loss of generality, we can then assume that the last element of our sequence is $[\![X|y^*]\!]$. By Property 3 of Definition 14, every conditional range in the sequence must be contained in some set of the δ-overlap family $\mathcal{D}$. Since $[\![X|y^*]\!] \subseteq \mathscr{D}(x)$ and $[\![X|y_1]\!] \not\subseteq \mathscr{D}(x)$, it follows that there exist two consecutive conditional ranges along the sequence and two sets of the δ-overlap family covering them, such that $[\![X|y_{i-1}]\!] \subseteq \mathscr{D}(x_{i-1})$, $[\![X|y_i]\!] \subseteq \mathscr{D}(x_i)$, and $\mathscr{D}(x_{i-1}) \neq \mathscr{D}(x_i)$.

Then, we have

$$\begin{aligned} & m_{\mathscr{X}}(\mathscr{D}(x_{i-1}) \cap \mathscr{D}(x_i)) \\ &= m_{\mathscr{X}}((\llbracket X|y_{i-1} \rrbracket \cap \llbracket X|y_i \rrbracket) \cup (\mathscr{D}(x_{i-1}^*) \cap \mathscr{D}(x_i^*))) \\ &\overset{(a)}{\geq} m_{\mathscr{X}}(\llbracket X|y_{i-1} \rrbracket \cap \llbracket X|y_i \rrbracket) \\ &\overset{(b)}{>} \delta m_{\mathscr{X}}(\llbracket X \rrbracket), \end{aligned} \tag{9.47}$$

where (a) follows from (9.9) and (b) follows from (9.34). It follows that

$$\frac{m_{\mathscr{X}}(\mathscr{D}(x_{i-1}) \cap \mathscr{D}(x_i))}{m_{\mathscr{X}}(\llbracket X \rrbracket)} > \delta, \tag{9.48}$$

and Property 2 of Definition 14 is violated. Thus, $\tilde{x}$ does not exists, which implies $\mathscr{C}(x) \subseteq \mathscr{D}(x)$. Combining (9.44) and (9.45), we conclude that the δ-overlap family $\mathcal{C}$ is unique.

We now turn to the proof of the second part of the Theorem. Since by (9.43) the uncertainty associated to the overlap between any two sets of the δ-overlap family $\mathcal{C}$ is zero, it follows that $\mathcal{C}$ is also a partition.

Now, we show that $\mathcal{C}$ is also a δ-isolated partition. This can be proven by contradiction. Assume $\mathcal{C}$ is not a δ-isolated partition. Then, there exists two distinct sets $\mathscr{C}(x_1), \mathscr{C}(x_2) \in \mathcal{C}$ such that $\mathscr{C}(x_1)$ and $\mathscr{C}(x_2)$ are not δ-isolated. This implies that there exists a point $\bar{x}_1 \in \mathscr{C}(x_1)$ and $\bar{x}_2 \in \mathscr{C}(x_2)$ such that $\bar{x}_1 \overset{\delta}{\nsim} \bar{x}_2$. Using the fact that $\mathscr{C}(x_1)$ and $\mathscr{C}(x_2)$ are δ-connected and Lemma 34 in Appendix 9.12.4, this implies that all points in the set $\mathscr{C}(x_1)$ are δ-connected to all points in the set $\mathscr{C}(x_2)$. Now, let x_1^* and x_2^* be points in a singly δ-connected set contained in $\mathscr{C}(x_1)$ and $\mathscr{C}(x_2)$ respectively, namely $x_1^* \in \llbracket X|y_1^* \rrbracket \subseteq \mathscr{C}(x_1)$ and $x_2^* \in \llbracket X|y_2^* \rrbracket \subseteq \mathscr{C}(x_2)$. Since $x_1^* \overset{\delta}{\nsim} x_2^*$, there exists a sequence of conditional ranges $\{\llbracket X|y_i \rrbracket\}_{i=1}^N$ satisfying (9.34), such that $x_1 \in \llbracket X|y_1 \rrbracket$ and $x_2 \in \llbracket X|y_N \rrbracket$. Without loss of generality, we can assume $\llbracket X|y_1 \rrbracket = \llbracket X|y_1^* \rrbracket$ and $\llbracket X|y_2 \rrbracket = \llbracket X|y_2^* \rrbracket$. Since $\mathcal{C}$ is a partition, we have that $\llbracket X|y_1^* \rrbracket \subseteq \mathscr{C}(x_1)$ and $\llbracket X|y_2^* \rrbracket \not\subseteq \mathscr{C}(x_1)$. It follows that there exist two consecutive conditional ranges along the sequence $\{\llbracket X|y_i \rrbracket\}_{i=1}^N$

and two sets of the δ-overlap family $\mathcal{C}$ covering them, such that $[\![X|y_{i-1}]\!] \subseteq \mathscr{C}(x_{i-1})$ and $[\![X|y_i]\!] \subseteq \mathscr{C}(x_i)$, and $\mathscr{C}(x_{i-1}) \neq \mathscr{C}(x_i)$. Similar to (9.47), we have

$$\frac{m_{\mathscr{X}}(\mathscr{C}(x_{i-1}) \cap \mathscr{C}(x_i))}{m_{\mathscr{X}}([\![X]\!])} > \delta, \tag{9.49}$$

and Property 2 of Definition 14 is violated. Thus, $\mathscr{C}(x_1)$ and $\mathscr{C}(x_2)$ do not exist, which implies $\mathcal{C}$ is δ-isolated partition.

Let $\mathcal{P}$ be any other δ-isolated partition. We wish to show that $|\mathcal{C}| \geq |\mathcal{P}|$, and that the equality holds if and only if $\mathcal{P} = \mathcal{C}$. First, note that every set $\mathscr{C}(x) \in \mathcal{C}$ can intersect at most one set in $\mathcal{P}$, otherwise the sets in $\mathcal{P}$ would not be δ-isolated. Second, since $\mathcal{C}$ is a cover of $[\![X]\!]$, every set in $\mathcal{P}$ must be intersected by at least one set in $\mathcal{C}$. It follows that

$$|\mathcal{C}| \geq |\mathcal{P}|. \tag{9.50}$$

Now, assume the equality holds. In this case, there is a one-to-one correspondence $\mathscr{P} : \mathcal{C} \to \mathcal{P}$, such that for all $x \in [\![X]\!]$, we have $\mathscr{C}(x) \subseteq \mathscr{P}(\mathscr{C}(x))$, and since both $\mathcal{C}$ and $\mathcal{P}$ are partitions of $[\![X]\!]$, it follows that $\mathcal{C} = \mathcal{P}$. Conversely, assuming $\mathcal{C} = \mathcal{P}$, then $|\mathcal{C}| = |\mathcal{P}|$ follows trivially. $\square$

We have introduced the notion of mutual information from Y to X in terms of the conditional range $[\![X|Y]\!]$. Since in general we have $[\![X|Y]\!] \neq [\![Y|X]\!]$, one may expect the definition of mutual information to be asymmetric in its arguments. Namely, the amount of information provided about X by the knowledge of Y may not be the same as the amount of information provided about Y by the knowledge of X. Although this is true in general, we show that for disassociated UVs symmetry is retained, provided that when swapping X with Y one also rescales δ appropriately. The following theorem establishes the symmetry in the mutual information under the appropriate scaling of the parameters δ_1 and δ_2. The proof requires the introduction of the notions of taxicab connectedness, taxicab family, and taxicab partition, which are given in Appendix 9.12.6.

Theorem 39. *If* $(X, Y) \overset{d}{\leftrightarrow} (\delta_1, \delta_2)$*, and a* (δ_1, δ_2)*-taxicab family of* $[\![X, Y]\!]$ *exists, then we have*

$$I_{\delta_1}(X;Y) = I_{\delta_2}(Y;X). \tag{9.51}$$

9.5 (ϵ,δ)-Capacity

We now give an operational definition of capacity of a communication channel and relate it to the notion of mutual information between UVs introduced above. Let $\mathscr{X}$ be a totally bounded, normed metric space such that for all $x \in \mathscr{X}$ we have $\|x\| \leq 1$, where $\|.\|$ represents norm. This normalization is for convenience of notation and all results can easily be extended to metric spaces of any bounded norm. Let $\mathcal{X} \subseteq \mathscr{X}$ be a discrete set of points in the space, which represents a codebook. Any point $x \in \mathcal{X}$ represents a codeword that can be selected at the transmitter, sent over the channel, and received with noise perturbation at most ϵ. Namely, for any transmitted codeword $x \in \mathcal{X}$, we receive a point in the set

$$S_\epsilon(x) = \{y \in \mathscr{X} : \|x - y\| \leq \epsilon\}. \tag{9.52}$$

It follows that all received codewords lie in the set $\mathcal{Y} = \bigcup_{x \in \mathcal{X}} S_\epsilon(x)$, where $\mathcal{Y} \subseteq \mathscr{Y} = \mathscr{X}$. Transmitted codewords can be decoded correctly as long as the corresponding uncertainty sets at the receiver do not overlap. This can be done by simply associating the received codeword to the point in the codebook that is closest to it.

For any $x_1, x_2 \in \mathcal{X}$, we now let

$$e_\epsilon(x_1, x_2) = \frac{m_{\mathscr{Y}}(S_\epsilon(x_1) \cap S_\epsilon(x_2))}{m_{\mathscr{Y}}(\mathscr{Y})}, \tag{9.53}$$

where $m_{\mathscr{Y}}(.)$ is an uncertainity function defined over the space $\mathscr{Y}$. We also assume without loss of generality that the uncertainty associated to the whole space $\mathscr{Y}$ of received codewords is $m_{\mathscr{Y}}(\mathscr{Y}) = 1$. Finally, we let $V_\epsilon \subseteq \mathscr{Y}$ be the smallest uncertainty set corresponding to a

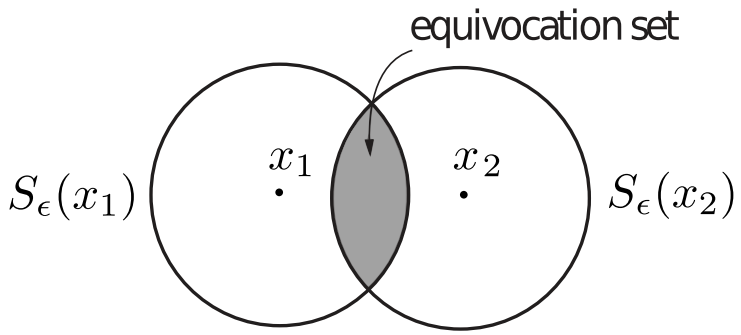


Figure 9.3. The size of the equivocation set is inversely proportional to the amount of adversarial effort required to induce an error.

transmitted codeword, namely $V_\epsilon = S_\epsilon(x^*)$, where $x^* = \text{argmin}_{x \in \mathscr{X}} m_{\mathscr{Y}}(S_\epsilon(x))$. The quantity $1 - e_\epsilon(x_1, x_2)$ can be viewed as the confidence we have of not confusing x_1 and x_2 in any transmission, or equivalently as the amount of adversarial effort required to induce a confusion between the two codewords. For example, if the uncertainty function is constructed using a measure, then all the erroneous codewords generated by an adversary to decode x_2 instead than x_1 must lie inside the equivocation set depicted in Figure 9.3, whose relative size is given by (9.53). The smaller the equivocation set is, the larger must be the effort required by the adversary to induce an error. If the uncertainty function represents the diameter of the set, then all the erroneous codewords generated by an adversary to decode x_2 instead than x_1 will be close to each other, in the sense of (9.53). Once again, the closer the possible erroneous codewords are, the harder must be for the adversary to generate an error, since any small deviation allows the decoder to correctly identify the transmitted codeword.

We now introduce the notion of *distinguishable codebook*, ensuring that every codeword cannot be confused with any other codeword, rather than with a specific one, at a given level of confidence.

Definition 17. (ϵ, δ)*-distinguishable codebook.*
For any $0 < \epsilon \leq 1$, $0 \leq \delta < m_{\mathscr{Y}}(V_\epsilon)$*, a codebook* $\mathcal{X} \subseteq \mathscr{X}$ *is* (ϵ, δ)*-distinguishable if for all* $x_1, x_2 \in \mathcal{X}$*, we have* $e_\epsilon(x_1, x_2) \leq \delta / |\mathcal{X}|$.

For any (ϵ, δ)-distinguishable codebook $\mathcal{X}$ and $x \in \mathcal{X}$, we let

$$e_\epsilon(x) = \sum_{x' \in \mathcal{X} : x' \neq x} e_\epsilon(x, x'). \tag{9.54}$$

It now follows from Definition 17 that

$$e_\epsilon(x) \leq \delta, \tag{9.55}$$

and each codeword in an (ϵ, δ)-distinguishable codebook can be decoded correctly with confidence at least $1 - \delta$. Definition 17 guarantees even more, namely that the confidence of not confusing any pair of codewords is uniformly bounded by $1 - \delta/|\mathcal{X}|$. This stronger constraint implies that we cannot "balance" the error associated to a codeword transmission by allowing some decoding pair to have a lower confidence and enforcing other pairs to have higher confidence. This is the main difference between our definition and the one used in [132] which bounds the average confidence, and allows us to relate the notion of capacity to the mutual information between pairs of codewords.

Definition 18. (ϵ, δ)-*capacity.*
For any totally bounded, normed metric space $\mathscr{X}$*,* $0 < \epsilon \leq 1$*, and* $0 \leq \delta < m_{\mathscr{Y}}(V_\epsilon)$*, the* (ϵ, δ)-*capacity of* $\mathscr{X}$ *is*

$$C_\epsilon^\delta = \sup_{\mathcal{X} \in \mathscr{X}_\epsilon^\delta} \log_2 |\mathcal{X}| \text{ bits}, \tag{9.56}$$

where $\mathscr{X}_\epsilon^\delta = \{\mathcal{X} : \mathcal{X}$ *is* (ϵ, δ)-*distinguishable*$\}$ *is the set of* (ϵ, δ)-*distinguishable codebooks.*

The (ϵ, δ)-capacity represents the largest number of bits that can be communicated by using any (ϵ, δ)-distinguishable codebook. The corresponding geometric picture is illustrated in Figure 9.4. For $\delta = 0$, our notion of capacity reduces to Kolmogorov's ϵ-capacity, that is the logarithm of the packing number of the space with balls of radius ϵ.

In the definition of capacity, we have restricted $\delta < m_{\mathscr{Y}}(V_\epsilon)$ to rule out the case when the

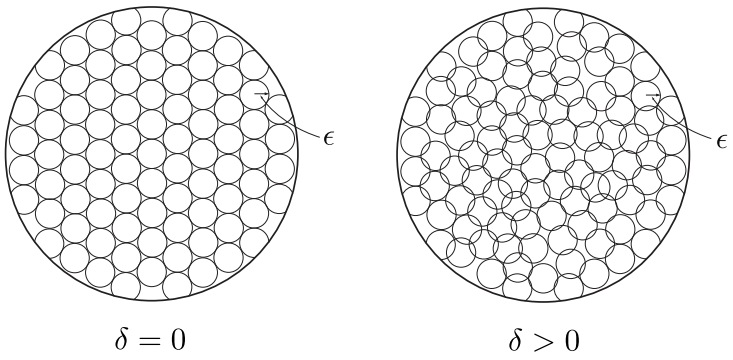


Figure 9.4. Illustration of the (ϵ, δ)-capacity in terms of packing ϵ-balls with maximum overlap δ.

decoding error can be at least as large as the error introduced by the channel, and the (ϵ, δ)-capacity is infinite. Also, note that $m_{\mathscr{Y}}(V_\epsilon) \leq 1$ since $V_\epsilon \subseteq \mathscr{Y}$ and (9.9) holds.

We now relate our operational definition of capacity to the notion of UVs and mutual information introduced in Section 9.4. Let X be the UV corresponding to the transmitted codeword. This is a map $X : \mathscr{X} \to \mathcal{X}$ and $[\![X]\!] = \mathcal{X} \subseteq \mathscr{X}$. Likewise, let Y be the UV corresponding to the received codeword. This is a map $Y : \mathscr{Y} \to \mathcal{Y}$ and $[\![Y]\!] = \mathcal{Y} \subseteq \mathscr{Y}$. For our ϵ-perturbation channel, these UVs are such that for all $y \in [\![Y]\!]$ and $x \in [\![X]\!]$, we have

$$[\![Y|x]\!] = \{y \in [\![Y]\!] : \|x - y\| \leq \epsilon\}, \tag{9.57}$$

$$[\![X|y]\!] = \{x \in [\![X]\!] : \|x - y\| \leq \epsilon\}, \tag{9.58}$$

see Figure 9.5. Clearly, the set in (9.57) is continuous, while the set in (9.58) is discrete.

To measure the levels of association and disassociation between X and Y, we use an uncertainty function $m_{\mathscr{X}}(.)$ defined over $\mathscr{X}$, and $m_{\mathscr{Y}}(.)$ defined over $\mathscr{Y}$. We introduce the feasible set

$$\begin{aligned}\mathscr{F}_\delta = \{X : [\![X]\!] \subseteq \mathscr{X}, \text{ and either } (X, Y) \overset{d}{\leftrightarrow} (0, \delta/|[\![X]\!]|) \\ \text{or } (X, Y) \overset{a}{\leftrightarrow} (1, \delta/|[\![X]\!]|)\},\end{aligned} \tag{9.59}$$

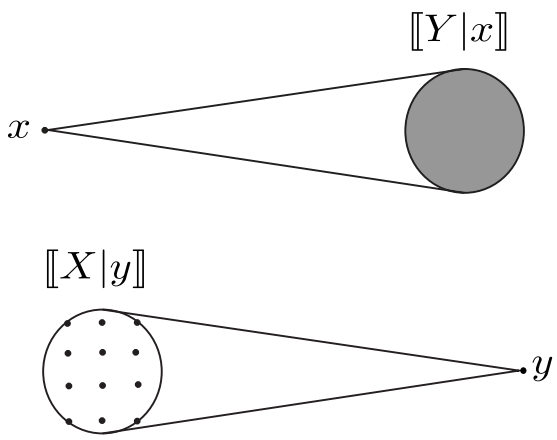


Figure 9.5. Conditional ranges $[\![Y|x]\!]$ and $[\![X|y]\!]$ due to the ϵ-perturbation channel.

representing the set of UVs X such that the marginal range $[\![X]\!]$ is a discrete set representing a codebook, and the UV can either achieve $(0, \delta/|[\![X]\!]|)$ levels of disassociation or $(1, \delta/|[\![X]\!]|)$ levels of association with Y. In our channel model, this feasible set also depends on the ϵ-perturbation through (9.57) and (9.58).

We can now state the non-stochastic channel coding theorem for our ϵ-perturbation channel.

Theorem 40. *For any totally bounded, normed metric space $\mathscr{X}$, ϵ-perturbation channel satisfying* (9.57) *and* (9.58)*, $0 < \epsilon \leq 1$ and $0 \leq \delta < m_{\mathscr{Y}}(V_\epsilon)$, we have*

$$C_\epsilon^\delta = \sup_{X \in \mathscr{F}_{\tilde{\delta}}, \tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])} I_{\tilde{\delta}/|[\![X]\!]|}(Y; X) \text{ bits}. \tag{9.60}$$

Proof. First, we show that there exists a UV X and $\tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])$ such that $X \in \mathscr{F}_{\tilde{\delta}}$, which implies that the supremum is well defined. Second, for all X and $\tilde{\delta}$ such that

$$X \in \mathscr{F}_{\tilde{\delta}}, \tag{9.61}$$

and

$$\tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!]), \tag{9.62}$$

we show that

$$I_{\tilde{\delta}/|[\![X]\!]|}(Y;X) \leq C_{\epsilon}^{\delta}. \tag{9.63}$$

Finally, we show the existence of $X \in \mathscr{F}_{\tilde{\delta}}$ and $\tilde{\delta} \leq \delta/m_{\mathscr{Y}}([\![Y]\!])$ such that $I_{\tilde{\delta}/|[\![X]\!]|}(Y;X) = C_{\epsilon}^{\delta}$.

Let us begin with the first step. Consider a point $x \in \mathscr{X}$. Let X be a UV such that

$$[\![X]\!] = \{x\}. \tag{9.64}$$

Then, we have that the marginal range of the UV Y corresponding to the received variable is

$$[\![Y]\!] = [\![Y|x]\!], \tag{9.65}$$

and therefore for all $y \in [\![Y]\!]$, we have

$$[\![X|y]\!] = \{x\}. \tag{9.66}$$

Using Definition 10 and (9.64), we have that

$$\mathscr{A}(Y;X) = \emptyset, \tag{9.67}$$

because $[\![X]\!]$ consists of a single point, and therefore the set in (9.11) is empty.

On the other hand, using Definition 10 and (9.66), we have

$$\mathscr{A}(X;Y) = \begin{cases} \{1\} \text{ if } \exists y_1, y_2 \in [\![Y]\!], \\ \emptyset \quad \text{otherwise.} \end{cases} \tag{9.68}$$

Using (9.67) and since $\mathscr{A} \preceq \delta$ holds for $\mathscr{A} = \emptyset$, we have

$$\mathscr{A}(Y;X) \preceq \delta/(|[\![X]\!]|m_{\mathscr{Y}}([\![Y]\!])). \tag{9.69}$$

Similarly, using (9.68), we have

$$\mathscr{A}(X;Y) \preceq 1. \tag{9.70}$$

Now, combining (9.69) and (9.70), we have

$$(X,Y) \overset{a}{\leftrightarrow} (1, \delta/(|[\![X]\!]|m_{\mathscr{Y}}([\![Y]\!]))). \tag{9.71}$$

Letting $\tilde{\delta} = \delta/m_{\mathscr{Y}}([\![Y]\!])$, this implies that $X \in \mathscr{F}_{\tilde{\delta}}$ and the first step of the proof is complete.

To prove the second step, we define the set of discrete UVs

$$\begin{aligned}\mathscr{G} = \{X : [\![X]\!] \subseteq \mathscr{X}, \exists \tilde{\delta} \leq \delta/m_{\mathscr{Y}}([\![Y]\!]) \text{ such that } \forall \mathscr{S}_1, \mathscr{S}_2 \in [\![Y|X]\!], \\ m_{\mathscr{Y}}(\mathscr{S}_1 \cap \mathscr{S}_2)/m_{\mathscr{Y}}([\![Y]\!]) \leq \tilde{\delta}/|[\![X]\!]|\},\end{aligned} \tag{9.72}$$

which is a larger set than the one containing all UVs X that are $(1, \tilde{\delta}/|[\![X]\!]|)$ associated to Y. Now, we will show that if an UV $X \in \mathscr{G}$, then the corresponding codebook $\mathcal{X} \in \mathscr{X}_{\epsilon}^{\delta}$. If $X \in \mathscr{G}$, then there exists a $\tilde{\delta} \leq \delta/m_{\mathscr{Y}}([\![Y]\!])$ such that for all $\mathscr{S}_1, \mathscr{S}_2 \in [\![Y|X]\!]$ we have

$$\frac{m_{\mathscr{Y}}(\mathscr{S}_1 \cap \mathscr{S}_2)}{m_{\mathscr{Y}}([\![Y]\!])} \leq \frac{\tilde{\delta}}{|[\![X]\!]|}. \tag{9.73}$$

It follows that for all $x_1, x_2 \in [\![X]\!]$, we have

$$\frac{m_{\mathscr{Y}}([\![Y|x_1]\!] \cap [\![Y|x_2]\!])}{m_{\mathscr{Y}}([\![Y]\!])} \leq \frac{\tilde{\delta}}{|[\![X]\!]|}. \tag{9.74}$$

Using $\mathcal{X} = [\![X]\!]$, (9.57), $[\![Y]\!] = \mathcal{Y} = \bigcup_{x \in \mathcal{X}} S_{\epsilon}(x)$ and $m_{\mathscr{Y}}(\mathscr{Y}) = 1$, for all $x_1, x_2 \in \mathcal{X}$ we have

$$\begin{aligned}\frac{m_{\mathscr{Y}}(S_{\epsilon}(x_1) \cap S_{\epsilon}(x_2))}{m_{\mathscr{Y}}(\mathscr{Y})} &\leq \frac{\tilde{\delta} m_{\mathscr{Y}}([\![Y]\!])}{|\mathcal{X}|} \\ &\overset{(a)}{\leq} \frac{\delta}{|\mathcal{X}|},\end{aligned} \tag{9.75}$$

where (a) follows from $\tilde{\delta} \leq \delta/m_{\mathscr{Y}}([\![Y]\!])$. Putting things together, it follows that

$$X \in \mathscr{G} \implies \mathcal{X} \in \mathscr{X}_{\epsilon}^{\delta} \tag{9.76}$$

Consider now a pair X and $\tilde{\delta}$ such that $\tilde{\delta} \leq \delta/m_{\mathscr{Y}}([\![Y]\!])$ and $X \in \mathscr{F}_{\tilde{\delta}}$

If $(X, Y) \overset{d}{\leftrightarrow} (0, \tilde{\delta}/|[\![X]\!]|)$, then using Lemma 33 in Appendix 9.12.4 there exist two UVs $\bar{X}$ and $\bar{Y}$, and $\bar{\delta} \leq \delta/m_{\mathscr{Y}}([\![\bar{Y}]\!])$ such that

$$(\bar{X}, \bar{Y}) \overset{a}{\leftrightarrow} (1, \bar{\delta}/|[\![\bar{X}]\!]|), \tag{9.77}$$

and

$$|[\![Y|X]\!]^{*}_{\tilde{\delta}/|[\![X]\!]|}| = |[\![\bar{Y}|\bar{X}]\!]^{*}_{\bar{\delta}/|[\![\bar{X}]\!]|}|. \tag{9.78}$$

On the other hand, if $(X, Y) \overset{a}{\leftrightarrow} (1, \tilde{\delta}/|[\![X]\!]|)$, then (9.77) and (9.78) also trivially hold. It then follows that (9.77) and (9.78) hold for all $X \in \mathscr{F}_{\tilde{\delta}}$. We now have

$$\begin{aligned} I_{\tilde{\delta}/|[\![X]\!]|}(Y; X) &= \log(|[\![Y|X]\!]^{*}_{\tilde{\delta}/|[\![X]\!]|}|) \\ &\overset{(a)}{=} \log(|[\![\bar{Y}|\bar{X}]\!]^{*}_{\bar{\delta}/|[\![\bar{X}]\!]|}|) \\ &\overset{(b)}{\leq} \log(|[\![\bar{X}]\!]|) \\ &\overset{(c)}{=} \log(|\bar{\mathcal{X}}|) \\ &\overset{(d)}{\leq} C_{\epsilon}^{\delta}, \end{aligned} \tag{9.79}$$

where (a) follows from (9.77) and (9.78), (b) follows from Lemma 35 in Appendix 9.12.4 since $\bar{\delta} \leq \delta/m_{\mathscr{Y}}([\![\bar{Y}]\!]) < m_{\mathscr{Y}}(V_{\epsilon})/m_{\mathscr{Y}}([\![\bar{Y}]\!])$, (c) follows by defining the codebook $\bar{\mathcal{X}}$ corresponding to the UV $\bar{X}$, and (d) follows from the fact that using (9.77) and Lemma 31, we have $\bar{X} \in \mathcal{G}$, which implies by (9.76) that $\bar{\mathcal{X}} \in \mathscr{X}_{\epsilon}^{\delta}$.

Finally, let

$$\mathcal{X}^{*} = \text{argsup}_{\mathcal{X} \in \mathscr{X}_{\epsilon}^{\delta}} \log(|\mathcal{X}|), \tag{9.80}$$

which achieves the capacity C_ϵ^δ. Let X^* be the UV whose marginal range corresponds to the codebook $\mathcal{X}^*$. It follows that for all $\mathscr{S}_1, \mathscr{S}_1 \in [\![Y^*|X^*]\!]$, we have

$$\frac{m_{\mathscr{Y}}(\mathscr{S}_1 \cap \mathscr{S}_1)}{m_{\mathscr{Y}}(\mathscr{Y})} \leq \frac{\delta}{|[\![X^*]\!]|}, \tag{9.81}$$

which implies using the fact that $m_{\mathscr{Y}}(\mathscr{Y}) = 1$,

$$\frac{m_{\mathscr{Y}}(\mathscr{S}_1 \cap \mathscr{S}_1)}{m_{\mathscr{Y}}([\![Y^*]\!])} \leq \frac{\delta}{|[\![X^*]\!]| m_{\mathscr{Y}}([\![Y^*]\!])}. \tag{9.82}$$

Letting $\delta^* = \delta / m_{\mathscr{Y}}([\![Y^*]\!])$, and using Lemma 31, we have that $(X^*, Y^*) \overset{a}{\leftrightarrow} (1, \delta^*/|[\![X^*]\!]|)$, which implies $X^* \in \cup_{\tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y^*]\!])} \mathscr{F}_{\tilde{\delta}}$ and the proof is complete.

□

Theorem 40 characterizes the capacity as the supremum of the mutual information over all UVs in the feasible set. The following theorem shows that the same characterization is obtained if we optimize the right hand side in (9.60) over all UVs in the space. It follows that by Theorem 40, rather than optimizing over all UVs representing all the codebooks in the space, a capacity achieving codebook can be found within the smaller class $\cup_{\tilde{\delta} \leq \delta / m_{\mathscr{Y}}(V_\epsilon)} \mathscr{F}_{\tilde{\delta}}$ of feasible sets with error at most $\delta / m_{\mathscr{Y}}(V_\epsilon)$, since for all $[\![Y]\!] \subseteq \mathscr{Y}$, $m_{\mathscr{Y}}(V_\epsilon) \leq m_{\mathscr{Y}}([\![Y]\!])$.

Theorem 41. *The* (ϵ, δ)*-capacity in* (9.60) *can also be written as*

$$C_\epsilon^\delta = \sup_{\substack{X : [\![X]\!] \subseteq \mathscr{X}, \\ \tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])}} I_{\tilde{\delta}/|[\![X]\!]|}(Y; X) \textit{ bits}. \tag{9.83}$$

Proof. Consider an UV $X \notin \cup_{\tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])} \mathscr{F}_{\tilde{\delta}}$, where Y is the corresponding UV at the receiver. The idea of the proof is to show the existence of an UV $\bar{X} \in \cup_{\tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![\bar{Y}]\!])} \mathscr{F}_{\tilde{\delta}}$ and the corresponding UV $\bar{Y}$ at the receiver, and

$$\bar{\delta} = \tilde{\delta} m_{\mathscr{Y}}([\![Y]\!]) / m_{\mathscr{Y}}([\![\bar{Y}]\!]) \leq \delta / m([\![\bar{Y}]\!]), \tag{9.84}$$

such that the cardinality of the overlap partitions

$$|[\![\bar{Y}|\bar{X}]\!]^*_{\tilde{\delta}/|[\![\bar{X}]\!]|}| = |[\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}|. \tag{9.85}$$

Let the cardinality

$$|[\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}| = K. \tag{9.86}$$

By Property 1 of Definition 14, we have that for all $\mathscr{S}_i \in [\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}$, there exists an $x_i \in [\![X]\!]$ such that $[\![Y|x_i]\!] \subseteq \mathscr{S}_i$. Now, consider another UV $\bar{X}$ whose marginal range is composed of K elements of $[\![X]\!]$, namely

$$[\![\bar{X}]\!] = \{x_1, \dots x_K\}. \tag{9.87}$$

Let $\bar{Y}$ be the UV corresponding to the received variable. Using the fact that for all $x \in \mathscr{X}$, we have $[\![\bar{Y}|x]\!] = [\![Y|x]\!]$ since (9.57) holds, and using Property 2 of Definition 14, for all $x, x' \in [\![\bar{X}]\!]$, we have that

$$\begin{aligned} \frac{m_{\mathscr{Y}}([\![\bar{Y}|x]\!] \cap [\![\bar{Y}|x']\!])}{m_{\mathscr{Y}}([\![Y]\!])} &\leq \frac{\tilde{\delta}}{|[\![X]\!]|} \\ &\overset{(a)}{\leq} \frac{\tilde{\delta}}{|[\![\bar{X}]\!]|}, \end{aligned} \tag{9.88}$$

where (a) follows from the fact that $[\![\bar{X}]\!] \subseteq [\![X]\!]$ using (9.87). Then, for all $x, x' \in [\![\bar{X}]\!]$, we have that

$$\begin{aligned} \frac{m_{\mathscr{Y}}([\![\bar{Y}|x]\!] \cap [\![\bar{Y}|x']\!])}{m_{\mathscr{Y}}([\![\bar{Y}]\!])} &\leq \frac{\tilde{\delta} m_{\mathscr{Y}}([\![Y]\!])}{|[\![\bar{X}]\!]| m_{\mathscr{Y}}([\![\bar{Y}]\!])} \\ &= \frac{\bar{\delta}}{|[\![\bar{X}]\!]|}, \end{aligned} \tag{9.89}$$

since $\bar{\delta} = \tilde{\delta} m_{\mathscr{Y}}([\![Y]\!])/m_{\mathscr{Y}}([\![\bar{Y}]\!])$. Then, by Lemma 31 it follows that

$$(\bar{X}, \bar{Y}) \overset{a}{\leftrightarrow} (1, \bar{\delta}/|[\![\bar{X}]\!]|). \tag{9.90}$$

Since $\tilde{\delta} \leq \delta/m_{\mathscr{Y}}([\![Y]\!])$, we have

$$\bar{\delta} \leq \delta/m_{\mathscr{Y}}([\![\bar{Y}]\!]) < m_{\mathscr{Y}}(V_\epsilon)/m_{\mathscr{Y}}([\![\bar{Y}]\!]). \tag{9.91}$$

Therefore, $\bar{X} \in \mathscr{F}_{\bar{\delta}}$ and $\bar{\delta} \leq \delta/m_{\mathscr{Y}}([\![\bar{Y}]\!])$. We now have that

$$\begin{aligned} |[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}| &\overset{(a)}{=} |[\![\bar{X}]\!]| \\ &\overset{(b)}{=} |[\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}|, \end{aligned} \tag{9.92}$$

where (a) follows by applying Lemma 36 in Appendix 9.12.4 using (9.90) and (9.91), and (b) follows from (9.86) and (9.87). Combining (9.92) with Theorem 40, the proof is complete. □

Finally, we make some considerations with respect to previous results in the literature. First, we note that for $\delta = 0$, all of our definitions reduce to Nair's ones and Theorem 40 recovers Nair's coding theorem [158, Theorem 4.1] for the zero-error capacity of an additive ϵ-perturbation channel.

Second, we point out that the (ϵ, δ)-capacity considered in [132] defines the set of (ϵ, δ)-distinguishable codewords such that the *average* overlap among all codewords is at most δ. In contrast, our definition requires the overlap for *each* pair of codewords to be at most $\delta/|\mathcal{X}|$. The following theorem provides the relationship between our C_ϵ^δ and the capacity $\tilde{C}_\epsilon^\delta$ considered in [132], which is defined using the Euclidean norm.

Theorem 42. *Let $\tilde{C}_\epsilon^\delta$ be the (ϵ, δ)-capacity defined in [132]. We have*

$$C_\epsilon^\delta \leq \tilde{C}_\epsilon^{\delta/(2m_{\mathscr{Y}}(V_\epsilon))}, \tag{9.93}$$

and

$$\tilde{C}_\epsilon^\delta \leq C_\epsilon^{\delta m_{\mathscr{Y}}(V_\epsilon) 2^{2\tilde{C}_\epsilon^\delta + 1}}. \tag{9.94}$$

Proof. For every codebook $\mathcal{X} \in \mathscr{X}_\epsilon^\delta$ and $x_1, x_2 \in \mathcal{X}$, we have

$$e_\epsilon(x_1, x_2) \le \delta/|\mathcal{X}|. \tag{9.95}$$

Since $m_{\mathscr{Y}}(\mathscr{Y}) = 1$, this implies that for all $x_1, x_2 \in \mathcal{X}$, we have

$$m_{\mathscr{Y}}(S_\epsilon(x_1) \cap S_\epsilon(x_2)) \le \delta/|\mathcal{X}|. \tag{9.96}$$

For all $\mathcal{X} \in \mathscr{X}$, the average overlap defined in [132, (53)] is

$$\Delta = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \frac{e_\epsilon(x)}{2m_{\mathscr{Y}}(V_\epsilon)}. \tag{9.97}$$

Then, we have

$$\begin{aligned} \Delta &= \frac{1}{2|\mathcal{X}|m_{\mathscr{Y}}(V_\epsilon)} \sum_{x_1, x_2 \in \mathcal{X}} m_{\mathscr{Y}}(S_\epsilon(x_1) \cap S_\epsilon(x_2)) \\ &\overset{(a)}{\le} \frac{\delta|\mathcal{X}|^2}{2|\mathcal{X}|^2 m_{\mathscr{Y}}(V_\epsilon)} \\ &\le \frac{\delta}{2m_{\mathscr{Y}}(V_\epsilon)}, \end{aligned} \tag{9.98}$$

where (a) follows from (9.96). Thus, we have

$$C_\epsilon^\delta \le \tilde{C}_\epsilon^{\delta/(2m_{\mathscr{Y}}(V_\epsilon))}, \tag{9.99}$$

and (9.93) follows.

Now, let $\mathcal{X}$ be a codebook with average overlap at most δ, namely

$$\frac{1}{2|\mathcal{X}|m_{\mathscr{Y}}(V_\epsilon)} \sum_{x_1, x_2 \in \mathcal{X}} m_{\mathscr{Y}}(S_\epsilon(x_1) \cap S_\epsilon(x_2)) \le \delta. \tag{9.100}$$

This implies that for all $x_1, x_2 \in \mathcal{X}$, we have

$$\begin{aligned}\frac{|\mathcal{X}| m_{\mathscr{Y}}(S_\epsilon(x_1) \cap S_\epsilon(x_2))}{m_{\mathscr{Y}}(\mathscr{Y})} &\leq \frac{2\delta |\mathcal{X}|^2 m_{\mathscr{Y}}(V_\epsilon)}{m_{\mathscr{Y}}(\mathscr{Y})} \\ &\overset{(a)}{=} 2\delta |\mathcal{X}|^2 m_{\mathscr{Y}}(V_\epsilon) \\ &\leq \delta 2^{2\tilde{C}_\epsilon^\delta + 1} m_{\mathscr{Y}}(V_\epsilon), \end{aligned} \tag{9.101}$$

where (a) follows from the fact that $m_{\mathscr{Y}}(\mathscr{Y}) = 1$. Thus, we have

$$\tilde{C}_\epsilon^\delta \leq C_\epsilon^{\delta 2^{2\tilde{C}_\epsilon^\delta + 1} m_{\mathscr{Y}}(V_\epsilon)}, \tag{9.102}$$

and (9.94) follows. □

9.6 (N, δ)-Capacity of General Channels

We now extend our results to more general channels where the noise can be different across codewords, and not necessarily contained within a ball of radius ϵ.

Let $\mathcal{X} \subseteq \mathscr{X}$ be a discrete set of points in the space, which represents a codebook. Any point $x \in \mathcal{X}$ represents a codeword that can be selected at the transmitter, sent over the channel, and received with perturbation. A channel with transition mapping $N : \mathscr{X} \to \mathscr{Y}$ associates to any point in $\mathscr{X}$ a set in $\mathscr{Y}$, such that the received codeword lies in the set

$$S_N(x) = \{y \in \mathscr{Y} : y \in N(x)\}. \tag{9.103}$$

Figure 9.6 illustrates possible uncertainty sets associated to three different codewords.

All received codewords lie in the set $\mathcal{Y} = \bigcup_{x \in \mathcal{X}} S_N(x)$, where $\mathcal{Y} \subseteq \mathscr{Y}$. For any $x_1, x_2 \in \mathcal{X}$, we now let

$$e_N(x_1, x_2) = \frac{m_{\mathscr{Y}}(S_N(x_1) \cap S_N(x_2))}{m_{\mathscr{Y}}(\mathscr{Y})}, \tag{9.104}$$

where $m_{\mathscr{Y}}(.)$ is an uncertainty function defined over $\mathscr{Y}$. We also assume without loss of generality

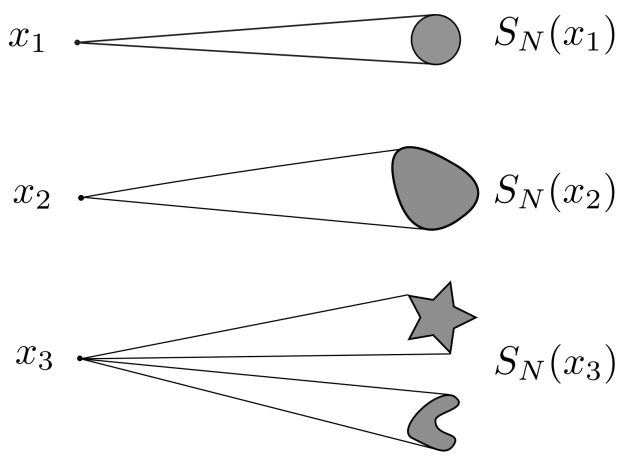


Figure 9.6. Uncertainty sets associated to three different codewords. Sets are not necessarily balls, they can be different across codewords, and also be composed of disconnected subsets.

that the uncertainty associated with the space $\mathscr{Y}$ of received codewords is $m_{\mathscr{Y}}(\mathscr{Y}) = 1$. We also let $V_N = N(x^*)$, where $x^* = \text{argmin}_{x \in \mathscr{X}} m_{\mathscr{Y}}(N(x))$. Thus, V_N is the set corresponding to the minimum uncertainty introduced by the noise mapping N.

Definition 19. (N, δ)-*distinguishable codebook.*

For any $0 \leq \delta < m_{\mathscr{Y}}(V_N)$*, a codebook* $\mathcal{X} \subseteq \mathscr{X}$ *is* (N, δ)*-distinguishable if for all* $x_1, x_2 \in \mathcal{X}$*, we have* $e_N(x_1, x_2) \leq \delta/|\mathcal{X}|$.

Definition 20. (N, δ)-*capacity.*

For any totally bounded, normed metric space $\mathscr{X}$*, channel with transition mapping* N*, and* $0 \leq \delta < m_{\mathscr{Y}}(V_N)$*, the* (N, δ)*-capacity of* $\mathscr{X}$ *is*

$$C_N^\delta = \sup_{\mathcal{X} \in \mathscr{X}_N^\delta} \log_2 |\mathcal{X}| \text{ bits}, \tag{9.105}$$

where $\mathscr{X}_N^\delta = \{\mathcal{X} : \mathcal{X}$ *is* (N, δ)*-distinguishable*$\}$.

We now relate our operational definition of capacity to the notion of UVs and mutual information introduced in Section 9.4. As usual, let X be the UV corresponding to the transmitted codeword and Y be the UV corresponding to the received codeword. For a channel with transition

mapping N, these UVs are such that for all $y \in [\![Y]\!]$ and $x \in [\![X]\!]$, we have

$$[\![Y|x]\!] = \{y \in [\![Y]\!] : y \in N(x)\}, \tag{9.106}$$

$$[\![X|y]\!] = \{x \in [\![X]\!] : y \in N(x)\}. \tag{9.107}$$

To measure the levels of association and disassociation between UVs X and Y, we use an uncertainty function $m_{\mathscr{X}}(.)$ defined over $\mathscr{X}$, and $m_{\mathscr{Y}}(.)$ is defined over $\mathscr{Y}$. The definition of feasible set is the same as the one given in (9.59). In our channel model, this feasible set depends on the transition mapping N through (9.106) and (9.107).

We can now state the non-stochastic channel coding theorem for channels with transition mapping N.

Theorem 43. *For any totally bounded, normed metric space $\mathscr{X}$, channel with transition mapping N satisfying* (9.106) *and* (9.107)*, and $0 \leq \delta < m_{\mathscr{Y}}(V_N)$, we have*

$$C_N^{\delta} = \sup_{X \in \mathscr{F}_{\tilde{\delta}}, \tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])} I_{\tilde{\delta}/|[\![X]\!]|}(Y;X) \ \ bits. \tag{9.108}$$

The proof is along the same lines as the one of Theorem 40 and is omitted.

Theorem 43 characterizes the capacity as the supremum of the mutual information over all codebooks in the feasible set. The following theorem shows that the same characterization is obtained if we optimize the right hand side in (9.108) over all codebooks in the space. It follows that by Theorem 43, rather than optimizing over all codebooks, a capacity achieving codebook can be found within the smaller class $\cup_{\tilde{\delta} \leq \delta / m_{\mathscr{Y}}(V_N)} \mathscr{F}_{\tilde{\delta}}$ of feasible sets with error at most $\delta / m_{\mathscr{Y}}(V_N)$.

Theorem 44. *The (N, δ)-capacity in* (9.108) *can also be written as*

$$C_N^{\delta} = \sup_{\substack{X : [\![X]\!] \subseteq \mathscr{X}, \\ \tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])}} I_{\tilde{\delta}/|[\![X]\!]|}(Y;X) \ bits. \tag{9.109}$$

The proof is along the same lines as the one of Theorem 41 and is omitted.

9.7 Capacity of Stationary Memoryless Uncertain Channels

We now consider the special case of stationary, memoryless, uncertain channels.

Let $\mathscr{X}^\infty$ be the space of $\mathscr{X}$-valued discrete-time functions $x : \mathbb{Z}_{>0} \to \mathscr{X}$, where $\mathbb{Z}_{>0}$ is the set of positive integers denoting the time step. Let $x(a : b)$ denote the function $x \in \mathscr{X}^\infty$ restricted over the time interval $[a, b]$. Let $\mathcal{X} \subseteq \mathscr{X}^\infty$ be a discrete set which represents a codebook. Also, let $\mathcal{X}(1 : n) = \cup_{x\in\mathcal{X}} x(1 : n)$ denote the set of all codewords up to time n, and $\mathcal{X}(n) = \cup_{x\in\mathcal{X}} x(n)$ the set of all codeword symbols in the codebook at time n. The codeword symbols can be viewed as the coefficients representing a continuous signal in an infinite-dimensional space. For example, transmitting one symbol per time step can be viewed as transmitting a signal of unit spectral support over time. The perturbation of the signal at the receiver due to the noise can be described as a displacement experienced by the corresponding codeword symbols $x(1), x(2), \ldots$. To describe this perturbation we consider the set-valued map $N^\infty : \mathscr{X}^\infty \to \mathscr{Y}^\infty$, associating any point in $\mathscr{X}^\infty$ to a set in $\mathscr{Y}^\infty$. For any transmitted codeword $x \in \mathcal{X} \subseteq \mathscr{X}^\infty$, the corresponding received codeword lies in the set

$$S_{N^\infty}(x) = \{y \in \mathscr{Y}^\infty : y \in N^\infty(x)\}. \tag{9.110}$$

Additionally, the noise set associated to $x(1 : n) \in \mathcal{X}(1 : n)$ is

$$S_{N^\infty}(x(1 : n)) = \{y(1 : n) \in \mathscr{Y}^n : y \in N^\infty(x)\}, \tag{9.111}$$

where

$$\mathscr{Y}^n = \underbrace{\mathscr{Y} \times \mathscr{Y} \times \cdots \times \mathscr{Y}}_{n}. \tag{9.112}$$

We are now ready to define stationary, memoryless, uncertain channels.

Definition 21. *A stationary memoryless uncertain channel is a transition mapping* $N^\infty : \mathscr{X}^\infty \to \mathscr{Y}^\infty$ *that can be factorized into identical terms describing the noise experienced by the codeword symbols. Namely, there exists a set-valued map* $N : \mathscr{X} \to \mathscr{Y}$ *such that for all* $n \in \mathbb{Z}_{>0}$ *and* $x(1:n) \in \mathscr{X}^\infty$*, we have*

$$S_{N^\infty}(x(1:n)) = N(x(1)) \times \ldots \times N(x(n)). \tag{9.113}$$

According to the definition, a stationary memoryless uncertain channel maps the nth input symbol into the nth output symbol in a way that does not depend on the symbols at other time steps, and the mapping is the same at all time steps. Since the channel can be characterized by the mapping N instead of N^∞, to simplify the notation in the following we use $S_N(.)$ instead of $S_{N^\infty}(.)$.

Another important observation is that the ϵ-noise channel that we considered in Section 9.5 may not admit a factorization like the one in (9.113). For example, consider the space to be equipped with the L^2 norm, the codeword symbols to represent the coefficients of an orthogonal representation of a transmitted signal, and the noise experienced by any codeword to be within a ball of radius ϵ. In this case, if a codeword symbol is perturbed by a value close to ϵ the perturbation of all other symbols must be close to zero. On the other hand, the general channels considered in Section 9.6 can be stationary and memoryless, if the noise acts on the coefficients in a way that satisfies (9.113).

For stationary memoryless uncertain channels, all received codewords lie in the set $\mathcal{Y} = \cup_{x\in\mathcal{X}} S_N(x)$, and the received codewords up to time n lie in the set $\mathcal{Y}(1:n) = \cup_{x\in\mathcal{X}} S_N(x(1:n))$. Then, for any $x_1(1:n), x_2(1:n) \in \mathcal{X}(1:n)$, we let

$$e_N(x_1(1:n), x_2(1:n)) = \frac{m_{\mathscr{Y}}(S_N(x_1(1:n)) \cap S_N(x_2(1:n)))}{m_{\mathscr{Y}}(\mathscr{Y}^n)}. \tag{9.114}$$

We also assume without loss of generality that at any time step n, the uncertainty associated

to the space $\mathscr{Y}^n$ of received codewords is $m_{\mathscr{Y}}(\mathscr{Y}^n) = 1$. We also let $V_N = N(x^*)$, where $x^* = \text{argmin}_{x\in\mathscr{X}} m_{\mathscr{Y}}(N(x))$. Thus, V_N is the set corresponding to the minimum uncertainty introduced by the noise mapping at a single time step. Finally, we let

$$V_N^n = \underbrace{V_N \times V_N \times \cdots \times V_N}_{n}. \tag{9.115}$$

Definition 22. (N, δ_n)*-distinguishable codebook.*

For all $n \in \mathbb{Z}_{>0}$ *and* $0 \leq \delta_n < m_{\mathscr{Y}}(V_N^n)$*, a codebook* $\mathcal{X}_n = \mathcal{X}(1:n)$ *is* (N, δ_n)*-distinguishable if for all* $x_1(1:n), x_2(1:n) \in \mathcal{X}_n$*, we have*

$$e_N(x_1(1:n), x_2(1:n)) \leq \delta_n/|\mathcal{X}_n|. \tag{9.116}$$

It immediately follows that for any (N, δ_n)-distinguishable codebook $\mathcal{X}_n$, we have

$$e_N(x(1:n)) = \sum_{\substack{x'(1:n)\in\mathcal{X}_n: \\ x'(1:n)\neq x(1:n)}} e_N(x(1:n), x'(1:n)) \leq \delta_n, \tag{9.117}$$

so that each codeword in $\mathcal{X}_n$ can be decoded correctly with confidence at least $1 - \delta_n$. Definion 22 guarantees even more namely, that the confidence of not confusing any pair of codewords is at least $1 - \delta_n/|\mathcal{X}_n|$.

We now associate to any sequence $\{\delta_n\}$ the largest distinguishable rate sequence $\{R_{\delta_n}\}$, whose elements represent the largest rates that satisfy that confidence sequence.

Definition 23. *Largest* $\{\delta_n\}$*-distinguishable rate sequence.*

For any sequence $\{\delta_n\}$*, the largest* $\{\delta_n\}$*-distinguishable rate sequence* $\{R_{\delta_n}\}$ *is such that for all* $n \in \mathbb{Z}_{>0}$ *we have*

$$R_{\delta_n} = \sup_{\mathcal{X}_n \in \mathscr{X}_N^{\delta_n}(n)} \frac{\log|\mathcal{X}_n|}{n} \textit{ bits per symbol}, \tag{9.118}$$

where

$$\mathscr{X}_N^{\delta_n}(n) = \{\mathcal{X}_n : \mathcal{X}_n \textit{ is } (N, \delta_n)\textit{-distinguishable}\}. \tag{9.119}$$

We say that any constant rate R that lays below the largest $\{\delta_n\}$-distinguishable rate sequence is $\{\delta_n\}$-distinguishable. Such a $\{\delta_n\}$-distinguishable rate ensures the existence of a sequence of distinguishable codes that, for all $n \in \mathbb{Z}_{>0}$, have rate at least R and confidence at least $1 - \delta_n$.

Definition 24. $\{\delta_n\}$*-distinguishable rate.*
For any sequence $\{\delta_n\}$*, a constant rate* R *is said to be* $\{\delta_n\}$*-distinguishable if for all* $n \in \mathbb{Z}_{>0}$*, we have*

$$R \leq R_{\delta_n}. \tag{9.120}$$

We call any $\{\delta_n\}$-distinguishable rate R achievable, if $\delta_n \to 0$ as $n \to \infty$. An achievable rate R ensures the existence of a sequence of distinguishable codes of rate at least R whose confidence tends to one as $n \to \infty$. It follows that in this case we can achieve communication at rate R with arbitrarily high confidence by choosing a sufficiently large codebook.

Definition 25. *Achievable rate.*
A constant rate R *is achievable if there exists a sequence* $\{\delta_n\}$ *such that* $\delta_n \to 0$ *as* $n \to \infty$*, and* R *is* $\{\delta_n\}$*-distinguishable.*

We now give a first definition of capacity as the supremum of the $\{\delta_n\}$-distinguishable rates. Using this definition, transmitting at constant rate below capacity ensures the existence of a sequence of codes that, for all $n \in \mathbb{Z}_{>0}$, have confidence at least $1 - \delta_n$.

Definition 26. $(N, \{\delta_n\})_*$ *capacity.*
For any stationary memoryless uncertain channel with transition mapping N*, and any given*

sequence $\{\delta_n\}$*, we let*

$$C_N(\{\delta_n\})_* = \sup\{R : R \textit{ is } \{\delta_n\}\textit{-distinguishable}\} \tag{9.121}$$

$$= \inf_{n\in\mathbb{Z}_{>0}} R_{\delta_n} \quad \textit{bits per symbol}. \tag{9.122}$$

Another definition of capacity arises if rather than the largest lower bound to the sequence of rates one considers the least upper bound for which we can transmit satisfying a given confidence sequence. Using this definition, transmitting at constant rate below capacity ensures the existence of a code that satisfies at least one confidence value along the sequence $\{\delta_n\}$.

Definition 27. $(N, \{\delta_n\})^*$ *capacity.*
For any stationary memoryless uncertain channel with transition mapping N*, and any given sequence* $\{\delta_n\}$*, we define*

$$C_N(\{\delta_n\})^* = \sup_{n\in\mathbb{Z}_{>0}} R_{\delta_n} \quad \textit{bits per symbol}. \tag{9.123}$$

Definitions 26 and 27 lead to non-stochastic analogues of Shannon's probabilistic and zero-error capacities, respectively.

First, consider Definition 26 and take the supremum of the achievable rates, rather than the supremum of the $\{\delta_n\}$-distinguishable rates. This means that we can pick any confidence sequence such that δ_n tends to zero as $n \to \infty$. In this way, we obtain the non-stochastic analogue of Shannon's probabilistic capacity, where δ_n plays the role of the probability of error and the capacity is the largest rate that can be achieved by a sequence of codebooks with an arbitrarily high confidence level.

Definition 28. $(N, \{\downarrow 0\})_*$ *capacity.*
For any stationary memoryless uncertain channel with transition mapping N*, we define the*

$(N, \{\downarrow 0\})_*$ *capacity as*

$$C_N(\{\downarrow 0\})_* = \sup\{R : R \textit{ is achievable}\} \tag{9.124}$$

$$= \sup_{\{\delta_n\}:\delta_n = o(1)} C_N(\{\delta_n\})_*. \tag{9.125}$$

Next, consider Definition 27 in the case $\{\delta_n\}$ is a constant sequence, namely for all $n \in \mathbb{Z}_{>0}$ we have $\delta_n = \delta \geq 0$. In this case, transmitting below capacity ensures the existence of a finite-length code that has confidence at least $1 - \delta$.

Definition 29. $(N, \delta)^*$ *capacity.*

For any stationary memoryless uncertain channel with transition mapping N, and any sequence $\{\delta_n\}$ such that for all $n \in \mathbb{Z}_{>0}$ we have $\delta_n = \delta \geq 0$, we define

$$C_N^{\delta *} = \sup_{n \in \mathbb{Z}_{>0}} R_{\delta_n} \quad \textit{bits per symbol}. \tag{9.126}$$

Letting $\delta = 0$, we obtain the zero-error capacity. In this case, below capacity there exists at a code with which we can transmit with full confidence.

We point out the key difference between Definitions 28 and 29. Transmitting below the $(N, \{\downarrow 0\})_*$ capacity, allows to achieve arbitrarily high confidence by increasing the codeword size. In contrast, transmitting below the $(N, \delta)^*$ capacity, ensures the existence of a fixed codebook that has confidence at least $1 - \delta$.

We now relate our notions of capacity to the *mutual information rate* between transmitted and received codewords. Let X be the UV corresponding to the transmitted codeword. This is a map $X : \mathscr{X}^\infty \to \mathcal{X}$ and $[\![X]\!] = \mathcal{X} \subseteq \mathscr{X}^\infty$. Restricting this map to a finite time $n \in \mathbb{Z}_{>0}$ yields another UV $X(n)$ and $[\![X(n)]\!] = \mathcal{X}(n) \subseteq \mathscr{X}$. Likewise, a codebook segment is an UV $X(a:b) = \{X(n)\}_{a \leq n \leq b}$, of marginal range

$$[\![X(a:b)]\!] = \mathcal{X}(a:b) \subseteq \mathscr{X}^{b-a+1}. \tag{9.127}$$

Likewise, let Y be the UV corresponding to the received codeword. It is a map $Y : \mathscr{Y}^\infty \to \mathcal{Y}$ and $[\![Y]\!] = \mathcal{Y} \subseteq \mathscr{Y}^\infty$. $Y(n)$ and $Y(a:b)$ are UVs, and $[\![Y(n)]\!] = \mathcal{Y} \subseteq \mathscr{Y}^\infty$ and $[\![Y(a:b)]\!] = \mathcal{Y}(a:b) \subseteq \mathscr{Y}^{b-a+1}$. For a stationary memoryless channel with transition mapping N, these UVs are such that for all $n \in \mathbb{Z}_{>0}$, $y(1:n) \in [\![Y(1:n)]\!]$ and $x(1:n) \in [\![X(1:n)]\!]$, and we have

$$[\![Y(1:n)|x(1:n)]\!] = \{y(1:n) \in [\![Y(1:n)]\!] : y(1:n) \in S_N(x(1:n))\}, \tag{9.128}$$

$$[\![X(1:n)|y(1:n)]\!] = \{x(1:n) \in [\![X(1:n)]\!] : y(1:n) \in S_N(x(1:n))\}. \tag{9.129}$$

Now, we define the largest δ_n-*mutual information rate* as the supremum mutual information per unit-symbol transmission that a codeword $X(1:n)$ can provide about $Y(1:n)$ with confidence at least $1 - \delta_n/|[\![X(1:n)]\!]|$.

Definition 30. *Largest δ_n-information rate.*
For all $n \in \mathbb{Z}_{>0}$, the largest δ_n-information rate from $X(1:n)$ to $Y(1:n)$ is

$$R^I_{\delta_n} = \sup_{\substack{X(1:n):[\![X(1:n)]\!] \subseteq \mathscr{X}^n, \\ \tilde{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![Y(1:n)]\!])}} \frac{I_{\tilde{\delta}/|[\![X(1:n)]\!]|}(Y(1:n); X(1:n))}{n}. \tag{9.130}$$

We let the feasible set at time n be

$$\begin{aligned} \mathscr{F}_\delta(n) = \{X(1:n) : [\![X(1:n)]\!] \subseteq \mathscr{X}^n, \text{ and either} \\ (X(1:n), Y(1:n)) \overset{d}{\leftrightarrow} (0, \delta/|[\![X(1:n)]\!]|) \text{ or} \\ (X(1:n), Y(1:n)) \overset{a}{\leftrightarrow} (1, \delta/|[\![X(1:n)]\!]|)\}. \end{aligned} \tag{9.131}$$

In the following theorem we establish the relationship between R_{δ_n} and $R^I_{\delta_n}$.

Theorem 45. *For any totally bounded, normed metric space $\mathscr{X}$, disrete-time space $\mathscr{X}^\infty$,*

stationary memoryless uncertain channel with transition mapping N *satisfying* (9.128) *and* (9.129)*, and sequence* $\{\delta_n\}$ *such that for all* $n \in \mathbb{Z}_{>0}$ *we have* $0 \leq \delta_n < m_{\mathscr{Y}}(V_N^n)$*, we have*

$$R_{\delta_n} = \sup_{\substack{X(1:n) \in \mathscr{F}_{\tilde{\delta}}(n), \\ \tilde{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![Y(1:n)]\!])}} \frac{I_{\tilde{\delta}/|[\![X(1:n)]\!]|}(Y(1:n); X(1:n))}{n}. \tag{9.132}$$

We also have

$$R_{\delta_n} = R_{\delta_n}^I. \tag{9.133}$$

Proof. The proof of the theorem is similar to the one of Theorem 40 and is given in Appendix 9.12.2. □

The following coding theorem is now an immediate consequence of Theorem 45 and of our capacity definitions.

Theorem 46. *For any totally bounded, normed metric space* $\mathscr{X}$*, disrete-time space* $\mathscr{X}^\infty$*, stationary memoryless uncertain channel with transition mapping* N *satisfying* (9.128) *and* (9.129)*, and sequence* $\{\delta_n\}$ *such that for all* $n \in \mathbb{Z}_{>0}$*,* $0 \leq \delta_n < m_{\mathscr{Y}}(V_N^n)$ *and* $0 \leq \delta < m_{\mathscr{Y}}(V_N^n)$*, we have*

$$1) \quad C_N(\{\delta_n\})_* = \inf_{n \in \mathbb{Z}_{>0}} R_{\delta_n}^I, \tag{9.134}$$

$$2) \quad C_N(\{\delta_n\})^* = \sup_{n \in \mathbb{Z}_{>0}} R_{\delta_n}^I, \tag{9.135}$$

$$3) \quad C_N(\{\downarrow 0\})_* = \sup_{\{\delta_n\}: \delta_n = o(1)} \inf_{n \in \mathbb{Z}_{>0}} R_{\delta_n}^I, \tag{9.136}$$

$$4) \quad C_N^{\delta *} = \sup_{n \in \mathbb{Z}_{>0}} R_{\delta_n}^I : \forall n \in \mathbb{Z}_{>0}, \delta_n = \delta. \tag{9.137}$$

Theorem 46 provides a multi-letter expressions of capacity, since the information rate $R_{\delta_n}^I$ depends on $I_{\tilde{\delta}/|[\![X(1:n)]\!]|}(Y(1:n); X(1:n))$ according to (9.130). Next, we establish conditions on the uncertainty function, confidence sequence, and class of stationary, memoryless channels

leading to the factorization of the mutual information and to single-letter expressions.

9.7.1 Factorization of the Mutual Information

To obtain sufficient conditions for the factorization of the mutual information we need to assume to work with a *product uncertainty function*.

Assumption 3. *(Product uncertainty function). The uncertainty function of a cartesian product of n sets can be factorized in the product of its terms, namely for any $n \in \mathbb{Z}_{>0}$ and $\mathscr{S} \subseteq [\![Y]\!]$ such that*

$$\mathscr{S} = \mathscr{S}_1 \times \mathscr{S}_2 \times \ldots \times \mathscr{S}_n, \tag{9.138}$$

we have

$$m_{\mathscr{Y}}(\mathscr{S}) = m_{\mathscr{Y}}(\mathscr{S}_1) \times m_{\mathscr{Y}}(\mathscr{S}_2) \times \ldots \times m_{\mathscr{Y}}(\mathscr{S}_n). \tag{9.139}$$

We also need to assume that the product uncertainty function satisfies a union bound.

Assumption 4. *(Union bound). For all $\mathscr{S}_1, \mathscr{S}_2 \subseteq [\![Y]\!]$, we have*

$$m_{\mathscr{Y}}(\mathscr{S}_1 \cup \mathscr{S}_2) \leq m_{\mathscr{Y}}(\mathscr{S}_1) + m_{\mathscr{Y}}(\mathscr{S}_2). \tag{9.140}$$

Before stating the main result of this section, we prove the following useful lemma.

Lemma 32. *Let $X(1:n)$ and $Y(1:n)$ be two UVs such that*

$$[\![X(1:n)]\!] = [\![X(1)]\!] \times [\![X(2)]\!] \ldots \times [\![X(n)]\!], \tag{9.141}$$

and for all $x(1:n) \in [\![X(1:n)]\!]$, we have

$$[\![Y(1:n)|x(1:n)]\!] = [\![Y(1)|x(1)]\!] \times \ldots [\![Y(n)|x(n)]\!]. \tag{9.142}$$

Let

$$0 \leq \delta < \min_{\substack{1 \leq i \leq n, \\ x(i) \in [\![X(i)]\!]}} m_{\mathscr{Y}}([\![Y(i)|x(i)]\!]). \tag{9.143}$$

Finally, let either

$$(X(1:n), Y(1;n)) \overset{d}{\leftrightarrow} (0, \delta^n), \text{ or} \tag{9.144}$$

$$(X(1:n), Y(1;n)) \overset{a}{\leftrightarrow} (1, \delta^n). \tag{9.145}$$

Under Assumption 3, we have:

1. *The cartesian product $\prod_{i=1}^n [\![Y(i)|X(i)]\!]_\delta^*$ is a covering of $[\![Y(1:n)]\!]$.*

2. *Every $\mathscr{S} \in \prod_{i=1}^n [\![Y(i)|X(i)]\!]_\delta^*$ is δ^n-connected and contains at least one singly δ^n-connected set of the form $[\![Y(1:n)|x(1:n)]\!]$.*

3. *For every singly δ^n-connected set of the form $[\![Y(1:n)|x(1:n)]\!]$, there exist a set in $\prod_{i=1}^n [\![Y(i)|X(i)]\!]_\delta^*$ containing it, namely for all $x(1:n) \in [\![X(1:n)]\!]$, there exists a set $\mathscr{S} \in \prod_{i=1}^n [\![Y(i)|X(i)]\!]_\delta^*$ such that $[\![Y(1:n)|x(1:n)]\!] \subseteq \mathscr{S}$.*

4. *For all $\mathscr{S}_1, \mathscr{S}_2 \in \prod_{i=1}^n [\![Y(i)|X(i)]\!]_\delta^*$, we have*

$$\frac{m_{\mathscr{Y}}(\mathscr{S}_1 \cap \mathscr{S}_2)}{m_{\mathscr{Y}}([\![Y(1:n)]\!])} \leq \delta(\hat{\delta}(n))^{n-1}, \tag{9.146}$$

where

$$\hat{\delta}(n) = \max_{\substack{1 \leq i \leq n, \\ \mathscr{S} \in [\![Y(i)|X(i)]\!]_\delta^*}} \frac{m_{\mathscr{Y}}(\mathscr{S})}{m_{\mathscr{Y}}([\![Y(i)]\!])}. \tag{9.147}$$

Proof. The proof is given in Appendix 9.12.3. □

Under Assumption 3 and Assumption 4, given two UVs that can be written in Cartesian product form and that are either associated at level $(0, \delta^n)$ or disassociated at level $(1, \delta^n)$, we now obtain an upper bound on the mutual information at level δ^n in terms of the sum of the mutual

information at level δ of their components. An analogous result in the stochastic setting states that the mutual information between two n-dimensional random variables $X^n = \{X_1, \ldots, X_n\}$ and $Y^n = \{Y_1, \ldots, Y_n\}$ is at most the sum of the component-wise mutual information, namely

$$I(X^n; Y^n) \leq \sum_{i=1}^{n} I(X_i; Y_i), \tag{9.148}$$

where $I(X;Y)$ represents the Shannon mutual information between two random variables X and Y. In contrast to the stochastic setting, here the mutual information is associated to a confidence parameter δ^n that is re-scaled to δ when this is decomposed into the sum of n terms.

Theorem 47. *Let $X(1:n)$ and $Y(1:n)$ be two UVs such that*

$$[\![X(1:n)]\!] = [\![X(1)]\!] \times [\![X(2)]\!] \ldots [\![X(n)]\!], \tag{9.149}$$

and for all $x(1:n) \in [\![X(1:n)]\!]$, we have

$$[\![Y(1:n)|x(1:n)]\!] = [\![Y(1)|x(1)]\!] \times \ldots [\![Y(n)|x(n)]\!]. \tag{9.150}$$

Also, let

$$0 \leq \delta < \frac{\min_{1\leq i\leq n, x(i)\in[\![X(i)]\!]} m_{\mathscr{Y}}([\![Y(i)|x(i)]\!])}{\max_{1\leq i\leq n} |[\![X(i)]\!]|}. \tag{9.151}$$

Finally, let either

$$(X(1:n), Y(1:n)) \overset{d}{\leftrightarrow} (0, \delta^n), \text{ or} \tag{9.152}$$

$$(X(1:n), Y(1:n)) \overset{a}{\leftrightarrow} (1, \delta^n). \tag{9.153}$$

Under Assumptions 3 and 2, we have

$$I_{\delta^n}(Y(1:n); X(1:n)) \leq \sum_{i=1}^{n} I_\delta(Y(i); X(i)). \tag{9.154}$$

Proof. First, we will show that for all $\mathscr{S} \in \prod_{i=1}^{n} [\![Y(i)|X(i)]\!]_{\delta}^{*}$, there exists a point $x_{\mathscr{S}}(1:n) \in [\![X(1:n)]\!]$ and a set $\mathscr{D}(\mathscr{S}) \in [\![Y(1:n)|X(1:n)]\!]_{\delta^n}^{*}$ such that

$$[\![Y(1:n)|x_{\mathscr{S}}(1:n)]\!] \subseteq \mathscr{S}, \tag{9.155}$$

$$[\![Y(1:n)|x_{\mathscr{S}}(1:n)]\!] \subseteq \mathscr{D}(\mathscr{S}). \tag{9.156}$$

Using this result, we will then show that

$$|\prod_{i=1}^{n} [\![Y(i)|X(i)]\!]_{\delta}^{*}| \geq |[\![Y(1:n)|X(1:n)]\!]_{\delta^n}^{*}|, \tag{9.157}$$

which immediately implies (9.154).

Let us begin with the first step. We have

$$\begin{aligned} \delta &\overset{(a)}{<} \frac{\min_{1\leq i\leq n, x(i)\in[\![X(i)]\!]} m_{\mathscr{Y}}([\![Y(i)|x(i)]\!])}{\max_{1\leq i\leq n} |[\![X(i)]\!]|} \\ &\overset{(b)}{\leq} \min_{1\leq i\leq n, x(i)\in[\![X(i)]\!]} m_{\mathscr{Y}}([\![Y(i)|x(i)]\!]), \end{aligned} \tag{9.158}$$

where (a) follows from (9.151), and (b) follows from the fact that for all $1 \leq i \leq n$, we have $|[\![X(i)]\!]| \geq 1$.

Now, consider a set $\mathscr{S} \in \prod_{i=1}^{n} [\![Y(i)|X(i)]\!]_{\delta}^{*}$. Using (9.158), by Lemma 32 part $2)$ we have that there exist a point $x'(1:n) \in [\![X(1:n)]\!]$ such that

$$[\![Y(1:n)|x'(1:n)]\!] \subseteq \mathscr{S}. \tag{9.159}$$

Now, using (9.158), part $1)$ in Lemma 32, and Definition 14, we have

$$\cup_{\mathscr{S}\in\prod_{i=1}^{n}[\![Y(i)|X(i)]\!]_{\delta}^{*}}\mathscr{S} = [\![Y(1:n)]\!] \\ = \cup_{\mathscr{D}\in[\![Y(1:n)|X(1:n)]\!]_{\delta^n}^{*}}\mathscr{D}. \tag{9.160}$$

Using (9.160) and Property 3 of Definition 14, there exists a set $\mathscr{D}(x'(1:n)) \in [\![Y(1:n)|X(1:n)]\!]_{\delta^n}^{*}$ such that

$$[\![Y(1:n)|x'(1:n)]\!] \subseteq \mathscr{D}(x'(1:n)). \tag{9.161}$$

Letting $x_{\mathscr{S}}(1:n) = x'(1:n)$ and $\mathscr{D}(\mathscr{S}) = \mathscr{D}(x'(1:n))$ in (9.159) and (9.161), we have that (9.155) and (9.156) follow.

We now proceed with proving (9.157). We distinguish two cases. In the first case, there exists two sets $\mathscr{S} \in \prod_{i=1}^{n}[\![Y(i)|X(i)]\!]_{\delta}^{*}$ and $\mathscr{D}_1 \in [\![Y(1:n)|X(1:n)]\!]_{\delta^n}^{*}$ such that

$$\mathscr{D}_1 \cap \mathscr{S} \setminus \mathscr{D}(\mathscr{S}) \neq \emptyset. \tag{9.162}$$

In the second case, the sets $\mathscr{S}$ and $\mathscr{D}_1$ satisfying (9.162) do not exist. We will show that the first case is not possible, and in the second case, we have that (9.157) holds.

To rule out the first case, consider two points

$$y_1(1:n) \in [\![Y(1:n)|x_{\mathscr{S}}(1:n)]\!] \subseteq \mathscr{D}(\mathscr{S}) \tag{9.163}$$

and

$$y_2(1:n) \in \mathscr{D}_1 \cap \mathscr{S} \setminus \mathscr{D}(\mathscr{S}). \tag{9.164}$$

If $(X(1:n), Y(1:n) \overset{a}{\leftrightarrow} (1, \delta^n)$, then we have

$$\begin{aligned}
&\delta^n |[\![X(1:n)]\!]| \\
&\overset{(a)}{<} \left(\frac{\min_{1\leq i\leq n, x(i)\in[\![X(i)]\!]} m_{\mathscr{Y}}([\![Y(i)|x(i)]\!])}{\max_{1\leq i\leq n} |[\![X(i)]\!]|} \right)^n |[\![X(1:n)]\!]| \\
&\overset{(b)}{\leq} \left(\min_{\substack{1\leq i\leq n, \\ x(i)\in[\![X(i)]\!]}} m_{\mathscr{Y}}([\![Y(i)|x(i)]\!]) \right)^n \\
&\overset{(c)}{\leq} \min_{x(1:n)\in[\![X(1:n)]\!]} m_{\mathscr{Y}}([\![Y(1:n)|x(1:n)]\!]) \\
&\overset{(d)}{\leq} \frac{\min_{x(1:n)\in[\![X(1:n)]\!]} m_{\mathscr{Y}}([\![Y(1:n)|x(1:n)]\!])}{m_{\mathscr{Y}}([\![Y(1:n)]\!])},
\end{aligned} \tag{9.165}$$

where (a) follows from (9.151), (b) follows from (9.149), and the fact that

$$|[\![X(1)]\!] \times [\![X(2)]\!] \ldots [\![X(n)]\!]| \leq (\max_{1\leq i\leq n} |[\![X(i)]\!]|)^n, \tag{9.166}$$

(c) follows from (9.150) and Assumption 3, and (d) follows from (9.9), and the facts that $[\![Y(1:n)]\!] \subseteq \mathscr{Y}^n$ and $m_{\mathscr{Y}}(\mathscr{Y}^n) = 1$. Combining (9.165), Assumption 4 and Lemma 37 in Appendix 9.12.4, we have that there exists a point $y(1:n) \in [\![Y(1:n)|x_{\mathscr{S}}(1:n)]\!]$ such that for all $[\![Y(1:n)|x(1:n)]\!] \in [\![Y(1:n)|X(1:n)]\!] \setminus \{[\![Y(1:n)|x_{\mathscr{S}}(1:n)]\!]\}$, we have.

$$y(1:n) \notin [\![Y(1:n)|x(1:n)]\!]. \tag{9.167}$$

Without loss of generality, let

$$y_1(1:n) = y(1:n). \tag{9.168}$$

It now follows that $y_1(1:n)$ and $y_2(1:n)$ cannot be δ^n-connected. This follows because

$$y_1(1:n) \in [\![Y(1:n)|x_{\mathscr{S}}(1:n)]\!], \tag{9.169}$$

$$y_2(1:n) \notin [\![Y(1:n)|x_{\mathscr{S}}(1:n)]\!], \tag{9.170}$$

(9.167) and $(X(1:n), Y(1:n) \overset{a}{\nleftrightarrow} (1, \delta^n)$, so that there does not exist a sequence $\{[\![Y(1:n)|x_i(1:n)]\!]\}_{i=1}^N$ such that for all $1 < i \leq N$

$$\frac{m_{\mathscr{Y}}([\![Y(1:n)|x_i(1:n)]\!] \cap [\![Y(1:n)|x_{i-1}(1:n)]\!])}{m_{\mathscr{Y}}([\![Y(1:n)]\!])} > \delta^n. \tag{9.171}$$

On the other hand, if $(X(1:n), Y(1:n) \overset{d}{\nleftrightarrow} (0, \delta^n)$, then using Theorem 38, we have that $[\![Y(1:n)|X(1:n)]\!]^*_{\delta^n}$ is a δ^n-isolated partition. Thus, $y_1(1:n)$ and $y_2(1:n)$ are not δ^n-connected, since $y_1(1:n) \in \mathscr{D}(\mathscr{S})$ and $y_2(1:n) \in \mathscr{D}_1 \cap \mathscr{S} \setminus \mathscr{D}(\mathscr{S})$. However, since $y_1(1:n), y_2(1:n) \in \mathscr{S}$ and $\mathscr{S}$ is δ^n-connected using (9.158) and $2)$ in Lemma 32, we have that $y_1(1:n) \overset{\delta^n}{\leftsquigarrow} y_2(1:n)$. This contradiction implies that $\mathscr{D}_1$ and $\mathscr{S}$ do not exist.

In the second case, if $\mathscr{S}$ and $\mathscr{D}_1$ do not exist, then for all $\mathscr{S}' \in \prod_{i=1}^n [\![Y(i)|X(i)]\!]^*_\delta$ and $\mathscr{D}' \in [\![Y(1:n)|X(1:n)]\!]^*_{\delta^n}$, we have

$$\mathscr{D}' \cap \mathscr{S}' \setminus \mathscr{D}(\mathscr{S}') = \emptyset, \tag{9.172}$$

which implies that

$$\begin{aligned}
\mathscr{S}' &\overset{(a)}{=} \cup_{\mathscr{D}' \in [\![Y(1:n)|X(1:n)]\!]^*_{\delta^n}} (\mathscr{S}' \cap \mathscr{D}') \\
&\overset{(b)}{\subseteq} \cup_{\mathscr{D}' \in [\![Y(1:n)|X(1:n)]\!]^*_{\delta^n}} \Big(\mathscr{D}(\mathscr{S}') \cup (\mathscr{S}' \cap \mathscr{D}' \setminus \mathscr{D}(\mathscr{S}')) \Big) \\
&= \mathscr{D}(\mathscr{S}') \cup_{\mathscr{D}' \in [\![Y(1:n)|X(1:n)]\!]^*_{\delta^n}} (\mathscr{S}' \cap \mathscr{D}' \setminus \mathscr{D}(\mathscr{S}')) \\
&\overset{(c)}{=} \mathscr{D}(\mathscr{S}'),
\end{aligned} \tag{9.173}$$

where (a) follows from (9.160), (b) follows from the trivial fact that for any three sets $\mathscr{A}$, $\mathscr{B}$

and $\mathscr{C}$,

$$\mathscr{A} \cap \mathscr{B} \subseteq \mathscr{C} \cup (\mathscr{A} \cap \mathscr{B} \setminus \mathscr{C}), \tag{9.174}$$

and (c) follows from (9.172). Combining (9.173) and (9.160), we have that

$$|\prod_{i=1}^{n} [\![Y(i)|X(i)]\!]_{\delta}^{*}| \geq |[\![Y(1:n)|X(1:n)]\!]_{\delta^n}^{*}|. \tag{9.175}$$

The statement of the theorem now follows. □

The following corollary shows that the bound in Theorem 47 is tight in the zero-error case.

Corollary 47.1. *Let $X(1:n)$ and $Y(1:n)$ satisfy (9.149) and* (9.150)*. Under Assumptions 3 and 4, we have that*

$$I_0(Y(1:n); X(1:n)) = \sum_{i=1}^{n} I_0(Y(i); X(i)). \tag{9.176}$$

Proof. The proof is along the same lines as the one of Theorem 47. For all $X(1:n)$ and $Y(1:n)$, if $\mathcal{A}(Y;X) = \emptyset$, then

$$(X(1:n), Y(1:n)) \overset{a}{\leftrightarrow} (1,0), \tag{9.177}$$

otherwise

$$(X(1:n), Y(1:n)) \overset{d}{\leftrightarrow} (0,0). \tag{9.178}$$

Hence, either (9.152) or (9.153) holds for $\delta = 0$. Now, by replacing $\delta = 0$ in $1) - 4)$ of Lemma 32, we have that $\prod_{i=1}^{n} [\![Y(i)|X(i)]\!]_0^*$ satisfies all the properties of a 0-overlap family. Combining this fact and Theorem 47, the statement of the corollary follows. □

9.7.2 Single letter expressions

We are now ready to present sufficient conditions leading to single-letter expressions for $C(\{\delta_n\})_*$, $C_N^{\delta *}$, $C(\{\delta_n\})^*$ and $C_N(\{\downarrow 0\})_*$. Under these conditions, the multi dimensional

optimization problem of searching for a codebook that achieves capacity over a time interval of size n can be reduced to searching for a codebook over a single time step.

First, we start with the single-letter expression for $C(\{\delta_n\})^*$.

Theorem 48. *For any stationary memoryless uncertain channel N and for any $0 \leq \delta_1 < m_{\mathscr{Y}}(V_N)$, let $\bar{X} \in \mathscr{F}_{\bar{\delta}}(1)$ be an UV over one time step associated with a one-dimensional codebook that achieves the capacity $C_N(\{\delta_1\})^* = C_N(\{\delta_1\})_* = R_{\delta_1}$, and let $\bar{Y}$ be the UV corresponding to the received codeword, namely*

$$\begin{aligned} C_N(\{\delta_1\})^* &= I_{\bar{\delta}/|[\![\bar{X}]\!]|}(\bar{Y};\bar{X}) \\ &= \sup_{\substack{X(1)\in\mathscr{F}_{\tilde{\delta}}(1):\\ \tilde{\delta}\leq\delta_1/m_{\mathscr{Y}}([\![Y(1)]\!])}} I_{\tilde{\delta}/|[\![X(1)]\!]|}(Y(1);X(1)). \end{aligned} \tag{9.179}$$

If for all one-dimensional codewords $x \in \mathscr{X} \setminus [\![\bar{X}]\!]$ there exists a set $\mathscr{S} \in [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}$ such that the uncertainty region $[\![Y|x]\!] \subseteq \mathscr{S}$, $\bar{\delta}(1+1/|[\![\bar{X}]\!]|) \leq \delta_1/m_{\mathscr{Y}}([\![\bar{Y}]\!])$, and for all $n > 1$ we have $0 \leq \delta_n \leq (\bar{\delta} m_{\mathscr{Y}}(V_N)/|[\![\bar{X}]\!]|)^n$, then under Assumptions 3 and 4 we have that the n-dimensional capacity

$$C_N(\{\delta_n\})^* = I_{\bar{\delta}/|[\![\bar{X}]\!]|}(\bar{Y};\bar{X}). \tag{9.180}$$

Proof. Let

$$[\![\bar{X}(1:n)]\!] = \underbrace{[\![\bar{X}]\!] \times \cdots \times [\![\bar{X}]\!]}_{n}, \tag{9.181}$$

and

$$[\![\bar{Y}(1:n)]\!] = \underbrace{[\![\bar{Y}]\!] \times \cdots \times [\![\bar{Y}]\!]}_{n}. \tag{9.182}$$

For all $n > 0$, we have

$$\begin{aligned}
\delta_n &\leq \left(\frac{\bar{\delta} m_{\mathscr{Y}}(V_N)}{|[\![\bar{X}]\!]|}\right)^n \\
&\overset{(a)}{\leq} \left(\frac{\delta_1 m_{\mathscr{Y}}(V_N)}{|[\![\bar{X}]\!]| m_{\mathscr{Y}}([\![\bar{Y}]\!])}\right)^n \\
&\overset{(b)}{\leq} \left(\frac{\delta_1}{|[\![\bar{X}]\!]|}\right)^n \\
&\overset{(c)}{<} (m_{\mathscr{Y}}(V_N))^n \\
&\overset{(d)}{=} m_{\mathscr{Y}}(V_N^n),
\end{aligned} \tag{9.183}$$

where (a) follows from the fact that $\bar{\delta} \leq \delta_1 / m_{\mathscr{Y}}([\![\bar{Y}]\!])$, (b) follows from $m_{\mathscr{Y}}(V_N) \leq m_{\mathscr{Y}}([\![\bar{Y}]\!])$, (c) follows from $\delta_1 < m_{\mathscr{Y}}(V_N)$ and $|[\![\bar{X}]\!]| \geq 1$, and (d) follows from Assumption 3.

We now proceed in three steps. First, using (9.183) and Theorem 45, we have

$$C_N(\{\delta_n\})^* = \sup_{n \in \mathbb{Z}_{>0}} R_{\delta_n} \geq R_{\delta_1} = R^I_{\delta_1} = I_{\bar{\delta}/|[\![\bar{X}]\!]|}(\bar{Y}; \bar{X}). \tag{9.184}$$

Second, we will show that for all $n \in \mathbb{Z}_{>0}$, we have

$$\sup_{\substack{X(1:n):[\![X(1:n)]\!] \subseteq \mathscr{X}^n, \\ \tilde{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![Y(1:n)]\!])}} I_{\tilde{\delta}/|[\![X(1:n)]\!]|}(Y(1:n); X(1:n)) \leq n I_{\bar{\delta}/|[\![\bar{X}]\!]|}(\bar{Y}; \bar{X}). \tag{9.185}$$

Finally, using (9.183), Theorem 45 and (9.185), for all $n \in \mathbb{Z}_{>0}$, we have that

$$R_{\delta_n} = R^I_{\delta_n} \leq I_{\bar{\delta}/|[\![\bar{X}]\!]|}(\bar{Y}; \bar{X}), \tag{9.186}$$

which implies

$$C_N(\{\delta_n\})^* = \sup_{n \in \mathbb{Z}_{>0}} R_{\delta_n} \leq I_{\bar{\delta}/|[\![\bar{X}]\!]|}(\bar{Y}; \bar{X}). \tag{9.187}$$

Using (9.184) and (9.187), the result (9.180) follows.

Now, we only need to prove (9.185). We will prove this by contradiction. Consider an

UV $X(1:n)$ and

$$\delta' \leq \delta_n / m_{\mathscr{Y}}([\![Y(1:n)]\!]), \tag{9.188}$$

such that

$$|[\![Y(1:n)|X(1:n)]\!]^*_{\delta'/|[\![X(1:n)]\!]|}| > |\prod_{i=1}^{n} [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}|. \tag{9.189}$$

We will show that (9.189) cannot hold using the following four claims, whose proofs appear in Appendix 9.12.5.

- **Claim 1:** If (9.189) holds, then there exists two UVs $\tilde{X}(1:n)$ and $\tilde{Y}(1:n)$ such that letting

$$\tilde{\delta} = \frac{\delta' m_{\mathscr{Y}}([\![Y(1:n)]\!])}{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!])}, \tag{9.190}$$

  we have

$$\tilde{\delta} \leq \frac{\delta_n}{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!])}, \tag{9.191}$$

$$(\tilde{X}(1:n), \tilde{Y}(1:n)) \overset{a}{\leftrightarrow} (1, \tilde{\delta}/|[\![\tilde{X}(1:n)]\!]|), \tag{9.192}$$

  and

$$|[\![\tilde{Y}(1:n)|\tilde{X}(1:n)]\!]^*_{\tilde{\delta}/|[\![\tilde{X}(1:n)]\!]|}| > |\prod_{i=1}^{n} [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}|. \tag{9.193}$$

- **Claim 2:** For all $\tilde{x}(1:n) \in [\![\tilde{X}(1:n)]\!]$, there exists a set $\mathscr{S} \in \prod_{i=1}^{n} [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}$ such that

$$[\![\tilde{Y}(1:n)|\tilde{x}(1:n)]\!] \subseteq \mathscr{S}. \tag{9.194}$$

- **Claim 3:** Using Claims 1 and 2, there exists a set $\mathscr{S} \in \prod_{i=1}^{n} [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}$ and two points

$\tilde{x}_1(1:n), \tilde{x}_2(1:n) \in [\![\tilde{X}(1:n)]\!]$ such that

$$[\![\tilde{Y}(1:n)|\tilde{x}_1(1:n)]\!], [\![\tilde{Y}(1:n)|\tilde{x}_2(1:n)]\!] \subset \mathscr{S}. \tag{9.195}$$

Also, there exists a $1 \leq i^* \leq n$ such that

$$\frac{m_{\mathscr{Y}}([\![\tilde{Y}(i^*)|\tilde{x}_1(i^*)]\!] \cap [\![\tilde{Y}(i^*)|\tilde{x}_2(i^*)]\!])}{(m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!]))^{1/n}} \leq \frac{\bar{\delta}}{|[\![\bar{X}]\!]|}. \tag{9.196}$$

- **Claim 4:** Using Claim 3, we have that there exist two UVs X' and Y', and $\delta^* \leq \delta_1/m_{\mathscr{Y}}([\![Y']\!])$ such that

$$|[\![Y'|X']\!]^*_{\delta^*/|[\![X']\!]|}| > |[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}|. \tag{9.197}$$

The result in Claim 4 contradicts (9.179). It follows that (9.189) cannot hold and the proof of Theorem 48 is complete.

□

Since $C_N^{\delta *}$ is a special case of $C_N(\{\delta_n\})^*$ for which the sequence δ_n is constant, it seem natural to use Theorem 48 to obtain a single-letter expression for $C_N^{\delta *}$ as well. However, the range of δ_n in Theorem 48 restricts the obtained single-letter expression for this case to the zero-error capacity C_N^{0*} only. To see this, note that δ_n in Theorem 48 is constrained to

$$\delta_n \leq (\bar{\delta} m_{\mathscr{Y}}(V_N)/|[\![\bar{X}]\!]|)^n < (m^2_{\mathscr{Y}}(V_N)/m_{\mathscr{Y}}([\![\bar{Y}]\!]))^n. \tag{9.198}$$

It follows that if $m_{\mathscr{Y}}(V_N) < m_{\mathscr{Y}}(\mathscr{Y}) = 1$, then we have $\delta_n = o(1)$ as $n \to \infty$. Hence, in this case a single letter expression for $C_N^{\delta *}$ can only be obtained for $\delta = 0$. On the other hand, if $m_{\mathscr{Y}}(V_N) = m_{\mathscr{Y}}(\mathscr{Y})$, then for any $0 \leq \delta < m_{\mathscr{Y}}(V_N)$ the codebook can only contain a single codeword and in this case we have $C_N^{\delta *} = 0$. We conclude that the only non-trivial single-letter

expression is obtained for the zero-error capacity, as stated next.

Corollary 48.1. *For any stationary memoryless uncertain channel N, let $\bar{X} \in \mathscr{F}_0(1)$ be an UV over one time step associated with a one-dimensional codebook that achieves the capacity $C_N(\{0\})^* = C_N(\{0\})_* = R_0$, and let $\bar{Y}$ be the UV corresponding to the received codeword, namely*

$$\begin{aligned} C_N(\{0\})^* &= I_0(\bar{Y};\bar{X}) \\ &= \sup_{X(1)\in\mathscr{F}_0(1)} I_0(Y(1);X(1)). \end{aligned} \tag{9.199}$$

If for all one-dimensional codewords $x \in \mathscr{X} \setminus [\![\bar{X}]\!]$, there exists a set $\mathscr{S} \in [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}$ such that the uncertainty region $[\![Y|x]\!] \subseteq \mathscr{S}$, then under Assumptions 3 and 4 we have that the n-dimensional zero-error capacity

$$C_N^{0*} = I_0(\bar{Y};\bar{X}). \tag{9.200}$$

Next, we present the sufficient conditions leading to the single letter expression for $C(\{\delta_n\})_*$.

Theorem 49. *For any stationary memoryless uncertain channel N, and for any $0 \leq \delta_1 < m_{\mathscr{Y}}(V_N)$, let $\bar{X} \in \mathscr{F}_{\bar{\delta}}(1)$ be an UV over one time step associated with a one-dimensional codebook that achieves the capacity $C_N(\{\delta_1\})_* = C_N(\{\delta_1\})^* = R_{\delta_1}$, and let $\bar{Y}$ be the UV corresponding to the received codeword, namely*

$$\begin{aligned} C_N(\{\delta_1\})_* &= I_{\bar{\delta}/|[\![\bar{X}]\!]|}(\bar{Y};\bar{X}) \\ &= \sup_{\substack{X(1)\in\mathscr{F}_{\tilde{\delta}}(1),\\ \tilde{\delta}\leq\delta_1/m_{\mathscr{Y}}([\![Y(1)]\!])}} I_{\tilde{\delta}/|[\![X(1)]\!]|}(Y(1);X(1)). \end{aligned} \tag{9.201}$$

Let

$$\hat{\delta} = \max_{\mathscr{S} \in [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}} \frac{m_{\mathscr{Y}}(\mathscr{S})}{m_{\mathscr{Y}}([\![\bar{Y}]\!])}. \tag{9.202}$$

If for all $n > 1$, we have $\bar{\delta}(\hat{\delta}|[\![\bar{X}]\!]|)^{n-1} \leq \delta_n < 1$, then under Assumption 3 we have that the n-dimensional capacity

$$C_N(\{\delta_n\})_* = I_{\bar{\delta}/|[\![\bar{X}]\!]|}(\bar{Y}; \bar{X}). \tag{9.203}$$

Proof. First, we show that for all $n \in \mathbb{Z}_{>0}$ and $\delta_n \geq \bar{\delta}(\hat{\delta}|[\![\bar{X}]\!]|)^{n-1}$, there exists a codebook $\tilde{\mathscr{X}}(1:n) \in \mathscr{X}_N^{\delta_n}(n)$ such that

$$|\tilde{\mathscr{X}}(1:n)| = |\prod_{i=1}^{n} [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}|. \tag{9.204}$$

This, along with Definition 23, implies that for all $n \in \mathbb{Z}_{>0}$ and $\delta_n \geq \bar{\delta}(\hat{\delta}|[\![\bar{X}]\!]|)^{n-1}$, we have

$$R_{\delta_n} \geq I_{\bar{\delta}/|[\![\bar{X}]\!]|}(\bar{Y}; \bar{X}), \tag{9.205}$$

and therefore

$$C_N(\{\delta_n\})_* = \inf_{n \in \mathbb{Z}_{>0}} R_{\delta_n} \geq I_{\bar{\delta}/|[\![\bar{X}]\!]|}(\bar{Y}; \bar{X}). \tag{9.206}$$

Second, we show that

$$C_N(\{\delta_n\})_* \leq I_{\bar{\delta}/|[\![\bar{X}]\!]|}(\bar{Y}; \bar{X}). \tag{9.207}$$

Thus, combining (9.206) and (9.207), we have that (9.203) follows and the proof is complete.

We now start with the first step of showing (9.206). Without loss of generality, we assume that $[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|} > 1$, otherwise

$$I_{\bar{\delta}/|[\![\bar{X}]\!]|}(\bar{Y}; \bar{X}) = 0, \tag{9.208}$$

and (9.203) holds trivially by the definition of $C_N(\{\delta_n\})_*$. Let

$$[\![\bar{X}(1:n)]\!] = \underbrace{[\![\bar{X}]\!] \times \cdots \times [\![\bar{X}]\!]}_{n}, \tag{9.209}$$

and

$$[\![\bar{Y}(1:n)]\!] = \underbrace{[\![\bar{Y}]\!] \times \cdots \times [\![\bar{Y}]\!]}_{n}. \tag{9.210}$$

Then, using $4)$ in Lemma 32 and the fact that N is a stationary memoryless channel, for all $\mathscr{S}_1, \mathscr{S}_2 \in \prod_{i=1}^{n} [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}$, we have

$$\begin{aligned} \frac{m_{\mathscr{Y}}(\mathscr{S}_1 \cap \mathscr{S}_2)}{m_{\mathscr{Y}}([\![\bar{Y}(1:n)]\!])} &\leq \frac{\bar{\delta}\hat{\delta}^{n-1}}{|[\![\bar{X}]\!]|} \\ &\overset{(a)}{\leq} \frac{\delta_n}{(|[\![\bar{X}]\!]|)^n} \\ &\overset{(b)}{=} \frac{\delta_n}{|[\![\bar{X}(1:n)]\!]|}, \end{aligned} \tag{9.211}$$

where (a) follows from the assumption in the theorem that $\delta_n \geq \bar{\delta}(\hat{\delta}|[\![\bar{X}]\!]|)^{n-1}$, and (b) follows from (9.209).

Using $2)$ in Lemma 32, we have that for all $\mathscr{S}_i \in \prod_{i=1}^{n} [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}$, there exists $x_i(1:n) \in [\![\bar{X}(1:n)]\!]$ such that

$$[\![\bar{Y}(1:n)|x_i(1:n)]\!] \subseteq \mathscr{S}_i. \tag{9.212}$$

Now, let

$$K = |\prod_{i=1}^{n} [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}|. \tag{9.213}$$

Consider a new UV $\tilde{X}(1:n)$ whose marginal range is composed of K elements of $[\![\bar{X}(1:n)]\!]$, namely

$$[\![\tilde{X}(1:n)]\!] = \{x_1(1:n), \ldots x_K(1:n)\}. \tag{9.214}$$

Let $\tilde{Y}(1:n)$ be the UV corresponding to the received variable. For all $x(1:n) \in [\![\tilde{X}(1:n)]\!]$,

we have $[\![\tilde{Y}(1:n)|x(1:n)]\!] = [\![\bar{Y}(1:n)|x(1:n)]\!]$ since N is a stationary memoryless channel. Using (9.211), (9.212), it now follows that for all $x(1:n), x'(1:n) \in [\![\tilde{X}(1:n)]\!]$, we have

$$\begin{aligned}\frac{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)|x(1:n)]\!] \cap [\![\tilde{Y}(1:n)|x'(1:n)]\!])}{m_{\mathscr{Y}}([\![\bar{Y}(1:n)]\!])} &\leq \frac{\delta_n}{|[\![\bar{X}(1:n)]\!]|} \\ &\overset{(a)}{\leq} \frac{\delta_n}{|[\![\tilde{X}(1:n)]\!]|}, \end{aligned} \tag{9.215}$$

where (a) follows from the fact that using (9.214), we have $[\![\tilde{X}(1:n)]\!] \subseteq [\![\bar{X}(1:n)]\!]$. This implies that for all $x(1:n), x'(1:n) \in [\![\tilde{X}(1:n)]\!]$,

$$\begin{aligned} e_N(x(1:n), x'(1:n)) &= \frac{m_{\mathscr{Y}}(S_N(x(1:n)) \cap S_N(x'(1:n)))}{m_{\mathscr{Y}}(\mathscr{Y}^n)} \\ &\overset{(a)}{=} \frac{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)|x(1:n)]\!] \cap [\![\tilde{Y}(1:n)|x'(1:n)]\!])}{m_{\mathscr{Y}}(\mathscr{Y}^n)} \\ &\overset{(b)}{\leq} \frac{\delta_n}{|[\![\tilde{X}(1:n)]\!]|} \frac{m_{\mathscr{Y}}([\![\bar{Y}(1:n)]\!])}{m_{\mathscr{Y}}(\mathscr{Y}^n)} \\ &\overset{(c)}{\leq} \frac{\delta_n}{|[\![\tilde{X}(1:n)]\!]|}, \end{aligned} \tag{9.216}$$

where (a) follows from the fact that N is stationary memoryless and for all $x(1:n) \in \mathscr{X}^n$, we have

$$[\![Y(1:n)|x(1:n)]\!] = S_N(x(1:n)), \tag{9.217}$$

(b) follows from (9.215), and (c) follows from (9.9) and $[\![\bar{Y}(1:n)]\!] \subseteq \mathscr{Y}^n$. This implies that the codebook $\tilde{\mathscr{X}}(1:n)$ corresponding to the UV $\tilde{X}(1:n)$ is (N, δ_n)-distinguishable.

It follows that (9.205) and (9.206) hold and the first step of the proof follows.

Now, we prove the second step. We have

$$\begin{aligned} C_N(\{\delta_n\})_* &= \inf_{n\in\mathbb{Z}_{>0}} R_{\delta_n} \\ &\overset{(a)}{\leq} R_{\delta_1} \\ &\overset{(b)}{=} R^I_{\delta_1} \\ &\overset{(c)}{=} I_{\bar{\delta}/|[\![\bar{X}]\!]|}(\bar{Y};\bar{X}), \end{aligned} \tag{9.218}$$

where (a) follows from the fact that

$$\inf_{n\in\mathbb{Z}_{>0}} R_{\delta_n} \leq R_{\delta_1}, \tag{9.219}$$

(b) follows from the fact that since $\delta_1 < m_{\mathscr{Y}}(V_N)$, we have that

$$R_{\delta_1} = R^I_{\delta_1}, \tag{9.220}$$

using Theorem 45, (c) follows from the fact that

$$R^I_{\delta_1} = I_{\bar{\delta}/|[\![\bar{X}]\!]|}(\bar{Y};\bar{X}), \tag{9.221}$$

using (9.132), (9.133) and (9.201). Hence, the second step of the proof follows. □

Finally, we present the sufficient conditions leading to the single letter expression for $C_N(\{\downarrow 0\})_*$.

Theorem 50. *Let $0 \leq \delta_1 < m_{\mathscr{Y}}(V_N)$. For any stationary memoryless uncertain channel N, let $X^* \in \mathscr{F}_{\delta^*}(1)$ be an UV over one time step associated with a one-dimensional codebook that achieves the largest one-dimensional δ_1-capacity, and let Y^* be the UV corresponding to the received codeword, namely X^* achieves* $\sup_{\delta_1<m_{\mathscr{Y}}(V_N)} C_N(\{\delta_1\})_* = \sup_{\delta_1<m_{\mathscr{Y}}(V_N)} C_N(\{\delta_1\})^* =$

$\sup_{\delta_1 < m_{\mathscr{Y}}(V_N)} R_{\delta_1}$, *and we have*

$$\begin{aligned}\sup_{\delta_1 < m_{\mathscr{Y}}(V_N)} C_N(\{\delta_1\})_* &= I_{\delta^*/|[\![X^*]\!]|}(Y^*; X^*) \\ &= \sup_{\delta_1 < m_{\mathscr{Y}}(V_N)} \sup_{\substack{X(1)\in\mathscr{F}_{\tilde{\delta}}(1),\\ \tilde{\delta}\le \delta_1/m_{\mathscr{Y}}([\![Y(1)]\!])}} I_{\tilde{\delta}/|[\![X(1)]\!]|}(Y(1); X(1)). \end{aligned} \tag{9.222}$$

Let

$$\hat{\delta}_* = \max_{\mathscr{S}\in[\![Y^*|X^*]\!]^*_{\delta^*/|[\![X^*]\!]|}} \frac{m_{\mathscr{Y}}(\mathscr{S})}{m_{\mathscr{Y}}([\![Y^*]\!])}. \tag{9.223}$$

If $\hat{\delta}_*|[\![X^*]\!]| < 1$*, then under Assumption 3 we have that the* n*-dimensional capacity*

$$C_N(\{\downarrow 0\})_* = I_{\delta^*/|[\![X^*]\!]|}(Y^*; X^*). \tag{9.224}$$

Proof. Consider a sequence of $\{\delta_n\}$ such that $\delta_1 = \delta^*$ and $\delta_n = \delta^*(\hat{\delta}_*|[\![X^*]\!]|)^{n-1}$ for $n > 1$. Then, using Theorem 49 for this sequence $\{\delta_n\}$, we have that

$$C_N(\{\delta_n\})_* = I_{\delta^*/|[\![X^*]\!]|}(Y^*; X^*). \tag{9.225}$$

Now, since $\hat{\delta}_*|[\![X^*]\!]| < 1$ using the assumption in the theorem, we have

$$\lim_{n\to\infty} \delta_n = 0. \tag{9.226}$$

Using (9.225) and (9.226), we have that

$$\begin{aligned} C_N(\{\downarrow 0\})_* &= \sup_{\{\delta'_n\}:\delta'_n = o(1)} C_N(\{\delta'_n\})_* \\ &\ge C_N(\{\delta_n\})_* \\ &= I_{\delta^*/|[\![X^*]\!]|}(Y^*; X^*). \end{aligned} \tag{9.227}$$

We also have

$$\begin{aligned} C_N(\{\downarrow 0\})_* &= \sup_{\{\delta'_n\}:\delta'_n=o(1)} \inf_{n\in\mathbb{Z}_{>0}} R_{\delta'_n} \\ &\overset{(a)}{\leq} \sup_{\{\delta'_n\}:\delta'_n=o(1)} R_{\delta'_1} \\ &\overset{(b)}{=} \sup_{\{\delta'_n\}:\delta'_n=o(1)} R^I_{\delta'_1} \\ &\overset{(c)}{=} \sup_{\delta'_1<m_{\mathscr{Y}}(V_N)} R^I_{\delta'_1} \\ &\overset{(d)}{=} I_{\delta^*/|[\![X^*]\!]|}(Y^*;X^*), \end{aligned} \tag{9.228}$$

where (a) follows from the fact that

$$\inf_{n\in\mathbb{Z}_{>0}} R_{\delta'_n} \leq R_{\delta'_1}, \tag{9.229}$$

(b) follows from the fact that since $\delta'_1 < m_{\mathscr{Y}}(V_N)$, we have

$$R_{\delta'_1} = R^I_{\delta'_1}, \tag{9.230}$$

using Theorem 45, (c) follows from the fact that $R^I_{\delta'_1}$ is only dependent on δ'_1 in the sequence $\{\delta'_n\}$, and (d) follows from (9.132), (9.133), and the definition of $I_{\delta^*/|[\![X^*]\!]|}(Y^*;X^*)$ in (9.222).

Combining (9.227) and (9.228), the statement of the theorem follows.

□

Table 9.1 shows a comparison between the sufficient conditions required to obtain single letter expressions for $C_N(\{\delta_n\})^*$, C_N^{0*}, $C_N(\{\delta_n\})_*$ and $C_N(\{\downarrow 0\})_*$. We point out that while in Theorems 48 and 49 any one-dimensional δ_1-capacity achieving codebook can be used to obtain the single-letter expression, in Theorem 50 the single-letter expression requires a codebook that achieves the largest capacity among all δ_1-capacity achieving codebooks. The sufficient conditions include in all cases Assumption 3, which is required to factorize the uncertainty

Table 9.1. Comparison of the Sufficient Conditions for the Existence of a Single-Letter expression

Sufficient Conditions	$C_N(\{\delta_n\})^*$ (Theorem 48)	C_N^{0*} (Corollary 48.1)	$C_N(\{\delta_n\})_*$ (Theorem 49)	$C_N(\{\downarrow 0\})_*$ (Theorem 50)
$m_{\mathscr{Y}}$ Satisfies Assumption 1	✓	✓	✓	✓
$m_{\mathscr{Y}}$ Satisfies Assumption 2	✓	✓		
1D Uncertainty Region Constraint	✓	✓		
Upper bound on δ_1	✓		✓	✓
Upper bound on $\{\delta_n\}_{n>1}$	✓			
Lower bound on $\{\delta_n\}_{n>1}$			✓	
Upper bound on $\bar{\delta}$	✓			
Upper bound on $\hat{\delta}_*$				✓

function over n dimensions, and leads to the key Lemma 32, and also to the upper bound on the mutual information between associated, or disassociated UVs in terms of the sum of the component-wise mutual information expressed by Theorem 47. The remaining conditions differ due to the different definitions of capacity.

9.8 Examples

To cast our sufficient conditions for the existence of single letter expressions of capacity in a concrete setting, we now provide some examples and compute the corresponding capacity.

In the following, we represent stationary memoryless uncertain channels in graph form. Let $\mathcal{G}(V, E)$ be a directed graph, where V is the set of vertices and E is the set of edges. The vertices represent input and output codeword symbols, namely $V = \mathscr{X} \cup \mathscr{Y}$. A directed edge from node $x \in \mathscr{X}$ to node $y \in \mathscr{Y}$, denoted by $x \to y$, shows that given symbol x is transmitted, y may be received at the output of the channel. It follows that for all $x \in \mathscr{X}$, the channel transition map representing the noise experienced by each codeword is given by

$$N(x) = \{y : (x \to y) \in E\}. \tag{9.231}$$

Example 3. *We consider a channel with*

$$\mathscr{X} = \mathscr{Y} = \{1, 2, 3, \ldots, 19\}. \tag{9.232}$$

To define the channel transition map, we let for all $x \in \{1, 2, 3, 4, 5, 6\}$

$$N(x) = \{1, 2, 3, 4, 5, 6, 11\}, \tag{9.233}$$

for all $x \in \{7, 8, 9, 10, 11, 12\}$

$$N(x) = \{7, 8, 9, 10, 11, 12, 2\}, \tag{9.234}$$

and for all $x \in \{13, 14, 15, 16, 17, 18, 19\}$

$$N(x) = \{13, 14, 15, 16, 17, 18, 19\}. \tag{9.235}$$

The corresponding graph is depicted in Figure 9.7. For any $\mathcal{Y}_n \subseteq \mathscr{Y}^n$*, we define the uncertainty function* $m_{\mathscr{Y}}(\mathcal{Y}_n)$ *in terms of cardinality*

$$m_{\mathscr{Y}}(\mathcal{Y}_n) = \frac{|\mathcal{Y}_n|}{|\mathscr{Y}^n|}. \tag{9.236}$$

Note that for all $n \in \mathbb{Z}_{>0}$*, we have* $m_{\mathscr{Y}}(\mathscr{Y}^n) = 1$.

It is easy to show that $m_{\mathscr{Y}}(.)$ *satisfies Assumption 3. Namely, for* $n = 1$*, we have that for all* $\mathcal{Y} \subseteq \mathscr{Y}$*,*

$$m_{\mathscr{Y}}(\mathcal{Y}) = \frac{|\mathcal{Y}|}{|\mathscr{Y}|}. \tag{9.237}$$

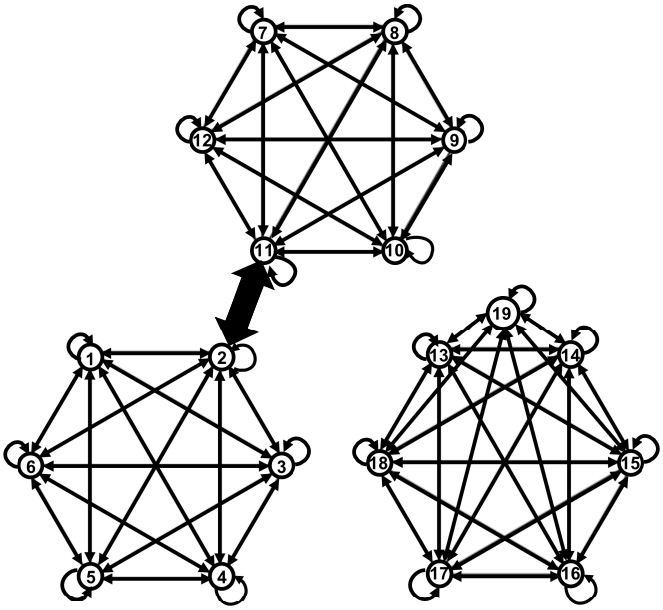


Figure 9.7. Channel described in Example 3. It consists of three complete graphs, and some additional edges. The thick solid arrow into node $y = 11$ represents multiple edges connecting all the nodes in the set $\{1, 2, 3, 4, 5, 6\}$ to node 11. Similarly, all the nodes in the set $\{7, 8, 9, 10, 11, 12\}$ are connected to node 2.

Let $\mathcal{Y}_n = \mathcal{Y}(1) \times \mathcal{Y}(2) \ldots \times \mathcal{Y}(n)$, where $\mathcal{Y}(i) \subseteq \mathscr{Y}$. Then, we have

$$\begin{aligned} m_{\mathscr{Y}}(\mathcal{Y}_n) &= m_{\mathscr{Y}}(\mathcal{Y}(1) \times \mathcal{Y}(2) \ldots \times \mathcal{Y}(n)) \\ &= \frac{|\mathcal{Y}(1) \times \mathcal{Y}(2) \ldots \times \mathcal{Y}(n)|}{|\mathscr{Y}^n|} \\ &\stackrel{(a)}{=} \frac{|\mathcal{Y}(1)|}{|\mathscr{Y}|} \frac{|\mathcal{Y}(2)|}{|\mathscr{Y}|} \cdots \frac{|\mathcal{Y}(n)|}{|\mathscr{Y}|} \\ &\stackrel{(b)}{=} m_{\mathscr{Y}}(\mathcal{Y}(1)) m_{\mathscr{Y}}(\mathcal{Y}(2)) \ldots m_{\mathscr{Y}}(\mathcal{Y}(n)), \end{aligned} \tag{9.238}$$

where (a) follows from the fact that for any two sets $\mathscr{S}_1$ and $\mathscr{S}_2$, $|\mathscr{S}_1 \times \mathscr{S}_2| = |\mathscr{S}_1||\mathscr{S}_2|$, and (b) follows from (9.237)*. It follows that $m_{\mathscr{Y}}(.)$ satisfies Assumption 3.*

A similar argument shows that $m_{\mathscr{Y}}(.)$ also satisfies Assumption 4. Namely, let $\mathcal{Y}_n =$

$\mathcal{Y}(1) \cup \mathcal{Y}(2) \ldots \cup \mathcal{Y}(n)$, *where* $\mathcal{Y}(i) \in \mathscr{Y}$. *Then, we have*

$$\begin{aligned} m_{\mathscr{Y}}(\mathcal{Y}_n) &= m_{\mathscr{Y}}(\mathcal{Y}(1) \cup \mathcal{Y}(2) \ldots \cup \mathcal{Y}(n)) \\ &= \frac{|\mathcal{Y}(1) \cup \mathcal{Y}(2) \ldots \cup \mathcal{Y}(n)|}{|\mathscr{Y}|} \\ &\overset{(a)}{\leq} \frac{|\mathcal{Y}(1)|}{|\mathscr{Y}|} + \frac{|\mathcal{Y}(2)|}{|\mathscr{Y}|} + \ldots + \frac{|\mathcal{Y}(n)|}{|\mathscr{Y}|} \\ &\overset{(b)}{=} m_{\mathscr{Y}}(\mathcal{Y}(1)) + m_{\mathscr{Y}}(\mathcal{Y}(2)) + \ldots + m_{\mathscr{Y}}(\mathcal{Y}(n)), \end{aligned} \tag{9.239}$$

where (a) *follows from the fact that for any two sets* $\mathscr{S}_1$ *and* $\mathscr{S}_2$, $|\mathscr{S}_1 \cup \mathscr{S}_2| \leq |\mathscr{S}_1| + |\mathscr{S}_2|$, *and* (b) *follows from* (9.237). *It follows that* $m_{\mathscr{Y}}(.)$ *satisfies Assumption 4.*

We now compute the capacity $C_N(\{\delta_n\})^*$ *for* $\delta_1 = 2/9$ *and for all* $n > 1$ $\delta_n = (7/342)^n$.

Since V_N *contains seven elements, we have that* $m_{\mathscr{Y}}(V_N) = 7/19$, *and* $\delta_1 < m_{\mathscr{Y}}(V_N)$. *Consider an UV* $\bar{X}$ *representing a one-dimensional codebook such that*

$$[\![\bar{X}]\!] = \{1, 7, 13\}. \tag{9.240}$$

It follows that the corresponding output UV $\bar{Y}$ *is such that*

$$[\![\bar{Y}]\!] = \{1, 2, 3, \ldots, 18, 19\}. \tag{9.241}$$

Letting $\bar{\delta} = 1/6$, *we have that* $\bar{\delta}/|[\![X]\!]| = 1/18$ *and the overlap family*

$$[\![\bar{Y}|\bar{X}]\!]^*_{1/18} = \{\mathscr{S}_1, \mathscr{S}_2\}, \tag{9.242}$$

where

$$\mathscr{S}_1 = \cup_{x \in \{1,7\}} N(x), \tag{9.243}$$

$$\mathscr{S}_2 = \cup_{x \in \{13\}} N(x). \tag{9.244}$$

We now show that $\bar{X}$ satisfies the sufficient conditions in Theorem 48. First, we note that for all $x \in \{2,3,4,5,6,8,9,10,11,12\}$, we have that $[\![\bar{Y}|x]\!] \subseteq \mathscr{S}_1$, and for all $x \in \{14,15,16,17,18,19\}$, we have $[\![\bar{Y}|x]\!] \subseteq \mathscr{S}_2$. It follows that for all $x \in \mathscr{X} \setminus [\![\bar{X}]\!] = \{2,3,4,5,6,8,9,10,11,12,14,15,16,17,18,19\}$, the uncertainity region $[\![\bar{Y}|x]\!] \subseteq \mathscr{S}$, where the set $\mathscr{S} \in [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}$. Second, we have that

$$\bar{\delta}(1 + 1/|[\![\bar{X}]\!]|) = 2/9 \leq \delta_1/m_{\mathscr{Y}}([\![\bar{Y}]\!]). \tag{9.245}$$

Third, we note that $\bar{\delta} = 1/6$,

$$\bar{\delta} m_{\mathscr{Y}}(V_N)/|[\![\bar{X}]\!]| = 7/342. \tag{9.246}$$

It follows that for all $n > 1$, we have that $\delta_n \leq (\bar{\delta} m_{\mathscr{Y}}(V_N)/|[\![\bar{X}]\!]|)^n$.

Since all the sufficient conditions in Theorem 48 are satisfied, we have

$$C_N(\{\delta_n\})^* = \log_2 |[\![\bar{Y}|\bar{X}]\!]^*_{1/18}| = 1. \tag{9.247}$$

Example 4. *We now consider the same channel as in Example 3, shown in Figure 9.7, and we compute the capacity C_N^{0*}. We consider the one-dimensional codebook $\bar{X}$ in* (9.240)*, and the corresponding output UV $\bar{Y}$ in* (9.241)*. Then, we have*

$$[\![\bar{Y}|\bar{X}]\!]^*_0 = \{\mathscr{S}_1, \mathscr{S}_2\}, \tag{9.248}$$

where

$$\mathscr{S}_1 = \cup_{x \in \{1,7\}} N(x), \tag{9.249}$$

$$\mathscr{S}_2 = \cup_{x \in \{13\}} N(x). \tag{9.250}$$

We now show that $\bar{X}$ satisfies the sufficient conditions in Corollary 48.1. We note that for all $x \in \{2,3,4,5,6,8,9,10,11,12\}$, we have that $[\![\bar{Y}|x]\!] \subseteq \mathscr{S}_1$, and for all $x \in \{14,15,16,17,18,19\}$, we have $[\![\bar{Y}|x]\!] \subseteq \mathscr{S}_2$. It follows that for all $x \in \mathscr{X} \setminus [\![\bar{X}]\!] = \{2,3,4,5,6,8,9,10,11,12,14,15,16,17,18,19\}$, the uncertainity region $[\![\bar{Y}|x]\!] \subseteq \mathscr{S}$, where $\mathscr{S} \in [\![\bar{Y}|\bar{X}]\!]_0^*$.

Since all the sufficient conditions in Corollary 48.1 are satisfied, we have

$$C_N^{0*} = \log_2 |[\![\bar{Y}|\bar{X}]\!]_0^*| = 1. \tag{9.251}$$

Example 5. *We now consider the same channel as in Example 3, shown in Figure 9.7, and we compute the capacity $C_N(\{\delta_n\})_*$ for $\delta_1 = (2/6)^3$ and for all $n > 1$ $\delta_n = (2/6)^3((7/19)^3 3)^{n-1}$.*

For any $\mathcal{Y}_n \subseteq \mathscr{Y}^n$, we define the uncertainty function $m_{\mathscr{Y}}(\mathcal{Y}_n)$ in terms of cardinality

$$m_{\mathscr{Y}}(\mathcal{Y}_n) = \left(\frac{|\mathcal{Y}_n|}{|\mathscr{Y}^n|}\right)^3. \tag{9.252}$$

Note that for all $n \in \mathbb{Z}_{>0}$, we have $m_{\mathscr{Y}}(\mathscr{Y}^n) = 1$. It is easy to show that $m_{\mathscr{Y}}(.)$ satisfies Assumption 3. For $n = 1$, we have that for all $\mathcal{Y} \subseteq \mathscr{Y}$,

$$m_{\mathscr{Y}}(\mathcal{Y}) = \left(\frac{|\mathcal{Y}|}{|\mathscr{Y}|}\right)^3. \tag{9.253}$$

Let $\mathcal{Y}_n = \mathcal{Y}(1) \times \mathcal{Y}(2) \ldots \times \mathcal{Y}(n)$, where $\mathcal{Y}(i) \in \mathscr{Y}$. Then, we have

$$\begin{aligned} m_{\mathscr{Y}}(\mathcal{Y}_n) &= m_{\mathscr{Y}}(\mathcal{Y}(1) \times \mathcal{Y}(2) \ldots \times \mathcal{Y}(n)) \\ &= \left(\frac{|\mathcal{Y}(1) \times \mathcal{Y}(2) \ldots \times \mathcal{Y}(n)|}{|\mathscr{Y}^n|}\right)^3 \\ &\overset{(a)}{=} \left(\frac{|\mathcal{Y}(1)|}{|\mathscr{Y}|}\right)^3 \left(\frac{|\mathcal{Y}(2)|}{|\mathscr{Y}|}\right)^3 \ldots \left(\frac{|\mathcal{Y}(n)|}{|\mathscr{Y}|}\right)^3 \\ &\overset{(b)}{=} m_{\mathscr{Y}}(\mathcal{Y}(1)) m_{\mathscr{Y}}(\mathcal{Y}(2)) \ldots m_{\mathscr{Y}}(\mathcal{Y}(n)), \end{aligned} \tag{9.254}$$

where (a) follows from the fact that for any two sets $\mathscr{S}_1$ and $\mathscr{S}_2$, $|\mathscr{S}_1 \times \mathscr{S}_2| = |\mathscr{S}_1||\mathscr{S}_2|$, and (b)

follows from (9.253)*. It follows that* $m_{\mathscr{Y}}(.)$ *satisfies Assumption 3.*

Since $m_{\mathscr{Y}}(V_N) = (7/19)^3$*, we have* $\delta_1 < m_{\mathscr{Y}}(V_N)$*. We consider an UV* $\bar{X}$ *representing a one-dimensional codebook, a corresponding output UV* $\bar{Y}$*, and* $\bar{\delta} = (2/6)^3$*, so that*

$$[\![\bar{X}]\!] = \{1, 7, 13\}, \tag{9.255}$$

$$[\![\bar{Y}]\!] = \{1, 2, \ldots, 19\}, \tag{9.256}$$

$$[\![\bar{Y}|\bar{X}]\!]^*_{1/81} = \{\mathscr{S}_1, \mathscr{S}_2, \mathscr{S}_3\}, \tag{9.257}$$

where

$$\mathscr{S}_1 = \cup_{x\in\{1\}} N(x), \tag{9.258}$$

$$\mathscr{S}_2 = \cup_{x\in\{7\}} N(x), \tag{9.259}$$

$$\mathscr{S}_3 = \cup_{x\in\{13\}} N(x). \tag{9.260}$$

Since $\bar{\delta} = \delta_1 = (2/6)^3$*, we have*

$$\hat{\delta} = \max_{\mathscr{S}\in[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}} \frac{m_{\mathscr{Y}}(\mathscr{S})}{m_{\mathscr{Y}}([\![\bar{Y}]\!])} = \left(\frac{7}{19}\right)^3. \tag{9.261}$$

It follows that for all $n > 1$*, we have* $\delta_n \geq \bar{\delta}(\hat{\delta}|[\![\bar{X}]\!]|)^{n-1}$ *and all the sufficient conditions in Theorem 49 are satisfied, so that*

$$C_N(\{\delta_n\})_* = \log_2 |[\![\bar{Y}|\bar{X}]\!]^*_{1/81}| = \log_2(3). \tag{9.262}$$

Example 6. *We again consider the same channel and the same uncertainty function as in Example 5, shown in Figure 9.7 and* (9.252)*, respectively. We compute the capacity* $C_N(\{\downarrow 0\})_*$*. Consider an UV* X^* *representing a one-dimensional codebook such that*

$$[\![X^*]\!] = \{1, 7, 13\}. \tag{9.263}$$

It follows that the corresponding output UV Y^* *is such that*

$$[\![Y^*]\!] = \{1, 2, 3 \ldots 19\}. \tag{9.264}$$

Letting $\delta^* = (2/6)^3$*, we have that* $\delta^*/|[\![X^*]\!]| = 1/81$ *and the overlap family is*

$$[\![Y^*|X^*]\!]^*_{\delta^*/|[\![X^*]\!]|=1/81} = \{\mathscr{S}_1, \mathscr{S}_2, \mathscr{S}_3\}, \tag{9.265}$$

where

$$\mathscr{S}_1 = \cup_{x\in\{1\}} N(x), \tag{9.266}$$

$$\mathscr{S}_2 = \cup_{x\in\{7\}} N(x), \tag{9.267}$$

$$\mathscr{S}_3 = \cup_{x\in\{13\}} N(x). \tag{9.268}$$

Since V_N *contains seven elements, we have that* $m_{\mathscr{Y}}(V_N) = (7/19)^3$*, and* $\delta^* < m_{\mathscr{Y}}(V_N)$*. Also, we have that*

$$\hat{\delta}_* = \max_{\mathscr{S}\in[\![Y^*|X^*]\!]^*_{\delta^*/|[\![X^*]\!]|}} \frac{m_{\mathscr{Y}}(\mathscr{S})}{m_{\mathscr{Y}}([\![Y^*]\!])} = \left(\frac{7}{19}\right)^3. \tag{9.269}$$

It follows that $\hat{\delta}_*|[\![X^*]\!]| < 1$.

Since all the sufficient conditions in Theorem 50 are satisfied, we have

$$C_N(\{\downarrow 0\})_* = \log_2 |[\![\bar{Y}|\bar{X}]\!]^*_{2/18}| = \log_2(3). \tag{9.270}$$

9.8.1 Discussion

The results in our examples show that for the channel presented in Figure 9.7, and the uncertainty function (9.236), there exists a vanishing sequence $\delta_1 = 2/9, \{\delta_n\}_2^\infty = \{(7/342)^n\}$ such that

$$C_N^{0*} = C_N(\{\delta_n\})^*. \tag{9.271}$$

For the same channel and uncertainty function, there is another vanishing sequence $\delta_1 = 4/9, \{\delta_n\}_2^\infty = \{(14/342)^n\}$, such that

$$C_N^{0*} < C_N(\{\delta_n\})^*. \tag{9.272}$$

On the other hand, for the same channel using the uncertainty function (9.252), there exists a vanishing sequence $\delta_1 = (2/6)^3, \{\delta_n\}_2^\infty = \{(2/6)^3(3(7/19)^3)^{n-1}\}$ such that

$$C_N(\{\delta_n\})_* = C_N(\{\downarrow 0\})_*. \tag{9.273}$$

For the same channel and uncertainty function (9.252), there exists another sequence $\delta_1 = (2/19)^3, \{\delta_n\}_2^\infty = (2/19)^3(3(12/19)^3)^{n-1}$ such that

$$C_N(\{\delta_n\})_* < C_N(\{\downarrow 0\})_*. \tag{9.274}$$

9.9 Applications

We now discuss some applications of the developed non-stochastic theory.

9.9.1 Error Correction in Adversarial Channels

Various adversarial channel models have been considered in the literature. A popular one considers a binary alphabet, and a codeword of length n that is sent from the transmitter to the receiver. The channel can flip at most a fraction $0 < \tau \leq 1$ of the n symbols in an arbitrary fashion [6]. In this case, the input and output spaces are $\mathscr{X}^n = \mathscr{Y}^n = \{0,1\}^n$, and a codebook is $\mathcal{X}_n \subseteq \mathscr{X}^n$. Due to the constraint on the total number of bit flips, the channel is non-stationary and with memory. For any $x \in \mathscr{X}^n$, we can let the norm be the Hamming distance from the n-dimensional all zero vector representing the origin of the space, namely

$$\|x\| = H(x, \{0\}_n) \leq n. \tag{9.275}$$

In this framework, for any transmitted codeword $x \in \mathcal{X}_n$, the set of possible received codewords is

$$S_{\tau n}(x) = \{y \in \mathscr{Y}^n : H(x,y) \leq \tau n\}, \tag{9.276}$$

where τn is the analogous of a noise range $\epsilon_n = \epsilon n$ in the non-stochastic channel model described in Section 9.5.

For all $x_1, x_2 \in \mathcal{X}_n$, the equivocation region corresponds to $S_{\tau n}(x_1) \cap S_{\tau n}(x_2)$ and for any $\mathscr{S} \subseteq \mathscr{Y}^n$, we can define the uncertainty function

$$m_{\mathscr{Y}^n}(\mathscr{S}) = \begin{cases} D(\mathscr{S}) + 1, \text{ if } \mathscr{S} \neq \emptyset, \\ 0, \text{ otherwise }, \end{cases} \tag{9.277}$$

where D indicates diameter, namely

$$D(\mathscr{S}) = \max_{y_1, y_2 \in \mathscr{S}} H(y_1, y_2). \tag{9.278}$$

With this definition, we have that

$$m_{\mathscr{Y}^n}(\mathscr{Y}^n) = n + 1. \tag{9.279}$$

For all $x_1, x_2 \in \mathcal{X}_n$, we let the error

$$e_{\tau n}(x_1, x_2) = \frac{m_{\mathscr{Y}^n}(S_{\tau n}(x_1) \cap S_{\tau n}(x_2))}{n+1}. \tag{9.280}$$

As usual, we say that a codebook $\mathcal{X}_n$ is $(\tau n, \delta_n)$-distinguishable if $e_{\tau n}(x_1, x_2) \leq \delta_n / |\mathcal{X}_n|$, and for all $n \in \mathbb{Z}_{>0}$ the $(\tau n, \delta_n)$ capacity is

$$C_{\tau n}^{\delta_n} = \sup_{\mathcal{X}_n \in \mathscr{X}_{\tau n}^{\delta_n}} \log_2(|\mathcal{X}_n|), \tag{9.281}$$

where $\mathscr{X}_{\tau n}^{\delta_n} = \{\mathcal{X}_n : \mathcal{X}_n \text{ is } (\tau n, \delta_n)\text{-distinguishable}\}$.

We now show that any $(\tau n, \delta_n)$-distinguishable codebook $\mathcal{X}_n$ can be used to correct a certain number of bit flips. This number depends on how far apart any two codewords are, and a lower bound on this distance can be expressed in terms of the diameter of the equivocation set and of the amount of perturbation introduced by the channel, see Figure 9.8. The following theorem provides a lower bound on the Hamming distance $H(x_1, x_2)$ between any two codewords. The number of bit flips that can be corrected can then be computed using the well-known formula $\lfloor (H(x_1, x_2) - 1)/2 \rfloor$. Finally, we point out that non-stochastic, adversarial, error correcting codes are of interest and have been studied in the context of multi-label classification in machine learning [65], and to improve the robustness of neural networks to adversarial attacks [218].

Theorem 51. *Given a channel satisfying* (9.276)*, if a codebook $\mathcal{X}_n$ is $(\tau n, \delta_n)$-distinguishable, then for all $x_1, x_2 \in \mathcal{X}_n$, we have*

$$H(x_1, x_2) \geq \left(\frac{2\tau n}{n+1} - \frac{\delta_n}{|\mathcal{X}_n|} \right)(n+1) + 1. \tag{9.282}$$

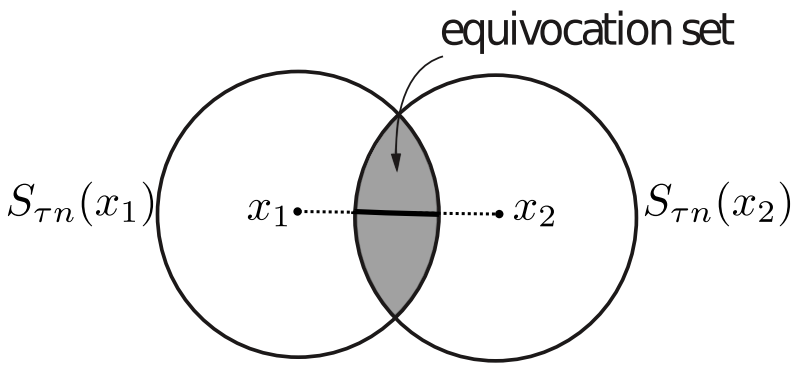


Figure 9.8. The Hamming distance between any two overlapping codewords depends on the code parameters τn and δ_n.

Proof. Let $\mathcal{X}_n \in \mathscr{X}_{\tau n}^{\delta}$. Then, for all $x_1, x_2 \in \mathcal{X}_n$, we have

$$\begin{aligned} e_{\tau n}(x_1, x_2) &= \frac{m_{\mathscr{Y}^n}(S_{\tau n}(x_1) \cap S_{\tau n}(x_2))}{n+1}, \\ &\leq \frac{\delta_n}{|\mathcal{X}_n|}. \end{aligned} \tag{9.283}$$

First, we consider the case when

$$x_1, x_2 \in S_{\tau n}(x_1) \cap S_{\tau n}(x_2). \tag{9.284}$$

Let $B(x_1, x_2)$ be the boundary of the equivocation set, namely

$$\begin{aligned} B(x_1, x_2) = \{x \in S_{\tau n}(x_1) \cap S_{\tau n}(x_2)) : \exists x' \notin S_{\tau n}(x_1) \cap S_{\tau n}(x_2)) \\ \text{such that } H(x, x') = 1\}. \end{aligned} \tag{9.285}$$

By (9.284), there exist $A, B \in B(x_1, x_2)$ such that

$$H(A, B) = H(A, x_1) + H(x_1, x_2) + H(x_2, B). \tag{9.286}$$

Then, we have

$$\begin{aligned} H(A,B) &= H(A,x_1) + H(x_1,x_2) + H(x_2,B) \\ &\overset{(a)}{=} 2\tau n - 2H(x_1,x_2) + H(x_1,x_2) \\ &= 2\tau n - H(x_1,x_2), \end{aligned} \tag{9.287}$$

where (a) follows from the fact that $A, B \in B(x_1, x_2)$, which implies

$$H(A,x_1) = H(A,x_2) - H(x_1,x_2) = \tau n - H(x_1,x_2), \tag{9.288}$$

and

$$H(B,x_2) = H(B,x_1) - H(x_1,x_2) = \tau n - H(x_1,x_2). \tag{9.289}$$

We now have that

$$\begin{aligned} H(x_1,x_2) &= 2\tau n - H(A,B) \\ &\overset{(a)}{\geq} 2\tau n - D(S_{\tau n}(x_1) \cap S_{\tau n}(x_2)), \end{aligned} \tag{9.290}$$

where (a) follows from the fact that using $A, B \in S_{\tau n}(x_1) \cap S_{\tau n}(x_2))$, we have $H(A,B) \leq D(S_{\tau n}(x_1) \cap S_{\tau n}(x_2))$.

Now, we consider the case when

$$x_1, x_2 \notin S_{\tau n}(x_1) \cap S_{\tau n}(x_2). \tag{9.291}$$

In this case, we have

$$H(x_1,x_2) \geq 2\tau n - D(S_{\tau n}(x_1) \cap S_{\tau n}(x_2)). \tag{9.292}$$

The result now follows by combining (9.290), (9.292) and (9.283).

□

9.9.2 Robustness of Neural Networks to Adversarial Attacks

Motivated by security considerations, the robustness of neural networks to adversarial examples has recently received great attention [228, 227, 224]. While different algorithms have been proposed to improve robustness [227, 224], studies quantifying the limits of robustness have been limited [228, 205]. We argue that the non-stochastic framework introduced in this paper can be a viable way to quantify robustness, and can be used as a baseline to evaluate the performance of different algorithmic solutions.

We follow the framework of [228] and consider a neural network trained to classify the incoming data among L possible labels in the set $\mathscr{L} = \{1, 2, \ldots, L\}$. Let $x_0 \in \mathbb{R}^d$ denote an input data point consisting of a feature vector of d dimensions. A neural network can be modelled using a classification function $f : \mathbb{R}^d \to \mathbb{R}^L$ whose ℓth component $f_\ell(\cdot)$ indicates the belief that a given data point is of label ℓ. The network classifies each input data point x_0 as being of label

$$c(x_0) = \text{argmax}_{\ell \in \{1,\ldots,L\}} f_\ell(x_0). \tag{9.293}$$

We let $\mathcal{D}(\ell) \subseteq \mathcal{D}$ denote the set of points in a data set $\mathcal{D}$ that are classified as being of label ℓ, namely

$$\mathcal{D}(\ell) = \{x_0 \in \mathcal{D} : c(x_0) = \ell\}. \tag{9.294}$$

When the points in this set are subject to an ϵ-attack, they become part of the perturbed input data set

$$S_\epsilon(\ell) = \{x \in \mathbb{R}^d : \ \|x - x_0\| \leq \epsilon, x_0 \in \mathcal{D}(\ell)\}. \tag{9.295}$$

An ϵ-attack on input point x_0 is successful if there exists a noise vector $e \in \mathbb{R}^d$ such that

$$c(x_0) \neq c(x_0 + e), \text{ and } \|e\| \leq \epsilon. \tag{9.296}$$

For any two labels $\ell_1, \ell_2 \in \{1, 2, \ldots, L\}$, we let

$$\mathcal{P}_\epsilon(\ell_1, \ell_2) = \{x \in S_\epsilon(\ell_1) : \text{ either } c(x) = \ell_1 \text{ or } c(x) = \ell_2\}. \tag{9.297}$$

Then, $\mathcal{P}_\epsilon(\ell_1, \ell_2) \cap \mathcal{P}_\epsilon(\ell_2, \ell_1)$ represents the set of points in $\mathcal{S}_\epsilon(\ell_1)$ and $\mathcal{S}_\epsilon(\ell_2)$ that can lead to a successful ϵ-attack.

For all $\ell_1, \ell_2 \in \{1, 2, \ldots, L\}$, we let the error

$$e_\epsilon(\ell_1, \ell_2) = \frac{m_{\mathscr{L}}(\mathcal{P}_\epsilon(\ell_1, \ell_2) \cap \mathcal{P}_\epsilon(\ell_2, \ell_1))}{m_{\mathscr{L}}(\mathscr{L})}. \tag{9.298}$$

Finally, we say that a subset of labels (viz. a codebook) $\mathcal{L} \subseteq \mathscr{L}$ is (ϵ, δ)-robust if for all $\ell_1, \ell_2 \in \mathcal{L}$, we have $e_\epsilon(\ell_1, \ell_2) \leq \delta/|\mathcal{L}|$. This implies that whenever a label in an (ϵ, δ)-robust codebook is assigned to any input point that is subject to an attack, this is the same as the label assigned to the same input in the absence of the attack, with confidence at least $1 - \delta$.

We can then define the (ϵ, δ)- robust capacity of the neural network as the logarithm of the maximum number of labels that can be robustly classified with confidence $1 - \delta$ in the presence of an ϵ-attack, namely

$$C_\epsilon^\delta(\mathcal{D}) = \sup_{\mathcal{L} \in \mathscr{L}_\epsilon^\delta(\mathcal{D})} \log_2(|\mathcal{L}|), \tag{9.299}$$

where $\mathscr{L}_\epsilon^\delta(\mathcal{D}) = \{\mathcal{L} : \mathcal{L} \text{ is } (\epsilon, \delta)\text{-robust}\}$. This capacity represents the largest amount of information that the labeling task of the neural network can convey, at a given level of confidence, under a perturbation attack. This information is independent of whether the neural network classifies the input data correctly or not.

For $\epsilon = 0$, for any two labels $\ell_1, \ell_2 \in \mathscr{L}$, we have that $\mathcal{D}(\ell_1) = S_\epsilon(\ell_1)$ and $\mathcal{P}_\epsilon(\ell_1, \ell_2) = \mathcal{D}(\ell_1)$, which implies that

$$\mathcal{P}_\epsilon(\ell_1, \ell_2) \cap \mathcal{P}_\epsilon(\ell_2, \ell_1) = \emptyset, \tag{9.300}$$

and the capacity $C_\epsilon^\delta(\mathcal{D}) = \log_2(L)$, regardless of the value of δ. This means that in the absence an attack, the amount of information conveyed by the network is simply the logarithm of the number of labels it classifies the data into.

The framework described above has been studied in the special case of $\delta = 0$ and for a single input data point x_0 in [228]. Our (ϵ, δ)- robust capacity generalizes the notion of robustness from a single point x_0 to the whole data set $\mathcal{D}$ and can quantify the overall robustness of a neural network.

9.9.3 Performance of Classification Systems

The non-stochastic (N, δ) capacity can also be used as a performance measure of classification systems operating on a given data set. Consider a system trained to classify the incoming data among L possible labels in the set $\mathscr{L} = \{1, 2, \ldots, L\}$. Let x_0 be an input data point, and $c(x_0)$ be the label assigned by the neural network to x_0. For a given data set $\mathcal{D}$, let $N(\ell) \subseteq \{1, 2, \ldots, L\}$ denote the subset of labels such that

$$\begin{aligned} N(\ell) = \{\ell' \in \mathscr{L} :\ & \text{there exists a data point } x_0 \in \mathcal{D} \text{ such that} \\ & \text{the correct label of } x_0 \text{ is } \ell \text{ and } c(x_0) = \ell'\}. \end{aligned} \tag{9.301}$$

If $\ell \in N(\ell)$, then there exists a data point that is correctly classified as ℓ. If $\ell' \in N(\ell)$ such that $\ell' \neq \ell$, then there exists a data point that is incorrectly classified as ℓ', and the correct label of this data point is ℓ.

For any two labels $\ell_1, \ell_2 \in \mathscr{L}$, we have the equivocation region $N(\ell_1) \cap N(\ell_2)$, and we let the error

$$e_N(\ell_1, \ell_2) = \frac{|N(\ell_1) \cap N(\ell_2)|}{|\mathscr{L}|}. \tag{9.302}$$

We say that a subset of labels $\mathcal{L} \subseteq \mathscr{L}$ is (N, δ)-classifiable if for all $\ell_1, \ell_2 \in \mathcal{L}$, we have $e_N(\ell_1, \ell_2) \leq \delta/|\mathcal{L}|$, and the (N, δ)-capacity of the classification system is

$$C_N^\delta = \sup_{\mathcal{L} \in \mathscr{L}_N^\delta} \log_2(|\mathcal{L}|), \tag{9.303}$$

where $\mathscr{L}_N^\delta = \{\mathcal{L} : \mathcal{L} \text{ is } (N, \delta)\text{-classifiable}\}$. Given the set of labels, this capacity quantifies the amount of information that the classifier is able to extract from a given data set, in terms of the logarithm of the largest number of labels that can be identified with confidence greater than $1 - \delta$. In contrast to the robust capacity described in Section 9.9.2, here the capacity refers to the ability of the network to perform the classification task *correctly* in the absence of an attack, rather than to its ability of performing classification *consistently* (but not necessarily correctly) in the presence of an attack.

9.10 Conclusion

In this paper, we presented a non-stochastic theory of information that is based on a notion of information with worst-case confidence that is independent of stochastic modeling assumptions. Using the non-stochastic variables framework of Nair [157], we showed that the capacity of several channel models equals the largest amount of information conveyed by the transmitter to the receiver, with a given level of confidence. These results are the natural generalization of Nair's ones, obtained in a zero-error framework, and provide an information-theoretic interpretation for the geometric problem of sphere packing with overlap, studied by Lim and Franceschetti [132]. More generally, they show that the path laid by Shannon can be extended to a non-stochastic setting, which is an idea that dates back to Kolmogorov [119].

Non-stochastic approaches to information, and their usage to quantify the performance of engineering systems have recently received attention in the context of estimation, control, security, communication over non-linear optical channels, and learning systems [188, 189, 230, 28, 228, 218, 65]. We hope that the theory developed here can be useful in some of these contexts.

To this end, we pointed out some possible applications in the context of classification systems and communication over adversarial channels.

While refinements and extension of the theory are certainly of interest, further exploration of application domains is of paramount importance. There is evidence in the literature for the need of a non-stochastic approach to study the flow of information in complex systems, and there is a certain tradition in computer science and especially in the field of online learning to study various problems in both a stochastic and a non-stochastic setting [17, 3, 181, 175, 174]. Nevertheless, it seems that only a few isolated efforts have been made towards the formal development of a non-stochastic information theory. A wider involvement of the community in developing alternative, even competing, theories is certainly advisable to eventually fulfill the need of several application areas.

9.11 Acknowledgement

Chapter 9, in part, is a reprint of the material as it appears in Anshuka Rangi and Massimo Franceschetti, "Non-stochastic Information Theory", *under preparation*, Anshuka Rangi and Massimo Franceschetti, "Towards a Non-Stochastic Information Theory", *IEEE International Symposium on Information Theory (ISIT)*, July 2019, and Anshuka Rangi and Massimo Franceschetti, "Channel Coding Theorems in Non-stochastic Information Theory", *IEEE International Symposium on Information Theory (ISIT)*, July 2021. The dissertation author was the primary investigator and author of this paper.

9.12 Appendix

9.12.1 Proof of Lemma 31

Proof. Let $(X, Y) \overset{a}{\leftrightarrow} (\delta_1, \delta_2)$. Then,

$$\mathscr{A}(X;Y) \preceq \delta_1, \tag{9.304}$$

$$\mathscr{A}(Y;X) \preceq \delta_2. \tag{9.305}$$

Let

$$\mathscr{S}_1 = \left\{(y_1, y_2) : \frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!]|)}{m_{\mathscr{X}}([\![X]\!])} = 0\right\}. \tag{9.306}$$

Then, for all $(y_1, y_2) \in \mathscr{S}_1$, we have

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!]|)}{m_{\mathscr{X}}([\![X]\!])} = 0 \leq \delta_1. \tag{9.307}$$

Also, if $(y_1, y_2) \in \mathscr{S}_1$, then

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!]|)}{m_{\mathscr{X}}([\![X]\!])} \notin \mathscr{A}(X;Y), \tag{9.308}$$

and if $(y_1, y_2) \notin \mathscr{S}_1$, then using (9.8), we have

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!]|)}{m_{\mathscr{X}}([\![X]\!])} \in \mathscr{A}(X;Y). \tag{9.309}$$

This along with (9.304) and (9.307) implies that (9.16) follows.

Likewise, let

$$\mathscr{S}_2 = \left\{(x_1, x_2) : \frac{m_{\mathscr{Y}}([\![Y|x_1]\!] \cap [\![Y|x_2]\!]|)}{m_{\mathscr{Y}}([\![Y]\!])} = 0\right\}. \tag{9.310}$$

Then, for all $(x_1, x_2) \in \mathscr{S}_2$,

$$\frac{m_{\mathscr{Y}}([\![Y|x_1]\!] \cap [\![Y|x_2]\!]|)}{m_{\mathscr{Y}}([\![Y]\!])} = 0 \leq \delta_2. \tag{9.311}$$

Also, if $(x_1, x_2) \in \mathscr{S}_2$, then

$$\frac{m_{\mathscr{Y}}([\![Y|x_1]\!] \cap [\![Y|x_2]\!]|)}{m_{\mathscr{Y}}([\![Y]\!])} \notin \mathscr{A}(Y;X), \tag{9.312}$$

and if $(y_1, y_2) \notin \mathscr{S}_2$, then using (9.8), we have

$$\frac{m_{\mathscr{Y}}([\![Y|x_1]\!] \cap [\![Y|x_2]\!]|)}{m_{\mathscr{Y}}([\![Y]\!])} \in \mathscr{A}(Y;X). \tag{9.313}$$

This along with (9.305) and (9.311) implies that (9.17) follows.

Now, we prove the opposite direction of the statement. Given that for all $y_1, y_2 \in [\![Y]\!]$, we have

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!]|)}{m_{\mathscr{X}}([\![X]\!])} \leq \delta_1, \tag{9.314}$$

and for all $x_1, x_2 \in [\![X]\!]$, we have

$$\frac{m_{\mathscr{Y}}([\![Y|x_1]\!] \cap [\![Y|x_2]\!]|)}{m_{\mathscr{Y}}([\![Y]\!])} \leq \delta_2. \tag{9.315}$$

Then, using the definition of $\mathscr{A}(X;Y)$ and $\mathscr{A}(Y;X)$, we have

$$\mathscr{A}(X;Y) \preceq \delta_1, \tag{9.316}$$

$$\mathscr{A}(Y;X) \preceq \delta_2. \tag{9.317}$$

The statement of the lemma follows. □

9.12.2 Proof of Theorem 45

Proof. We will show (9.132). Then, using Lemma 36 in Appendix 9.12.4, (9.133) follows using the same argument as in the proof of Theorem 44.

We proceed in three steps. First, we show that for all $n > 0$, there exists an UV $X(1:n)$ and $\tilde{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![Y(1:n)]\!])$ such that $X(1:n) \in \mathscr{F}_{\tilde{\delta}}(n)$, which implies $\mathscr{F}_{\tilde{\delta}}(n)$ is not empty, so that the supremum is well defined. Second, for all $n > 0$, and $X(1:n)$ and $\tilde{\delta}$ such that

$$X(1:n) \in \mathscr{F}_{\tilde{\delta}}(n), \tag{9.318}$$

and

$$\tilde{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![Y(1:n)]\!]), \tag{9.319}$$

we show that

$$\frac{I_{\tilde{\delta}/|[\![X(1:n)]\!]|}(Y(1:n); X(1:n))}{n} \leq R_{\delta_n}.$$

Finally, for all $n > 0$, we show the existence of $X(1:n) \in \mathscr{F}_{\tilde{\delta}}(n)$ and $\tilde{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![Y(1:n)]\!])$ such that

$$\frac{I_{\tilde{\delta}/|[\![X(1:n)]\!]|}(Y(1:n); X(1:n))}{n} = R_{\delta_n}. \tag{9.320}$$

Let us begin with the first step. Consider a point $x(1:n) \in \mathscr{X}^n$. Let $X(1:n)$ be a UV such that

$$[\![X(1:n)]\!] = \{x(1:n)\}. \tag{9.321}$$

Then, we have that the marginal range of the UV $Y(1:n)$ corresponding to the received variable is

$$[\![Y(1:n)]\!] = [\![Y(1:n)|x(1:n)]\!], \tag{9.322}$$

and therefore for all $y(1:n) \in [\![Y(1:n)]\!]$, we have

$$[\![X(1:n)|y(1:n)]\!] = \{x(1:n)\}. \tag{9.323}$$

Using Definition 10 and (9.321), we have that

$$\mathscr{A}(Y(1:n);X(1:n))=\emptyset, \tag{9.324}$$

because $[\![X(1:n)]\!]$ consists of a single point, and therefore the set in (9.11) is empty.

On the other hand, using Definition 10 and (9.323), we have

$$\mathscr{A}(X(1:n);Y(1:n))=\begin{cases}\{1\} \text{ if } \exists y_1(1:n), y_2(1:n)\in[\![Y(1:n)]\!],\\ \emptyset \quad \text{otherwise.}\end{cases} \tag{9.325}$$

Using (9.324) and since $\mathscr{A}\preceq\delta$ holds for $\mathscr{A}=\emptyset$, we have

$$\mathscr{A}(Y(1:n);X(1:n))\preceq\delta_n/(|[\![X(1:n)]\!]|m_{\mathscr{Y}}([\![Y(1:n)]\!])). \tag{9.326}$$

Similarly, using (9.325) we have

$$\mathscr{A}(X(1:n);Y(1:n))\preceq 1. \tag{9.327}$$

Now, combining (9.326) and (9.327), we have

$$(X(1:n),Y(1:n))\overset{a}{\leftrightarrow}(1,\delta_n/(|[\![X(1:n)]\!]|m_{\mathscr{Y}}([\![Y(1:n)]\!]))). \tag{9.328}$$

Letting $\tilde{\delta}=\delta_n/m_{\mathscr{Y}}([\![Y(1:n)]\!])$, this implies that $X(1:n)\in\mathscr{F}_{\tilde{\delta}}(n)$ and the first step of the proof is complete.

To prove the second step, we define

$$\mathscr{G}(n)=\Big\{X(1:n):[\![X(1:n)]\!]\subseteq\mathscr{X}^n,\exists\tilde{\delta}\leq\delta_n/m_{\mathscr{Y}}([\![Y(1:n)]\!]) \text{ such that}$$
$$\forall\mathscr{S}_1,\mathscr{S}_2\in[\![Y(1:n)|X(1:n)]\!],\frac{m_{\mathscr{Y}}(\mathscr{S}_1\cap\mathscr{S}_2)}{m_{\mathscr{Y}}([\![Y(1:n)]\!])}\leq\frac{\tilde{\delta}}{|[\![X(1:n)]\!]|}\Big\}, \tag{9.329}$$

which is a larger set than the one containing all UVs $X(1:n)$ that are $(1, \tilde{\delta}/|[\![X(1:n)]\!]|)$ associated to $Y(1:n)$. Similar to (9.76), it can be shown that

$$X(1:n) \in \mathscr{G}(n) \implies \mathcal{X}(1:n) \in \mathscr{X}_N^{\delta_n}(n) \tag{9.330}$$

Consider now a pair $X(1:n)$ and $\tilde{\delta}$ such that $\tilde{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![Y(1:n)]\!])$, and

$$X(1:n) \in \mathscr{F}_{\tilde{\delta}}(n). \tag{9.331}$$

If $(X(1:n), Y(1:n)) \overset{d}{\leftrightarrow} (0, \tilde{\delta}/|[\![X(1:n)]\!]|)$, then using Lemma 33 in Appendix 9.12.4, there exist UVs $\bar{X}(1:n)$ and $\bar{Y}(1:n)$ and $\bar{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![\bar{Y}(1:n)]\!])$ such that

$$(\bar{X}(1:n), \bar{Y}(1:n)) \overset{a}{\leftrightarrow} (1, \bar{\delta}/|[\![\bar{X}(1:n)]\!]|), \tag{9.332}$$

and

$$|[\![Y(1:n)|X(1:n)]\!]^*_{\tilde{\delta}/|[\![X(1:n)]\!]|}| = |[\![\bar{Y}(1:n)|\bar{X}(1:n)]\!]^*_{\bar{\delta}/|[\![\bar{X}(1:n)]\!]|}|. \tag{9.333}$$

On the other hand, if $(X(1:n), Y(1:n)) \overset{a}{\leftrightarrow} (1, \tilde{\delta}/|[\![X(1:n)]\!]|)$, then (9.332) and (9.333) also trivially hold. It then follows that (9.332) and (9.333) hold for all $X(1:n) \in \mathscr{F}_{\tilde{\delta}}(n)$. We now have

$$\begin{aligned}
I_{\tilde{\delta}/|[\![X(1:n)]\!]|}(Y(1:n); X(1:n)) &= \log(|[\![Y(1:n)|X(1:n)]\!]^*_{\tilde{\delta}/|[\![X(1:n)]\!]|}|) \\
&\overset{(a)}{=} \log(|[\![\bar{Y}(1:n)|\bar{X}(1:n)]\!]^*_{\bar{\delta}/|[\![\bar{X}(1:n)]\!]|}|) \\
&\overset{(b)}{\leq} \log(|[\![\bar{X}(1:n)]\!]|) \\
&\overset{(c)}{=} \log(|\bar{\mathcal{X}}(1:n)|) \\
&\overset{(d)}{\leq} nR_{\delta_n},
\end{aligned} \tag{9.334}$$

where (a) follows from (9.332) and (9.333), (b) follows from Lemma 35 in Appendix 9.12.4 since $\bar{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![\bar{Y}(1:n)]\!]) < m_{\mathscr{Y}}(V_N^n)/m_{\mathscr{Y}}([\![\bar{Y}(1:n)]\!])$, (c) follows by defining the codebook $\bar{\mathcal{X}}(1:n)$ corresponding to the UV $[\![\bar{X}(1:n)]\!]$, and (d) follows from the fact that using (9.332) and Lemma 31, we have $\bar{X}(1:n) \in \mathcal{G}(n)$, which implies by (9.330) that $\bar{\mathcal{X}}(1:n) \in \mathscr{X}_N^{\delta_n}(n)$.

For any $n \in \mathbb{Z}_{>0}$, let

$$\mathcal{X}_n^* = \text{argsup}_{\mathcal{X}_n \in \mathscr{X}_N^{\delta_n}(n)} \frac{\log(|\mathcal{X}_n|)}{n}, \tag{9.335}$$

which achieves the rate R_{δ_n}. Let X^* be the UV whose marginal range corresponds to the codebook $\mathcal{X}_n^*$. It follows that for all $\mathscr{S}_1, \mathscr{S}_1 \in [\![Y^*|X^*]\!]$, we have

$$\frac{m_{\mathscr{Y}}(\mathscr{S}_1 \cap \mathscr{S}_1)}{m_{\mathscr{Y}}(\mathscr{Y}^n)} \leq \frac{\delta_n}{|[\![X^*]\!]|}, \tag{9.336}$$

which implies using the fact that $m_{\mathscr{Y}}(\mathscr{Y}^n) = 1$,

$$\frac{m_{\mathscr{Y}}(\mathscr{S}_1 \cap \mathscr{S}_1)}{m_{\mathscr{Y}}([\![Y^*]\!])} \leq \frac{\delta_n}{(|[\![X^*]\!]| m_{\mathscr{Y}}([\![Y^*]\!]))}. \tag{9.337}$$

Letting $\delta^* = \delta_n / m_{\mathscr{Y}}([\![Y^*]\!])$, and using Lemma 31, we have that $(X^*, Y^*) \overset{a}{\leftrightarrow} (1, \delta^*/|[\![X^*]\!]|)$, which implies

$$X^* \in \cup_{\tilde{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![Y^*]\!])} \mathscr{F}_{\tilde{\delta}}(n), \tag{9.338}$$

and (9.132) follows. □

9.12.3 Proof of Lemma 32

Proof. Let us begin with part $1)$. We have

$$\begin{aligned}
[\![Y(1:n)]\!] &= \cup_{x(1:n)\in[\![X(1:n)]\!]}[\![Y(1:n)|x(1:n)]\!] \\
&\overset{(a)}{=} \cup_{x(1:n)\in[\![X(1:n)]\!]}[\![Y(1)|x(1)]\!]\times\ldots\times[\![Y(n)|x(n)]\!] \\
&\overset{(b)}{=} \cup_{x(1)\in[\![X(1)]\!]}[\![Y(1)|x(1)]\!]\times\ldots\times\cup_{x(n)\in[\![X(n)]\!]}[\![Y(n)|x(n)]\!] \\
&= [\![Y(1)]\!]\times[\![Y(2)]\!]\times\ldots\times[\![Y(n)]\!],
\end{aligned} \tag{9.339}$$

where (a) follows from (9.142), and (b) follows from (9.141). Now, we have

$$\begin{aligned}
\bigcup_{\mathscr{S}\in\prod_{i=1}^{n}[\![Y(i)|X(i)]\!]_\delta^*}\mathscr{S} &\overset{(a)}{=} \prod_{i=1}^{n}\Big(\bigcup_{\mathscr{S}\in[\![Y(i)|X(i)]\!]_\delta^*}\mathscr{S}\Big) \\
&\overset{(b)}{=} \prod_{i=1}^{n}[\![Y(i)]\!] \\
&\overset{(c)}{=} [\![Y(1:n)]\!],
\end{aligned} \tag{9.340}$$

where (a) follows from the fact that the cartesian product is distributive over union, namely

$$\cup_{(i,j)\in I\times J}A_i\times B_j = (\cup_{i\in I}A_i)\times(\cup_{j\in J}B_j), \tag{9.341}$$

(b) follows from the fact that for all $1\leq i\leq n$, $[\![Y(i)|X(i)]\!]_\delta^*$ is a covering of $[\![Y(i)]\!]$ by Definition 14, and (c) follows from (9.339). Hence, part $1)$ follows.

Now, we prove part $2)$. Here, we will first show that for all $\mathscr{S}\in\prod_{i=1}^{n}[\![Y(i)|X(i)]\!]_\delta^*$, we have that $\mathscr{S}$ is δ^n-connected. Second, we will show that $\mathscr{S}$ contains at least one singly δ^n-connected set.

Let us begin with the first step of part $2)$. Consider a set $\mathscr{S}\in\prod_{i=1}^{n}[\![Y(i)|X(i)]\!]_\delta^*$. Then,

there exists a sequence $\{\mathscr{S}_i\}_{i=1}^n$ such that

$$\mathscr{S} = \mathscr{S}_1 \times \mathscr{S}_2 \times \ldots \times \mathscr{S}_n, \tag{9.342}$$

and for all $1 \leq i \leq n$,

$$\mathscr{S}_i \in [\![Y(i)|X(i)]\!]_\delta^*. \tag{9.343}$$

Now, consider two points $y_1(1:n), y_2(1:n) \in \mathscr{S}$. Then, using (9.142), (9.339) and (9.342), for all $1 \leq i \leq n$, we have that

$$y_1(i), y_2(i) \in \mathscr{S}_i. \tag{9.344}$$

Also, since $\mathscr{S}_i$ is δ-connected using (9.343) and Property 1 of Definition 14, we have

$$y_1(i) \overset{\delta}{\leftrightsquigarrow} y_2(i), \tag{9.345}$$

namely there exists a sequence $\{[\![Y(i)|x_k(i)]\!]\}_{k=1}^{N(i)}$ such that

$$y_1(i) \in [\![Y(i)|x_1(i)]\!], y_2(i) \in [\![Y(i)|x_{N(i)}(i)]\!], \tag{9.346}$$

and for all $1 \leq k < N(i)$,

$$\frac{m_{\mathscr{Y}}([\![Y(i)|x_k(i)]\!] \cap [\![Y(i)|x_{k+1}(i)]\!])}{m_{\mathscr{Y}}([\![Y(i)]\!])} > \delta. \tag{9.347}$$

Without loss of generality, let

$$N(1) \leq N(2) \leq \ldots \leq N(n). \tag{9.348}$$

Now, consider the following sequence of conditional ranges

$$\begin{aligned}
&[\![Y(1:n)|x_1(1), x_1(2)\dots x_1(n)]\!],\\
&[\![Y(1:n)|x_2(1), x_2(2)\dots x_2(n)]\!],\\
&\dots\\
&[\![Y(1:n)|x_{N(1)}(1), x_{N(1)}(2)\dots x_{N(1)}(n)]\!],\\
&[\![Y(1:n)|x_{N(1)}(1), x_{N(1)+1}(2)\dots x_{N(1)+1}(n)]\!],\\
&\dots\\
&[\![Y(1:n)|x_{N(1)}(1), x_{N(2)}(2)\dots x_{N(n)}(n)]\!].
\end{aligned} \tag{9.349}$$

In this sequence, for all $1 \leq k < N(n)$, if $x_k(i) = x_{k+1}(i)$, then we have

$$\begin{aligned}
\frac{m_{\mathscr{Y}}([\![Y(i)|x_k(i)]\!] \cap [\![Y(i)|x_{k+1}(i)]\!])}{m_{\mathscr{Y}}([\![Y(i)]\!])} &\overset{(a)}{=} \frac{m_{\mathscr{Y}}([\![Y(i)|x_k(i)]\!])}{m_{\mathscr{Y}}([\![Y(i)]\!])}\\
&\overset{(b)}{>} \frac{\delta}{m_{\mathscr{Y}}([\![Y(i)]\!])}\\
&\overset{(c)}{\geq} \frac{\delta}{m_{\mathscr{Y}}(\mathscr{Y})}\\
&\overset{(d)}{>} \delta,
\end{aligned} \tag{9.350}$$

where (a) follows from the fact that $x_k(i) = x_{k+1}(i)$, (b) follows from (9.143), (c) follows from the fact that $[\![Y(i)]\!] \subseteq \mathscr{Y}$ and (9.9) holds, and (d) follows from the fact that $m_{\mathscr{Y}}(\mathscr{Y}) = 1$. Additionally, in the sequence (9.349), for all $1 \leq k < N(n)$, if $x_k(i) \neq x_{k+1}(i)$, then we have that (9.347) holds. This along with (9.350) implies that for all $1 \leq k < N(n)$, we have

$$\frac{m_{\mathscr{Y}}([\![Y(i)|x_k(i)]\!] \cap [\![Y(i)|x_{k+1}(i)]\!])}{m_{\mathscr{Y}}([\![Y(i)]\!])} > \delta. \tag{9.351}$$

Now, using (9.142) and (9.346), we have

$$y_1(1:n) \in [\![Y(1:n)|x_1(1), \ldots x_1(n)]\!], \tag{9.352}$$

and

$$y_2(1:n) \in [\![Y(1:n)|x_{N(1)}(1), \ldots x_{N(n)}(n)]\!]. \tag{9.353}$$

Also, for all $1 \leq k < N(n)$, the uncertainty associated with the intersection of the two consecutive conditional ranges in the sequence (9.349) is

$$\begin{aligned} &\frac{m_{\mathscr{Y}}([\![Y(1:n)|x_k(1:n)]\!] \cap [\![Y(1:n)|x_{k+1}(1:n)]\!])}{m_{\mathscr{Y}}([\![Y(1:n)]\!])} \\ &\overset{(a)}{=} \prod_{i=1}^{n} \frac{m_{\mathscr{Y}}([\![Y(i)|x_k(i)]\!] \cap [\![Y(i)|x_{k+1}(i)]\!])}{m_{\mathscr{Y}}([\![Y(i)]\!])} \\ &\overset{(b)}{>} \delta^n, \end{aligned} \tag{9.354}$$

where (a) follows from Assumption 3, (9.142), (9.339) and the fact that

$$\prod_{i=1}^{n} \mathscr{S}_i \cap \prod_{i=1}^{n} \mathscr{T}_i = (\mathscr{S}_1 \cap \mathscr{T}_1) \times \ldots \times (\mathscr{S}_n \cap \mathscr{T}_n), \tag{9.355}$$

(b) follows from (9.351). Hence, using (9.352), (9.353) and (9.354), we have

$$y_1(1:n) \overset{\delta^n}{\leftrightsquigarrow} y_2(1:n). \tag{9.356}$$

Hence, $\mathscr{S}$ is δ^n-connected.

Now, let us prove the second step of part 2). For all $1 \leq i \leq n$, since $[\![Y(i)|X(i)]\!]^*_\delta$ satisfies Property 1 of Definition 14, there exists an $x_i \in [\![X(i)]\!]$ such that

$$[\![Y(i)|x_i]\!] \subseteq \mathscr{S}_i. \tag{9.357}$$

Therefore, for $x(1:n) = [x_1, x_2, \ldots, x_n]$, we have

$$\begin{aligned}
[\![Y(1:n)|x(1:n)]\!] &\overset{(a)}{=} [\![Y(1)|x(1)]\!] \times \ldots \times [\![Y(n)|x(n)]\!] \\
&\overset{(b)}{=} [\![Y(1)|x_1]\!] \times \ldots \times [\![Y(n)|x_n]\!] \\
&\overset{(c)}{\subseteq} \mathscr{S}_1 \times \mathscr{S}_2 \times \ldots \mathscr{S}_n \\
&\overset{(d)}{=} \mathscr{S}, \qquad (9.358)
\end{aligned}$$

where (a) follows from (9.142), (b) follows from the fact that $x(1:n) = [x_1, x_2, \ldots, x_n]$, (c) follows from (9.357), and (d) follows from (9.342). Hence, $\mathscr{S}$ contains at least one singly δ^n-connected set, which concludes the second step of part $2)$.

Now, let us prove part $3)$. For all $1 \leq i \leq n$, since $[\![Y(i)|X(i)]\!]_\delta^*$ satisfies Property 3 of Definition 14, for all $x(i) \in [\![X(i)]\!]$, there exist a set $\mathscr{S}(x(i)) \in [\![Y(i)|X(i)]\!]_\delta^*$ such that

$$[\![Y(i)|x(i)]\!] \subseteq \mathscr{S}(x(i)). \qquad (9.359)$$

Then for all $x(1:n) \in [\![X(1:n)]\!]$, we have

$$\begin{aligned}
[\![Y(1:n)|x(1:n)]\!] &\overset{(a)}{=} [\![Y(1)|x(1)]\!] \times \ldots \times [\![Y(n)|x(n)]\!] \\
&\overset{(b)}{\subseteq} \mathscr{S}(x(1)) \times \ldots \times \mathscr{S}(x(n)) \\
&\in \prod_{i=1}^{n} [\![Y(i)|X(i)]\!]_\delta^*, \qquad (9.360)
\end{aligned}$$

where (a) follows from (9.142), and (b) follows from (9.359). Hence, part $3)$ follows.

Finally, let us prove part $4)$. Consider two distinct sets $\mathscr{S}_1, \mathscr{S}_2 \in \prod_{i=1}^{n} [\![Y(i)|X(i)]\!]_\delta^*$. Then, we have

$$\mathscr{S}_1 = \mathscr{S}_{11} \times \mathscr{S}_{12} \times \ldots \mathscr{S}_{1n}, \qquad (9.361)$$

$$\mathscr{S}_2 = \mathscr{S}_{21} \times \mathscr{S}_{22} \times \ldots \mathscr{S}_{2n}, \qquad (9.362)$$

where for all $1 \leq i \leq n$

$$\mathscr{S}_{1i}, \mathscr{S}_{2i} \in [\![Y(i)|X(i)]\!]^*_\delta. \tag{9.363}$$

Since $\mathscr{S}_1 \neq \mathscr{S}_2$, there exists $1 \leq i^* \leq n$ such that

$$\mathscr{S}_{1i^*} \neq \mathscr{S}_{2i^*}. \tag{9.364}$$

Then, by Property 2 of Definition 14 and (9.363), we have

$$\frac{m_{\mathscr{Y}}(\mathscr{S}_{1i^*} \cap \mathscr{S}_{2i^*})}{m_{\mathscr{Y}}([\![Y(i^*)]\!])} \leq \delta. \tag{9.365}$$

Also, using (9.147), we have that for all $1 \leq i \leq n$,

$$\frac{m_{\mathscr{Y}}(\mathscr{S}_{1i})}{m_{\mathscr{Y}}([\![Y(i)]\!])} \leq \hat{\delta}(n). \tag{9.366}$$

Then, we have

$$\begin{aligned}
\frac{m_{\mathscr{Y}}(\mathscr{S}_1 \cap \mathscr{S}_2)}{m_{\mathscr{Y}}([\![Y(1:n)]\!])} &\overset{(a)}{=} \frac{m_{\mathscr{Y}}((\mathscr{S}_{11} \times \ldots \mathscr{S}_{1n}) \cap (\mathscr{S}_{21} \times \ldots \mathscr{S}_{2n}))}{m_{\mathscr{Y}}([\![Y(1:n)]\!])} \\
&\overset{(b)}{=} \frac{m_{\mathscr{Y}}((\mathscr{S}_{11} \cap \mathscr{S}_{21}) \times \ldots \times (\mathscr{S}_{1n} \cap \mathscr{S}_{2n}))}{m_{\mathscr{Y}}([\![Y(1:n)]\!])} \\
&\overset{(c)}{=} \frac{m_{\mathscr{Y}}(\mathscr{S}_{11} \cap \mathscr{S}_{21}) \ldots m_{\mathscr{Y}}(\mathscr{S}_{1n} \cap \mathscr{S}_{2n})}{\prod_{i=1}^n m_{\mathscr{Y}}([\![Y(i)]\!])} \\
&\overset{(d)}{\leq} \frac{m_{\mathscr{Y}}(\mathscr{S}_{1i^*} \cap \mathscr{S}_{2i^*})}{m_{\mathscr{Y}}([\![Y(i^*)]\!])} \prod_{i \neq i^*} \frac{m_{\mathscr{Y}}(\mathscr{S}_{1i})}{m_{\mathscr{Y}}([\![Y(i)]\!])} \\
&\overset{(e)}{\leq} \delta(\hat{\delta}(n))^{n-1},
\end{aligned} \tag{9.367}$$

where (a) follows from (9.361) and (9.362), (b) follows from the fact that for all sequences of sets $\{\mathscr{S}_i\}_{i=1}^n$ and $\{\mathscr{T}_i\}_{i=1}^n$, we have

$$\prod_{i=1}^n \mathscr{S}_i \cap \prod_{i=1}^n \mathscr{T}_i = (\mathscr{S}_1 \cap \mathscr{T}_1) \times \ldots \times (\mathscr{S}_n \cap \mathscr{T}_n), \tag{9.368}$$

(c) follows from Assumption 3 and (9.339), (d) follows from (9.9) and the fact that for all $1 \leq i \leq n$, $\mathscr{S}_{1i} \cap \mathscr{S}_{2i} \subseteq \mathscr{S}_{1i}$, and (e) follows from (9.365) and (9.366). Hence, part $4)$ follows. $\square$

9.12.4 Auxiliary Results

Lemma 33. *Given a $\delta < m_{\mathscr{Y}}(V_N)$, two UVs X and Y satisfying* (9.57) *and* (9.58)*, and a $\tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])$ such that*

$$(X, Y) \overset{d}{\leftrightarrow} (0, \tilde{\delta}/|[\![X]\!]|). \tag{9.369}$$

Then, there exists two UVs $\bar{X}$ and $\bar{Y}$ satisfying (9.57) *and* (9.58)*, and there exists a $\bar{\delta} \leq \delta / m_{\mathscr{Y}}([\![\bar{Y}]\!])$ such that*

$$(\bar{X}, \bar{Y}) \overset{a}{\leftrightarrow} (1, \bar{\delta}/|[\![\bar{X}]\!]|), \tag{9.370}$$

and

$$|[\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}| = |[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}|. \tag{9.371}$$

Proof. Let the cardinality

$$|[\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}| = K. \tag{9.372}$$

By Property 1 of Definition 14, we have that for all $\mathscr{S}_i \in [\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}$, there exists a $x_i \in [\![X]\!]$ such that $[\![Y|x_i]\!] \subseteq \mathscr{S}_i$. Now, consider a new UV $\bar{X}$ whose marginal range is composed of K elements of $[\![X]\!]$, namely

$$[\![\bar{X}]\!] = \{x_1, x_2, \ldots, x_K\}. \tag{9.373}$$

Let $\bar{Y}$ be the UV corresponding to the received variable. Using the fact that for all $x \in \mathscr{X}$, we have $[\![\bar{Y}|x]\!] = [\![Y|x]\!]$ since (9.57) holds, and using Property 2 of Definition 14, for all $x, x' \in [\![\bar{X}]\!]$,

we have

$$\frac{m_{\mathscr{Y}}([\![\bar{Y}|x]\!] \cap [\![\bar{Y}|x']\!])}{m_{\mathscr{Y}}([\![Y]\!])} \leq \frac{\tilde{\delta}}{|[\![X]\!]|} \overset{(a)}{\leq} \frac{\tilde{\delta}}{|[\![\bar{X}]\!]|}, \tag{9.374}$$

where (a) follows from the fact that $[\![\bar{X}]\!] \subseteq [\![X]\!]$ using (9.373). Then, for all $x, x' \in [\![\bar{X}]\!]$, we have that

$$\frac{m_{\mathscr{Y}}([\![\bar{Y}|x]\!] \cap [\![\bar{Y}|x']\!])}{m_{\mathscr{Y}}([\![\bar{Y}]\!])} \leq \frac{\tilde{\delta} m_{\mathscr{Y}}([\![Y]\!])}{|[\![X]\!]| m_{\mathscr{Y}}([\![\bar{Y}]\!])} \overset{(a)}{\leq} \frac{\bar{\delta}}{|[\![\bar{X}]\!]|}, \tag{9.375}$$

where $\bar{\delta} = \tilde{\delta} m_{\mathscr{Y}}([\![Y]\!])/m_{\mathscr{Y}}([\![\bar{Y}]\!])$. Then, by Lemma 31 it follows that

$$(\bar{X}, \bar{Y}) \overset{a}{\leftrightarrow} (1, \bar{\delta}/|[\![\bar{X}]\!]|). \tag{9.376}$$

Since $\tilde{\delta} \leq \delta/m_{\mathscr{Y}}([\![Y]\!])$, we have

$$\bar{\delta} \leq \delta/m_{\mathscr{Y}}([\![\bar{Y}]\!]) < m_{\mathscr{Y}}(V_\epsilon)/m_{\mathscr{Y}}([\![\bar{Y}]\!]). \tag{9.377}$$

Using (9.376) and (9.377), we now have that

$$|[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}| \overset{(a)}{=} |[\![\bar{X}]\!]| \overset{(b)}{=} |[\![Y|X]\!]^*_{\bar{\delta}/|[\![X]\!]|}|, \tag{9.378}$$

where (a) follows from Lemma 36 in Appendix 9.12.4, and (b) follows from (9.372) and (9.373). Hence, the statement of the lemma follows. □

Lemma 34. *Let*

$$(X, Y) \overset{d}{\leftrightarrow} (\delta, \delta_2). \tag{9.379}$$

If $x \overset{\delta}{\leftrightsquigarrow} x_1$ *and* $x \overset{\delta}{\leftrightsquigarrow} x_2$*, then we have that* $x_1 \overset{\delta}{\leftrightsquigarrow} x_2$*.*

Proof. Let $\{[\![X|y_i]\!]\}_{i=1}^{N}$ be the sequence of conditional range connecting x and x_1. Likewise, let $\{[\![X|\tilde{y}_i]\!]\}_{i=1}^{\tilde{N}}$ be the sequence of conditional range connecting x and x_2.

Now, by Definition 13, we have

$$x_1 \in [\![X|y_N]\!], \tag{9.380}$$

$$x_2 \in [\![X|\tilde{y}_{\tilde{N}}]\!], \tag{9.381}$$

$$x \in [\![X|y_1]\!], \tag{9.382}$$

and

$$x \in [\![X|\tilde{y}_1]\!] \tag{9.383}$$

Then, using (9.8), we have that

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|\tilde{y}_1]\!]|}{m_{\mathscr{X}}([\![X]\!])} > 0, \tag{9.384}$$

which implies that

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|\tilde{y}_1]\!]|}{m_{\mathscr{X}}([\![X]\!])} \in \mathscr{A}(X;Y). \tag{9.385}$$

Using the fact that

$$(X,Y) \overset{d}{\leftrightarrow} (\delta, \delta_2), \tag{9.386}$$

we will now show that

$$\{[\![X|y_N]\!], [\![X|y_{N-1}]\!], \ldots [\![X|y_1]\!], [\![X|\tilde{y}_1]\!], \ldots [\![X|\tilde{y}_{\tilde{N}}]\!]\}, \tag{9.387}$$

is a sequence of conditional ranges connecting x_1 and x_2. Using (9.385) and (9.386), we have that

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|\tilde{y}_1]\!]|}{m_{\mathscr{X}}([\![X]\!])} > \delta. \tag{9.388}$$

Also, for all $1 < i \leq N$ and $1 < j \leq \tilde{N}$, we have

$$\frac{m_{\mathscr{X}}([\![X|y_i]\!] \cap [\![X|y_{i-1}]\!]|}{m_{\mathscr{X}}([\![X]\!])} > \delta, \tag{9.389}$$

and

$$\frac{m_{\mathscr{X}}([\![X|\tilde{y}_j]\!] \cap [\![X|\tilde{y}_{j-1}]\!]|}{m_{\mathscr{X}}([\![X]\!])} > \delta. \tag{9.390}$$

Also, we have

$$x_1 \in [\![X|y_N]\!], \text{ and } x_2 \in [\![X|\tilde{y}_{\tilde{N}}]\!]. \tag{9.391}$$

Hence, combining (9.388), (9.389), (9.390) and (9.391), we have that $x_1 \overset{\delta}{\rightsquigarrow} x_2$ via (9.387). □

Lemma 35. *Consider two UVs X and Y. Let*

$$\delta^* = \frac{\min_{y \in [\![Y]\!]} m_{\mathscr{X}}([\![X|y]\!])}{m_{\mathscr{X}}([\![X]\!])}. \tag{9.392}$$

If $\delta_1 < \delta^*$, then we have

$$|[\![X|Y]\!]^*_{\delta_1}| \leq |[\![Y]\!]|. \tag{9.393}$$

Proof. We will prove this by contradiction. Let

$$|[\![X|Y]\!]^*_{\delta_1}| > |[\![Y]\!]|. \tag{9.394}$$

Then, by Property 1 of Definition 14, there exists two sets $\mathscr{S}_1, \mathscr{S}_2 \in [\![X|Y]\!]^*_{\delta_1}$ and one singly δ_1-connected set $[\![X|y]\!]$ such that

$$[\![X|y]\!] \subseteq \mathscr{S}_1, \text{ and } [\![X|y]\!] \subseteq \mathscr{S}_2. \tag{9.395}$$

Then, we have

$$\begin{aligned}\frac{m_{\mathscr{X}}(\mathscr{S}_1 \cap \mathscr{S}_2)}{m_{\mathscr{X}}([\![X]\!])} &\overset{(a)}{\geq} \frac{m_{\mathscr{X}}([\![X|y]\!])}{m_{\mathscr{X}}([\![X]\!])} \\ &\overset{(b)}{\geq} \delta^* \\ &\overset{(c)}{>} \delta_1,\end{aligned} \tag{9.396}$$

where (a) follows from (9.395) and (9.9), (b) follows from (9.392), and (c) follows from the fact that $\delta_1 < \delta^*$. However, by Property 2 of Definition 14, we have

$$\frac{m_{\mathscr{X}}(\mathscr{S}_1 \cap \mathscr{S}_2)}{m_{\mathscr{X}}([\![X]\!])} \leq \delta_1. \tag{9.397}$$

Hence, we have that (9.396) and (9.397) contradict each other, which implies (9.394) does not hold. Hence, the statement of the theorem follows. □

Lemma 36. *Consider two UVs X and Y. Let*

$$\delta^* = \frac{\min_{y \in [\![Y]\!]} m_{\mathscr{X}}([\![X|y]\!])}{m_{\mathscr{X}}([\![X]\!])}. \tag{9.398}$$

For all $\delta_1 < \delta^*$ and $\delta_2 \leq 1$, if $(X, Y) \overset{a}{\leftrightarrow} (\delta_1, \delta_2)$, then we have

$$|[\![X|Y]\!]^*_{\delta_1}| = |[\![Y]\!]|. \tag{9.399}$$

Additionally, $[\![X|Y]\!]$ is a δ_1-overlap family.

Proof. We show that

$$[\![X|Y]\!] = \{[\![X|y]\!] : y \in [\![Y]\!]\} \tag{9.400}$$

is a δ_1-overlap family. First, note that $[\![X|Y]\!]$ is a cover of $[\![X]\!]$, since $[\![X]\!] = \cup_{y \in [\![Y]\!]} [\![X|y]\!]$. Second, each set in the family $[\![X|Y]\!]$ is singly δ_1-connected via $[\![X|Y]\!]$, since trivially any two points $x_1, x_2 \in [\![X|y]\!]$ are singly δ_1-connected via the same set. It follows that Property 1 of

Definition 14 holds.

Now, since $(X,Y) \overset{a}{\leftrightarrow} (\delta_1, \delta_2)$, then by Lemma 31 for all $y_1, y_2 \in [\![Y]\!]$ we have

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!])}{m_{\mathscr{X}}([\![X]\!])} \leq \delta_1, \tag{9.401}$$

which shows that Property 2 of Definition 14 holds. Finally, it is also easy to see that Property 3 of Definition 14 holds, since $[\![X|Y]\!]$ contains all sets $[\![X|y]\!]$. Hence, $[\![X|Y]\!]$ satisfies all the properties of δ_1-overlap family, which implies

$$|[\![X|Y]\!]| \leq |[\![X|Y]\!]^*_{\delta_1}|. \tag{9.402}$$

Since $|[\![X|Y]\!]| = |[\![Y]\!]|$, using Lemma 35, we also have

$$|[\![X|Y]\!]| \geq |[\![X|Y]\!]^*_{\delta_1}|. \tag{9.403}$$

Combining (9.402), (9.403) and the fact that $[\![X|Y]\!]$ satisfies all the properties of δ_1-overlap family, the statement of the lemma follows. □

Lemma 37. *Consider two UVs X and Y. Let*

$$\delta^* = \frac{\min_{y \in [\![Y]\!]} m_{\mathscr{X}}([\![X|y]\!])}{m_{\mathscr{X}}([\![X]\!])}. \tag{9.404}$$

For all $\delta_1 < \delta^*$ and $\delta_2 \leq 1$, if $(X,Y) \overset{a}{\leftrightarrow} (\delta_1/|[\![Y]\!]|, \delta_2)$, then under Assumption 4, for all $[\![X|y]\!] \in [\![X|Y]\!]$, there exists a point $x \in [\![X|y]\!]$ such that for all $[\![X|y']\!] \in [\![X|Y]\!] \setminus \{[\![X|y]\!]\}$,

$$x \notin [\![X|y']\!]. \tag{9.405}$$

Proof. We will prove this by contradiction. Consider a set $[\![X|y]\!]$. Let x satisfying (9.405) do

not exist. Then, for all $x' \in [\![X|y]\!]$, there exists a set $[\![X|y']\!] \in [\![X|Y]\!] \setminus \{[\![X|y]\!]\}$ such that

$$x' \in [\![X|y']\!]. \tag{9.406}$$

Thus, we have

$$\begin{aligned} m_{\mathscr{X}}(\cup_{[\![X|y']\!]\in[\![X|Y]\!]\setminus\{[\![X|y]\!]\}}([\![X|y]\!] \cap [\![X|y']\!])) &\overset{(a)}{\geq} m_{\mathscr{X}}([\![X|y]\!]) \\ &\overset{(b)}{\geq} \delta^* m_{\mathscr{X}}([\![X]\!]) \\ &\overset{(c)}{>} \delta_1 m_{\mathscr{X}}([\![X]\!]), \end{aligned} \tag{9.407}$$

where (a) follows from (9.406), (b) follows from (9.404), and (c) follows from the fact that $\delta_1 < \delta^*$. On the other hand, since $(X, Y) \overset{a}{\leftrightarrow} (\delta_1/|[\![Y]\!]|, \delta_2)$, we have

$$\begin{aligned} m_{\mathscr{X}}(\cup_{[\![X|y']\!]\in[\![X|Y]\!]\setminus\{[\![X|y]\!]\}}([\![X|y]\!] \cap [\![X|y']\!])) &\overset{(a)}{\leq} \sum_{[\![X|y']\!]\in[\![X|Y]\!]\setminus\{[\![X|y]\!]\}} m_{\mathscr{X}}([\![X|y]\!] \cap [\![X|y']\!]) \\ &\overset{(b)}{\leq} |[\![Y]\!]|\delta_1 m_{\mathscr{X}}([\![X]\!])/|[\![Y]\!]| \\ &= \delta_1 m_{\mathscr{X}}([\![X]\!]), \end{aligned} \tag{9.408}$$

where (a) follows from Assumption 4, (b) follows from Lemma 31. It follows that (9.407) and (9.408) contradict each other, and therefore x satisfying (9.405) exists. The statement of the lemma follows. □

9.12.5 Proof of 4 claims in Theorem 48

Proof of Claim 1. By Property 1 of Definition 14, we have that for all $\mathscr{S}_i \in [\![Y(1:n)|X(1:n)]\!]^*_{\delta'/|[\![X(1:n)]\!]|}$, there exists a $\tilde{x}_i(1:n) \in [\![X(1:n)]\!]$ such that $[\![Y(1:n)|\tilde{x}_i(1:n)]\!] \subseteq \mathscr{S}_i$. Now, consider a new UV $\tilde{X}(1:n)$ whose marginal range is composed of elements of $[\![X(1:n)]\!]$, namely

$$[\![\tilde{X}(1:n)]\!] = \{\tilde{x}_1(1:n), \ldots \tilde{x}_K(1:n)\}, \tag{9.409}$$

where

$$K = |[\![Y(1:n)|X(1:n)]\!]^*_{\delta'/|[\![X(1:n)]\!]|}|. \tag{9.410}$$

Let $\tilde{Y}(1:n)$ be the UV corresponding to the received variable. Then, similar to (9.88), by Property 2 of Definition 14 and since N is stationary memoryless channel, for all $x(1:n), x'(1:n) \in [\![\tilde{X}(1:n)]\!]$, we have

$$\begin{aligned}\frac{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)|x(1:n)]\!] \cap [\![\tilde{Y}(1:n)|x'(1:n)]\!])}{m_{\mathscr{Y}}([\![Y(1:n)]\!])} &\leq \frac{\delta'}{|[\![X(1:n)]\!]|} \\ &\overset{(a)}{\leq} \frac{\delta'}{|[\![\tilde{X}(1:n)]\!]|},\end{aligned} \tag{9.411}$$

where (a) follows from the fact that $[\![\tilde{X}(1:n)]\!] \subseteq [\![X(1:n)]\!]$. Similar to (9.89), for all $x(1:n), x'(1:n) \in [\![\tilde{X}(1:n)]\!]$, we have that

$$\begin{aligned}&\frac{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)|x(1:n)]\!] \cap [\![\tilde{Y}(1:n)|x'(1:n)]\!])}{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!])} \\ &\leq \frac{\delta' m_{\mathscr{Y}}([\![Y(1:n)]\!])}{|[\![\tilde{X}(1:n)]\!]| m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!])} \\ &= \frac{\tilde{\delta}}{|[\![\tilde{X}(1:n)]\!]|},\end{aligned} \tag{9.412}$$

where

$$\tilde{\delta} = \frac{\delta' m_{\mathscr{Y}}([\![Y(1:n)]\!])}{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!])}. \tag{9.413}$$

Then, by Lemma 31 it follows that

$$(\tilde{X}(1:n), \tilde{Y}(1:n)) \overset{a}{\leftrightarrow} (1, \tilde{\delta}/|[\![\tilde{X}(1:n)]\!]|). \tag{9.414}$$

Using (9.188), we also have

$$\tilde{\delta} \leq \frac{\delta_n}{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!])}. \tag{9.415}$$

Additionally, we have

$$\begin{aligned}\tilde{\delta} &\leq \frac{\delta_n}{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!])} \\ &\overset{(a)}{\leq} \left(\frac{\bar{\delta} m_{\mathscr{Y}}(V_N)}{|[\![\bar{X}]\!]|}\right)^n \frac{1}{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!])} \\ &\overset{(b)}{\leq} \frac{(\bar{\delta} m_{\mathscr{Y}}(V_N))^n}{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!])} \\ &\overset{(c)}{<} \frac{(m_{\mathscr{Y}}(V_N))^n}{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!])} \\ &\overset{(d)}{=} \frac{m_{\mathscr{Y}}(V_N^n)}{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!])},\end{aligned} \tag{9.416}$$

where (a) follows from the assumption in the theorem that

$$0 \leq \delta_n \leq (\bar{\delta} m_{\mathscr{Y}}(V_N)/|[\![\bar{X}]\!]|)^n, \tag{9.417}$$

(b) follows from the fact that $|[\![\bar{X}]\!]| \geq 1$, (c) follows from the fact that using $\delta_1 < m_{\mathscr{Y}}(V_N)$, we have

$$\bar{\delta} \leq \frac{\delta_1}{m_{\mathscr{Y}}([\![\bar{Y}]\!])} < \frac{m_{\mathscr{Y}}(V_N)}{m_{\mathscr{Y}}([\![\bar{Y}]\!])} \leq 1, \tag{9.418}$$

and (d) follows from Assumption 3. Now, we have

$$\begin{aligned}|[\![\tilde{Y}(1:n)|\tilde{X}(1:n)]\!]^*_{\tilde{\delta}/|[\![\tilde{X}(1:n)]\!]|}| &\overset{(a)}{=} |[\![\tilde{X}(1:n)]\!]| \\ &\overset{(b)}{=} |[\![Y(1:n)|X(1:n)]\!]^*_{\tilde{\delta}/|[\![X(1:n)]\!]|}|,\end{aligned} \tag{9.419}$$

where (a) follows by combining (9.414), (9.416) and Lemma 36, and (b) follows from (9.409) and (9.410). This along with (9.189) implies that we have

$$|[\![\tilde{Y}(1:n)|\tilde{X}(1:n)]\!]^*_{\tilde{\delta}/|[\![\tilde{X}(1:n)]\!]|}| > |\prod_{i=1}^{n} [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}|. \tag{9.420}$$

This concludes the proof of Claim 1. □

Proof of Claim 2. Since (9.414) and (9.416) holds, using Lemma 36, we have

$$[\![\tilde{Y}(1:n)|\tilde{X}(1:n)]\!]^*_{\tilde{\delta}/|[\![\tilde{X}(1:n)]\!]|} = [\![\tilde{Y}(1:n)|\tilde{X}(1:n)]\!]. \tag{9.421}$$

Using (9.421) and Property 1 of Definition 14, we have that for all $\mathscr{S} \in [\![\tilde{Y}(1:n)|\tilde{X}(1:n)]\!]^*_{\tilde{\delta}/|[\![\tilde{X}(1:n)]\!]|}$, there exists a $\tilde{x}(1:n) \in [\![\tilde{X}(1:n)]\!]$ such that

$$\mathscr{S} = [\![\tilde{Y}(1:n)|\tilde{x}(1:n)]\!]. \tag{9.422}$$

Now, for all $x \in \mathscr{X}$, let $\mathscr{S}(x) \in [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}$ be such that

$$[\![\bar{Y}|x]\!] \subseteq \mathscr{S}(x). \tag{9.423}$$

For all $x \in \mathscr{X} \setminus [\![\bar{X}]\!]$, the set $\mathscr{S}(x)$ exists from the assumption in the theorem. Also, for all $x \in [\![\bar{X}]\!]$, the set $\mathscr{S}(x)$ exists using Property 3 in Definition 14. Hence, for all $x \in \mathscr{X}$, we have that $\mathscr{S}(x)$ satisfying (9.423) exists.

Hence, for all $\tilde{x}(1:n) \in [\![\tilde{X}(1:n)]\!]$, we have that

$$\begin{aligned}
[\![\tilde{Y}(1:n)|\tilde{x}(1:n)]\!] &\overset{(a)}{=} [\![\tilde{Y}(1)|\tilde{x}(1)]\!] \times \ldots \times [\![\tilde{Y}(n)|\tilde{x}(n)]\!] \\
&\overset{(b)}{\subseteq} \mathscr{S}(x(1)) \times \ldots \times \mathscr{S}(x(n)) \\
&\overset{(c)}{\in} \prod_{i=1}^{n} [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|},
\end{aligned} \tag{9.424}$$

where (a) follows from the fact that N is a stationary memoryless uncertain channel, (b) follows from the fact that for all $x \in \mathscr{X}$, $\mathscr{S}(x)$ exists, and (c) follows from the fact that for all $x \in \mathscr{X}$, $\mathscr{S}(x) \in [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}$. Hence, Claim 2 is proved. □

Proof of Claim 3. Combining (9.420) and (9.421), we have that

$$|[\![\tilde{Y}(1:n)|\tilde{X}(1:n)]\!]| > \prod_{i=1}^{n} [\![\bar{Y}|\bar{X}]\!]^{*}_{\bar{\delta}/|[\![\bar{X}]\!]|}. \tag{9.425}$$

This along with (9.424) implies that there exists a set $\mathscr{S} \in \prod_{i=1}^{n} [\![\bar{Y}|\bar{X}]\!]^{*}_{\bar{\delta}/|[\![\bar{X}]\!]|}$ which contains at least two sets $\mathscr{D}_1, \mathscr{D}_2 \in [\![\tilde{Y}(1:n)|\tilde{X}(1:n)]\!]^{*}_{\bar{\delta}/|[\![\tilde{X}(1:n)]\!]|}$, namely

$$\mathscr{D}_1 \subset \mathscr{S}, \tag{9.426}$$

$$\mathscr{D}_2 \subset \mathscr{S}. \tag{9.427}$$

Using (9.422), without loss of generality, let

$$\mathscr{D}_1 = [\![\tilde{Y}(1:n)|\tilde{x}_1(1:n)]\!], \tag{9.428}$$

$$\mathscr{D}_2 = [\![\tilde{Y}(1:n)|\tilde{x}_2(1:n)]\!]. \tag{9.429}$$

Also, let

$$\mathscr{S} = \mathscr{S}_1 \times \ldots \times \mathscr{S}_n, \tag{9.430}$$

where $\mathscr{S}_1, \ldots, \mathscr{S}_n \in [\![\bar{Y}|\bar{X}]\!]^{*}_{\bar{\delta}/|[\![\bar{X}]\!]|}$. Also, we have

$$\frac{\bar{\delta}}{|[\![\bar{X}]\!]|} \leq \bar{\delta} \leq \frac{\delta_1}{m_{\mathscr{Y}}([\![\bar{Y}]\!])} < \frac{m_{\mathscr{Y}}(V_N)}{m_{\mathscr{Y}}([\![\bar{Y}]\!])}. \tag{9.431}$$

Now, we have

$$\begin{aligned}\frac{\tilde{\delta}}{|[\![\tilde{X}(1:n)]\!]|} &\leq \frac{\delta_n}{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!])} \\ &\overset{(a)}{\leq} \left(\frac{\bar{\delta} m_{\mathscr{Y}}(V_N)}{|[\![\bar{X}]\!]|}\right)^n \frac{1}{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!])} \\ &\overset{(b)}{\leq} \left(\frac{\bar{\delta}}{|[\![\bar{X}]\!]|}\right)^n, \end{aligned} \tag{9.432}$$

where (a) follows from the assumption in the theorem that

$$\delta_n \leq \left(\frac{\bar{\delta} m_{\mathscr{Y}}(V_N)}{|[\![\bar{X}]\!]|}\right)^n, \tag{9.433}$$

and (b) follows from the fact that using Assumption 3, we have

$$m_{\mathscr{Y}}(V_N^n) = (m_{\mathscr{Y}}(V_N))^n \leq m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!]). \tag{9.434}$$

Combining Lemma 31 and (9.414), we have

$$\begin{aligned}\frac{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)|\tilde{x}_1(1:n)]\!] \cap [\![\tilde{Y}(1:n)|\tilde{x}_2(1:n)]\!])}{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!])} &\leq \frac{\tilde{\delta}}{|[\![\tilde{X}(1:n)]\!]|} \\ &\overset{(a)}{\leq} \left(\frac{\bar{\delta}}{|[\![\bar{X}]\!]|}\right)^n, \end{aligned} \tag{9.435}$$

where (a) follows from (9.432). This implies that there exists a $1 \leq i^* \leq n$ such that

$$\frac{m_{\mathscr{Y}}([\![\tilde{Y}(i^*)|\tilde{x}_1(i^*)]\!] \cap [\![\tilde{Y}(i^*)|\tilde{x}_2(i^*)]\!])}{(m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!]))^{1/n}} \leq \frac{\bar{\delta}}{|[\![\bar{X}]\!]|}, \tag{9.436}$$

otherwise (9.435) does not hold, namely

$$\begin{aligned}
&\frac{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)|\tilde{x}_1(1:n)]\!] \cap [\![\tilde{Y}(1:n)|\tilde{x}_2(1:n)]\!])}{m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!])} \\
&\overset{(a)}{=} \prod_{i=1}^{n} \left(\frac{m_{\mathscr{Y}}([\![\tilde{Y}(i)|\tilde{x}_1(i)]\!] \cap [\![\tilde{Y}(i)|\tilde{x}_2(i)]\!])}{(m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!]))^{1/n}} \right), \\
&\overset{(b)}{>} \left(\frac{\bar{\delta}}{|[\![\bar{X}]\!]|} \right)^n,
\end{aligned} \tag{9.437}$$

where (a) follows from Assumption 3 and the fact that N is stationary memoryless, (b) follows from the hypothesis that i^* satisfying (9.436) does not exist. □

Proof of Claim 4. Now, consider a UV X' such that

$$[\![X']\!] = ([\![\bar{X}]\!] \setminus \{x \in \mathscr{X} : [\![Y|x]\!] \subseteq \mathscr{S}_{i^*}\}) \cup \{\tilde{x}_1(i^*)\} \cup \{\tilde{x}_2(i^*)\}. \tag{9.438}$$

For $[\![X'_1]\!] = ([\![\bar{X}]\!] \setminus \{x \in \mathscr{X} : [\![Y|x]\!] \subseteq \mathscr{S}_{i^*}\})$, the δ'_1- overlap family of $[\![Y'_1|X'_1]\!]$ satisfies

$$|[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}| - 1 \overset{(a)}{\leq} |[\![Y'_1|X'_1]\!]^*_{\delta'_1}|, \tag{9.439}$$

where

$$\delta'_1 = (\bar{\delta} m_{\mathscr{Y}}([\![\bar{Y}]\!]))/(|[\![\bar{X}]\!]| m_{\mathscr{Y}}([\![Y'_1]\!]), \tag{9.440}$$

and (a) follows from the fact that

$$\mathcal{S}_1 = \{\mathscr{S}' \in [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|} : \mathscr{S}' \neq \mathscr{S}_{i^*}\} \tag{9.441}$$

satisfies all the properties of $[\![Y'_1|X'_1]\!]^*_{\delta'_1}$ in Definition 14.

Now, consider the UV X' such that $[\![X']\!] = [\![X'_1]\!] \cup \{\tilde{x}_1(i^*)\} \cup \{\tilde{x}_2(i^*)\}$. We will show

that

$$\mathcal{S}_3 = \mathcal{S}_1 \cup \{[\![\tilde{Y}(i^*)|\tilde{x}_1(i^*)]\!]\} \cup \{[\![\tilde{Y}(i^*)|\tilde{x}_2(i^*)]\!]\}. \tag{9.442}$$

satisfies the property of $[\![Y'|X']\!]^*_{\delta^*/|[\![\bar{X}']\!]|}$, where

$$\delta^* = \frac{\bar{\delta}|[\![X']\!]|m_{\mathscr{Y}}([\![\bar{Y}]\!])}{|[\![\bar{X}]\!]|m_{\mathscr{Y}}([\![Y']\!])}. \tag{9.443}$$

Using (9.426), (9.427) and Claim 2, we have

$$[\![\tilde{Y}(i^*)|\tilde{x}_1(i^*)]\!], [\![\tilde{Y}(i^*)|\tilde{x}_2(i^*)]\!] \subseteq \mathscr{S}_{i^*}. \tag{9.444}$$

This along with the fact that $[\![\bar{Y}|\bar{X}]\!]_{\bar{\delta}/|[\![\bar{X}]\!]|}$ is an overlap family implies that for all $\mathscr{S}' \in \mathcal{S}_1$,

$$m_{\mathscr{Y}}([\![\tilde{Y}(i^*)|\tilde{x}_1(i^*)]\!] \cap \mathscr{S}') \leq \frac{\bar{\delta} m_{\mathscr{Y}}([\![\bar{Y}]\!])}{|[\![\bar{X}]\!]|}, \tag{9.445}$$

and

$$m_{\mathscr{Y}}([\![\tilde{Y}(i^*)|\tilde{x}_2(i^*)]\!] \cap \mathscr{S}') \leq \frac{\bar{\delta} m_{\mathscr{Y}}([\![\bar{Y}]\!])}{|[\![\bar{X}]\!]|}. \tag{9.446}$$

Also, we have that

$$\begin{aligned} m_{\mathscr{Y}}([\![\tilde{Y}(i^*)|\tilde{x}_1(i^*)]\!] \cap [\![\tilde{Y}(i^*)|\tilde{x}_2(i^*)]\!]) &\overset{(a)}{\leq} \frac{\bar{\delta}(m_{\mathscr{Y}}([\![\tilde{Y}(1:n)]\!]))^{1/n}}{|[\![\bar{X}]\!]|} \\ &\overset{(b)}{\leq} \frac{\bar{\delta}(m_{\mathscr{Y}}([\![\bar{Y}(1:n)]\!]))^{1/n}}{|[\![\bar{X}]\!]|} \\ &\overset{(c)}{=} \frac{\bar{\delta} m_{\mathscr{Y}}([\![\bar{Y}]\!])}{|[\![\bar{X}]\!]|}, \end{aligned} \tag{9.447}$$

where (a) follows from (9.436), (b) follows from (9.9) and $[\![\tilde{Y}(1:n)]\!] \subseteq [\![Y(1:n)]\!]$ by Claim 2, and (c) follows from Assumption 3 and (9.182). Additionally, $[\![\tilde{Y}(i^*)|\tilde{x}_1(i)]\!]$ and $[\![\tilde{Y}(i^*)|\tilde{x}_2(i)]\!]$ are singly $\delta^*/|[\![\bar{X}']\!]|$ connected sets. This along with (9.445), (9.446) and (9.447) implies that $\mathcal{S}_3$

satisfies all the properties of $[\![Y'|X']\!]^*_{\delta^*/|[\![X']\!]|}$. It follows that

$$\begin{aligned}|[\![Y'|X']\!]^*_{\delta^*/|[\![X']\!]|}| &\geq |\mathcal{S}_3| \\ &\overset{(a)}{=} |\mathcal{S}_1| + 2 \\ &\overset{(b)}{=} |[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}| + 1, \end{aligned} \tag{9.448}$$

where (a) follows from (9.442), and (b) follows from (9.441).

Now, we will show that

$$|[\![X']\!]| \leq |[\![\bar{X}]\!]| + 1. \tag{9.449}$$

We split the analysis into two mutually exclusive cases: $\tilde{x}_1(i^*) \in [\![\bar{X}]\!]$ or $\tilde{x}_2(i^*) \in [\![\bar{X}]\!]$; and $\tilde{x}_1(i^*), \tilde{x}_2(i^*) \notin [\![\bar{X}]\!]$. In the first case, if $\tilde{x}_1(i^*) \in [\![\bar{X}]\!]$ or $\tilde{x}_2(i^*) \in [\![\bar{X}]\!]$, then using (9.438), we have

$$|[\![X']\!]| \leq |[\![\bar{X}]\!]| + 1. \tag{9.450}$$

In the second case, if $\tilde{x}_1(i^*), \tilde{x}_2(i^*) \notin [\![\bar{X}]\!]$, then using (9.444), there exists a non-empty set $\mathscr{P} \subseteq [\![\bar{X}]\!]$ such that

$$[\![\tilde{Y}(i^*)|\tilde{x}_1(i^*)]\!] \cup [\![\tilde{Y}(i^*)|\tilde{x}_2(i^*)]\!] \subseteq \cup_{x \in \mathscr{P}} [\![\bar{Y}|x]\!]. \tag{9.451}$$

Also, there exists a $x' \in \mathcal{P}$ such that

$$[\![\bar{Y}|x']\!] \subseteq \mathscr{S}_{i^*}. \tag{9.452}$$

This can be proved by contradiction. Let $x' \in \mathcal{P}$ satisfying (9.452) does not exist. We have

$$\begin{aligned}
&m_{\mathscr{Y}}(\cup_{\mathscr{S}' \in [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}: \mathscr{S}' \neq \mathscr{S}_{i^*}} (\mathscr{S}_{i^*} \cap \mathscr{S}')) \\
&\overset{(a)}{\geq} m_{\mathscr{Y}}(\cup_{x: x \in \mathcal{P}} (\mathscr{S}_{i^*} \cap [\![\bar{Y}|x]\!])) \\
&\overset{(b)}{\geq} m_{\mathscr{Y}}([\![\tilde{Y}(i^*)|\tilde{x}_1(i^*)]\!] \cup [\![\tilde{Y}(i^*)|\tilde{x}_2(i^*)]\!]) \\
&\overset{(c)}{\geq} m_{\mathscr{Y}}(V_N) \\
&\overset{(d)}{>} \delta_1 \\
&\overset{(e)}{\geq} \bar{\delta} m_{\mathscr{Y}}([\![\bar{Y}]\!]),
\end{aligned} \tag{9.453}$$

where (a) follows from the fact that combining $\mathcal{P} \subseteq [\![X]\!]$, Property 3 of Definition 14, and the hypothesis that x' does not exist, we have

$$\cup_{x \in \mathscr{P}} [\![\bar{Y}|x]\!] \subseteq \mathscr{S}' \in \cup_{[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}: \mathscr{S}' \neq \mathscr{S}_{i^*}} \mathscr{S}', \tag{9.454}$$

(b) follows from (9.444) and (9.451), (c) follows from (9.9) and the fact that for all $x \in \mathscr{X}$,

$$m_{\mathscr{Y}}(V_N) \leq m_{\mathscr{Y}}([\![Y|x]\!]), \tag{9.455}$$

(d) follows from the fact that $\delta_1 < m_{\mathscr{Y}}(V_N)$, and (e) follows from the fact that $\bar{\delta} \leq \delta_1 / m_{\mathscr{Y}}([\![\bar{Y}]\!])$. On the other hand, since $[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}$ is an overlap family, we have

$$\begin{aligned}
m_{\mathscr{Y}}(\cup_{\mathscr{S}' \in [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}: \mathscr{S}' \neq \mathscr{S}_{i^*}} (\mathscr{S}_{i^*} \cap \mathscr{S}')) &\overset{(a)}{\leq} \sum_{\substack{\mathscr{S}' \in [\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}: \\ \mathscr{S}' \neq \mathscr{S}_{i^*}}} m_{\mathscr{Y}}(\mathscr{S}_{i^*} \cap \mathscr{S}') \\
&\overset{(b)}{\leq} \frac{\bar{\delta} |[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}| m_{\mathscr{Y}}([\![\bar{Y}]\!])}{|[\![\bar{X}]\!]|} \\
&\overset{(c)}{\leq} \bar{\delta} m_{\mathscr{Y}}([\![\bar{Y}]\!]),
\end{aligned} \tag{9.456}$$

where (a) follows from Assumption 4, (b) follows from Property 2 of Definition 14, and (c) follows from the fact that using (9.431), Lemma 35 holds. Hence, (9.453) and (9.456) contradict each other, which implies x' satisfying (9.452) exists. Now, using (9.452) and (9.438) , we have that

$$|[\![X']\!]| \leq |[\![\bar{X}]\!]| + 1. \tag{9.457}$$

Hence, (9.449) holds.

Finally, we have

$$\begin{aligned}\delta^* &= \frac{\bar{\delta}|[\![X']\!]|m_{\mathscr{Y}}([\![\bar{Y}]\!])}{|[\![\bar{X}]\!]|m_{\mathscr{Y}}([\![Y']\!])} \\ &\overset{(a)}{\leq} \frac{\bar{\delta}m_{\mathscr{Y}}([\![\bar{Y}]\!])}{m_{\mathscr{Y}}([\![Y']\!])}\left(1+\frac{1}{|[\![\bar{X}]\!]|}\right) \\ &\overset{(b)}{\leq} \frac{\delta_1}{m_{\mathscr{Y}}([\![Y']\!])},\end{aligned} \tag{9.458}$$

where (a) follows from (9.449), and (b) follows from the assumption in the theorem that $\bar{\delta}(1 + 1/|[\![\bar{X}]\!]|) \leq \delta_1/m_{\mathscr{Y}}([\![\bar{Y}]\!])$. Now, using (9.448) and (9.458), we have that there exists a $\delta^* \leq \delta_1/m_{\mathscr{Y}}([\![Y']\!])$ such that

$$|[\![Y'|X']\!]^*_{\delta^*/|[\![X']\!]|}| > |[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}|. \tag{9.459}$$

This concludes the proof of Claim 4. □

9.12.6 Taxicab symmetry of the mutual information

Definition 31. (δ_1, δ_2)*-taxicab connectedness and* (δ_1, δ_2)*-taxicab isolation.*

- *Points* $(x, y), (x', y') \in [\![X, Y]\!]$ *are* (δ_1, δ_2)*-taxicab connected via* $[\![X, Y]\!]$*, and are denoted by* $(x, y) \overset{\delta_1,\delta_2}{\leftrightsquigarrow} (x', y')$*, if there exists a finite sequence* $\{(x_i, y_i)\}_{i=1}^N$ *of points in* $[\![X, Y]\!]$ *such that* $(x, y) = (x_1, y_1)$*,* $(x', y') = (x_N, y_N)$ *and for all* $2 < i \leq N$*, we have either*

$$A_1 = \{x_i = x_{i-1} \text{ and } \frac{m_{\mathscr{X}}([\![X|y_i]\!] \cap [\![X|y_{i-1}]\!])}{m_{\mathscr{X}}([\![X]\!])} > \delta_1\},$$

or

$$A_2 = \{y_i = y_{i-1} \text{ and } \frac{m_{\mathscr{Y}}([\![Y|x_i]\!] \cap [\![Y|x_{i-1}]\!])}{m_{\mathscr{Y}}([\![Y]\!])} > \delta_2\}.$$

If $(x, y) \overset{\delta_1,\delta_2}{\leftrightsquigarrow} (x', y')$ *and* $N = 2$*, then we say that* (x, y) *and* (x', y') *are singly* (δ_1, δ_2)*-taxicab connected, i.e. either* $y = y'$ *and* $x, x' \in [\![X|y]\!]$ *or* $x = x'$ *and* $y, y' \in [\![Y|x]\!]$*.*

- *A set* $\mathscr{S} \subseteq [\![X, Y]\!]$ *is (singly)* (δ_1, δ_2)*-taxicab connected via* $[\![X, Y]\!]$ *if every pair of points in the set is (singly)* (δ_1, δ_2)*-taxicab connected in* $[\![X, Y]\!]$*.*
- *Two sets* $\mathscr{S}_1, \mathscr{S}_2 \subseteq [\![X, Y]\!]$ *are* (δ_1, δ_2)*-taxicab isolated via* $[\![X, Y]\!]$ *if no point in* $\mathscr{S}_1$ *is* (δ_1, δ_2)*-taxicab connected to any point in* $\mathscr{S}_2$*.*

Definition 32. *Projection of a set*

- *The projection* $\mathscr{S}_x^+$ *of a set* $\mathscr{S} \subseteq [\![X, Y]\!]$ *on the* x*-axis is defined as*

$$\mathscr{S}_x^+ = \{x : (x, y) \in \mathscr{S}\}. \tag{9.460}$$

- *The projection* $\mathscr{S}_y^+$ *of a set* $\mathscr{S} \subseteq [\![X, Y]\!]$ *on the* y*-axis is defined as*

$$\mathscr{S}_y^+ = \{y : (x, y) \in \mathscr{S}\}. \tag{9.461}$$

Definition 33. (δ_1, δ_2)*-taxicab family*

A (δ_1, δ_2)*-taxicab family of* $[\![X, Y]\!]$*, denoted by* $[\![X, Y]\!]^*_{(\delta_1,\delta_2)}$*, is a largest family of distinct sets covering* $[\![X, Y]\!]$ *such that:*

1. *Each set in the family is* (δ_1, δ_2)*-taxicab connected and contains at least one singly* δ_1*-connected set of form* $[\![X|y]\!] \times \{y\}$*, and at least one singly* δ_2*-connected set of the form* $[\![Y|x]\!] \times \{x\}$*.*

2. *The measure of overlap between the projections on the x-axis and y-axis of any two distinct sets in the family are at most $\delta_1 m_{\mathscr{X}}([\![X]\!])$ and $\delta_2 m_{\mathscr{Y}}([\![Y]\!])$ respectively.*

3. *For every singly (δ_1, δ_2)-connected set, there exists a set in the family containing it.*

We now show that when $(X, Y) \overset{d}{\leftrightarrow} (\delta_1, \delta_2)$ hold, the cardinality of (δ_1, δ_2)-taxicab family is same as the cardinality of the $[\![X|Y]\!]$ δ_1-overlap family and $[\![Y|X]\!]$ δ_2-overlap family.

Proof of Theorem 39

We will show that $|[\![X, Y]\!]^*_{(\delta_1,\delta_2)}| = |[\![X|Y]\!]^*_{\delta_1}|$. Then, $|[\![X, Y]\!]^*_{(\delta_1,\delta_2)}| = |[\![Y|X]\!]^*_{\delta_2}|$ can be derived along the same lines. Hence, the statement of the theorem follows.

First, we will show that

$$\mathcal{D} = \{\mathscr{S}^+_x : \mathscr{S} \in [\![X, Y]\!]^*_{(\delta_1,\delta_2)}\}, \tag{9.462}$$

satisfies all the properties of $[\![X|Y]\!]^*_{\delta_1}$.

Since $[\![X, Y]\!]^*_{(\delta_1,\delta_2)}$ is a covering of $[\![X, Y]\!]$, we have

$$\cup_{\mathscr{S}^+_x \in \mathcal{D}} \mathscr{S}^+_x = [\![X]\!], \tag{9.463}$$

which implies $\mathcal{D}$ is a covering of $[\![X]\!]$.

Consider a set $\mathscr{S} \in [\![X, Y]\!]^*_{(\delta_1,\delta_2)}$. For all $(x, y), (x', y') \in \mathscr{S}$, (x, y) and (x', y') are (δ_1, δ_2)-taxicab connected. Then, there exists a taxicab sequence of the form

$$(x, y), (x_1, y), (x_1, y_1), \ldots (x_{n-1}, y'), (x', y').$$

such that either A_1 or A_2 in Definition 31 is true. Then, the sequence $\{y, y_1, \ldots, y_{n-1}, y'\}$ yields

a sequence of conditional range $\{[\![X|\tilde{y}_j]\!]\}_{j=1}^{n+1}$ such that for all $1 < j \leq n+1$,

$$\frac{m_{\mathscr{X}}([\![X|\tilde{y}_j]\!] \cap [\![X|\tilde{y}_{j-1}]\!])}{m_{\mathscr{X}}([\![X]\!])} > \delta_1, \tag{9.464}$$

$$x \in [\![X|\tilde{y}_1]\!], \text{ and } x' \in [\![X|\tilde{y}_{n+1}]\!]. \tag{9.465}$$

Hence, $x \overset{\delta_1}{\rightsquigarrow} x'$ via $[\![X|Y]\!]$. Hence, $\mathscr{S}_x^+$ is δ_1-connected via $[\![X|Y]\!]$. Also, $\mathscr{S}$ contains at least one singly δ_1-connected set of the form $[\![X|y]\!] \times \{y\}$, which implies $[\![X|y]\!] \subseteq \mathscr{S}_x^+$. Hence, $\mathscr{S}_x^+$ contains at least one singly δ_1-connected set of the form $[\![X|y]\!]$. Hence, $\mathcal{D}$ satisfies Property 1 in Definition 14.

For all $\mathscr{S}_1, \mathscr{S}_2 \in [\![X,Y]\!]^*_{(\delta_1,\delta_2)}$, we have

$$m_{\mathscr{X}}(\mathscr{S}_{1,x}^+ \cap \mathscr{S}_{2,x}^+) \leq \delta_1 m_{\mathscr{X}}([\![X]\!]), \tag{9.466}$$

using Property 2 in Definition 33. Hence, $\mathcal{D}$ satisfies Property 2 in Definition 14.

Using Property 3 in Definition 33, we have that for all $[\![X|y]\!] \times \{y\}$, there exists a set $\mathscr{S}(y) \in [\![X,Y]\!]^*_{(\delta_1,\delta_2)}$ containing it. This implies that for all $[\![X|y]\!] \in [\![X|Y]\!]$, we have

$$[\![X|y]\!] \subseteq \mathscr{S}(y)_x^+. \tag{9.467}$$

Hence, $\mathcal{D}$ satisfies Property 3 in Definition 14.

Thus, $\mathcal{D}$ satisfies all the three properties of $[\![X|Y]\!]^*_{\delta_1}$. This implies along with Theorem 38 that

$$|\mathcal{D}| = |[\![X|Y]\!]^*_{\delta_1}|, \tag{9.468}$$

which implies $|[\![X,Y]\!]^*_{(\delta_1,\delta_2)}| = |[\![X|Y]\!]^*_{\delta_1}|$. Hence, the statement of the theorem follows. $\square$

Bibliography

[1] Ittai Abraham, Omar Alonso, Vasilis Kandylas, and Aleksandrs Slivkins. Adaptive crowdsourcing algorithms for the bandit survey problem. In *Conference on learning theory*, pages 882–910, 2013.

[2] Alekh Agarwal, Sarah Bird, Markus Cozowicz, Luong Hoang, John Langford, Stephen Lee, Jiaji Li, Dan Melamed, Gal Oshri, Oswaldo Ribas, et al. Making contextual decisions with low technical debt. *arXiv preprint arXiv:1606.03966*, 2016.

[3] Rajeev Agrawal. Sample mean based index policies with o (log n) regret for the multi-armed bandit problem. *Advances in Applied Probability*, pages 1054–1078, 1995.

[4] Shipra Agrawal and Nikhil Devanur. Linear contextual bandits with knapsacks. In *Advances in Neural Information Processing Systems*, pages 3450–3458, 2016.

[5] Shipra Agrawal and Nikhil R Devanur. Bandits with concave rewards and convex knapsacks. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 989–1006, 2014.

[6] Noga Alon, Boris Bukh, and Yury Polyanskiy. List-decodable zero-rate codes. *IEEE Transactions on Information Theory*, 65(3):1657–1667, 2018.

[7] Noga Alon, Nicolo Cesa-Bianchi, Ofer Dekel, and Tomer Koren. Online learning with feedback graphs: Beyond bandits. In *JMLR WORKSHOP AND CONFERENCE PROCEEDINGS*, volume 40. Microtome Publishing, 2015.

[8] Noga Alon, Nicolo Cesa-Bianchi, Claudio Gentile, Shie Mannor, Yishay Mansour, and Ohad Shamir. Nonstochastic multi-armed bandits with graph-structured feedback. *SIAM Journal on Computing*, 46(6):1785–1826, 2017.

[9] Noga Alon, Nicolo Cesa-Bianchi, Claudio Gentile, and Yishay Mansour. From bandits to experts: A tale of domination and independence. In *Advances in Neural Information Processing Systems*, pages 1610–1618, 2013.

[10] András Antos, Varun Grover, and Csaba Szepesvári. Active learning in multi-armed bandits. In *International Conference on Algorithmic Learning Theory*, pages 287–302. Springer, 2008.

[11] Raman Arora, Ofer Dekel, and Ambuj Tewari. Online bandit learning against an adaptive adversary: from regret to policy regret. *International Conference on Machine Learning*, 2012.

[12] Raman Arora, Teodor V Marinov, and Mehryar Mohri. Bandits with feedback graphs and switching costs. *Advances in Neural Information Processing Systems*, 32, 2019.

[13] Javed Aslam, Zack Butler, Florin Constantin, Valentino Crespi, George Cybenko, and Daniela Rus. Tracking a moving object with a binary sensor network. In *Proceedings of the 1st international conference on Embedded networked sensor systems*, pages 150–161. ACM, 2003.

[14] Nadarajah Asokan, Valtteri Niemi, and Kaisa Nyberg. Man-in-the-middle in tunnelled authentication protocols. In *International Workshop on Security Protocols*, pages 28–41. Springer, 2003.

[15] Luigi Atzori, Antonio Iera, and Giacomo Morabito. The internet of things: A survey. *Computer networks*, 54(15):2787–2805, 2010.

[16] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

[17] Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002.

[18] Peter Auer and Chao-Kai Chiang. An algorithm with nearly optimal pseudo-regret for both stochastic and adversarial bandits. In *Conference on Learning Theory*, pages 116–120, 2016.

[19] Peter Auer, Ronald Ortner, and Csaba Szepesvári. Improved rates for the stochastic continuum-armed bandit problem. In *International Conference on Computational Learning Theory*, pages 454–468. Springer, 2007.

[20] Ashwinkumar Badanidiyuru, Robert Kleinberg, and Aleksandrs Slivkins. Bandits with knapsacks. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 207–216. IEEE, 2013.

[21] Ashwinkumar Badanidiyuru, Robert Kleinberg, and Aleksandrs Slivkins. Bandits with knapsacks. *Journal of the ACM (JACM)*, 65(3):1–55, 2018.

[22] Vahid Behzadan and Arslan Munir. Vulnerability of deep reinforcement learning to policy induction attacks. In *International Conference on Machine Learning and Data Mining in Pattern Recognition*, pages 262–275. Springer, 2017.

[23] Felix Berkenkamp, Matteo Turchetta, Angela Schoellig, and Andreas Krause. Safe model-based reinforcement learning with stability guarantees. In *Advances in neural information processing systems*, pages 908–918, 2017.

[24] Lilian Besson and Emilie Kaufmann. What doubling tricks can and can't do for multi-armed bandits. *arXiv preprint arXiv:1803.06971*, 2018.

[25] Arpita Biswas, Shweta Jain, Debmalya Mandal, and Y Narahari. A truthful budget feasible multi-armed bandit mechanism for crowdsourcing time critical tasks. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 1101–1109. International Foundation for Autonomous Agents and Multiagent Systems, 2015.

[26] Rick S Blum, Saleem A Kassam, and H Vincent Poor. Distributed detection with multiple sensors II. Advanced topics. *Proceedings of the IEEE*, 85(1):64–79, 1997.

[27] Ilija Bogunovic, Arpan Losalka, Andreas Krause, and Jonathan Scarlett. Stochastic linear bandits robust to adversarial attacks. In *International Conference on Artificial Intelligence and Statistics*, pages 991–999. PMLR, 2021.

[28] Reza Rafie Borujeny and Frank R Kschischang. A signal-space distance measure for nondispersive optical fiber. *arXiv preprint arXiv:2001.08663*, 2020.

[29] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Randomized gossip algorithms. *IEEE/ACM Transactions on Networking (TON)*, 14(SI):2508–2530, 2006.

[30] Paolo Braca, Stefano Marano, and Vincenzo Matta. Enforcing consensus while monitoring the environment in wireless sensor networks. *IEEE Transactions on Signal Processing*, 56(7):3375–3380, 2008.

[31] Paolo Braca, Stefano Marano, and Vincenzo Matta. Running consensus in wireless sensor networks. In *11th International Conference on Information Fusion, 2008*, pages 1–6. IEEE, 2008.

[32] Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in multi-armed bandits problems. In *International conference on Algorithmic learning theory*, pages 23–37. Springer, 2009.

[33] Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvári. X-armed bandits. *Journal of Machine Learning Research*, 12(May):1655–1695, 2011.

[34] Sébastien Bubeck and Aleksandrs Slivkins. The best of both worlds: stochastic and adversarial bandits. In *Conference on Learning Theory*, pages 42–1, 2012.

[35] Mona Buisson-Fenet, Friedrich Solowjow, and Sebastian Trimpe. Actively learning gaussian process dynamics. In *Learning for Dynamics and Control*, pages 5–15. PMLR, 2020.

[36] Franco Callegati, Walter Cerroni, and Marco Ramilli. Man-in-the-middle attack to the https protocol. *IEEE Security & Privacy*, 7(1):78–81, 2009.

[37] Alvaro A Cardenas, Saurabh Amin, and Shankar Sastry. Secure control: Towards survivable cyber-physical systems. *System*, 1(a2):a3, 2008.

[38] Stéphane Caron, Branislav Kveton, Marc Lelarge, and Smriti Bhagat. Leveraging side observations in stochastic bandits. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, pages 142–151, 2012.

[39] F. S. Cattivelli and A. H. Sayed. Diffusion LMS strategies for distributed estimation. *IEEE Transactions on Signal Processing*, 58(3):1035–1048, March 2010.

[40] F. S. Cattivelli and A. H. Sayed. Distributed detection over adaptive networks using diffusion adaptation. *IEEE Transactions on Signal Processing*, 59(5):1917–1932, 2011.

[41] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

[42] Z Chair and PK Varshney. Optimal data fusion in multiple sensor detection systems. *IEEE Transactions on Aerospace and Electronic Systems*, (1):98–101, 1986.

[43] Yiding Chen and Xiaojin Zhu. Optimal attack against autoregressive models by manipulating the environment. *arXiv preprint arXiv:1902.00202*, 2019.

[44] Richard Cheng, Mohammad Javad Khojasteh, Aaron D Ames, and Joel W Burdick. Safe multi-agent interaction through robust control barrier functions with learned uncertainties. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 777–783. IEEE, 2020.

[45] Herman Chernoff. Sequential design of experiments. *The Annals of Mathematical Statistics*, 30(3):755–770, 1959.

[46] Amin Coja-Oghlan and Charilaos Efthymiou. On independent sets in random graphs. *Random Structures & Algorithms*, 47(3):436–486, 2015.

[47] Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the linear quadratic regulator. *Foundations of Computational Mathematics*, Aug 2019.

[48] Morris H DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[49] Ofer Dekel, Jian Ding, Tomer Koren, and Yuval Peres. Bandits with switching costs: T 2/3 regret. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 459–467. ACM, 2014.

[50] Bhuwan Dhingra, Lihong Li, Xiujun Li, Jianfeng Gao, Yun-Nung Chen, Faisal Ahmad, and Li Deng. Towards end-to-end reinforcement learning of dialogue agents for information access. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 484–495, 2017.

[51] Seyed Mehran Dibaji, Mohammad Pirani, David Bezalel Flamholz, Anuradha M Annaswamy, Karl Henrik Johansson, and Aranya Chakrabortty. A systems and control perspective of cps security. *Annual Reviews in Control*, 47:394–411, 2019.

[52] Wenkui Ding, Tao Qin, Xu-Dong Zhang, and Tie-Yan Liu. Multi-armed bandit with budget constraint and variable costs. In *Twenty-Seventh AAAI Conference on Artificial Intelligence*, 2013.

[53] Krishna Doddapaneni, Ravi Lakkundi, Suhas Rao, Sujay Gururaj Kulkarni, and Bhargav Bhat. Secure fota object for iot. In *2017 IEEE 42nd Conference on Local Computer Networks Workshops (LCN Workshops)*, pages 154–159. IEEE, 2017.

[54] Jiu-Gang Dong and Li Qiu. Flocking of the cucker-smale model on general digraphs. *IEEE Transactions on Automatic Control*, 62(10):5234–5239, 2017.

[55] Pinar Donmez, Jaime G Carbonell, and Jeff Schneider. Efficiently learning the accuracy of labeling sources for selective sampling. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 259–268. ACM, 2009.

[56] Joseph L Doob. *Stochastic processes*. John Wiley & Sons Inc, 1953.

[57] Vladimir P Dragalin, Alexander G Tartakovsky, and Venugopal V Veeravalli. Multihypothesis sequential probability ratio tests. II. Accurate asymptotic expansions for the expected sample size. *IEEE Transactions on Information Theory*, 46(4):1366–1383, 2000.

[58] VP Draglia, Alexander G Tartakovsky, and Venugopal V Veeravalli. Multihypothesis sequential probability ratio tests. I. Asymptotic optimality. *IEEE Transactions on Information Theory*, 45(7):2448–2461, 1999.

[59] John C Duchi, Alekh Agarwal, and Martin J Wainwright. Dual averaging for distributed optimization: Convergence analysis and network scaling. *IEEE Transactions on Automatic control*, 57(3):592–606, 2012.

[60] Paul Erdos. On a theorem of Hsu and Robbins. *The Annals of Mathematical Statistics*, 20(2):286–291, 1949.

[61] David D Fan, Ali-akbar Agha-mohammadi, and Evangelos A Theodorou. Deep learning tubes for tube mpc. *arXiv preprint arXiv:2002.01587*, 2020.

[62] Michal Feldman, Tomer Koren, Roi Livni, Yishay Mansour, and Aviv Zohar. Online pricing with strategic and patient buyers. In *Advances in Neural Information Processing Systems*, pages 3864–3872, 2016.

[63] Zhe Feng, David Parkes, and Haifeng Xu. The intrinsic robustness of stochastic bandits to strategic manipulation. In *International Conference on Machine Learning*, pages 3092–3101. PMLR, 2020.

[64] Aidin Ferdowsi and Walid Saad. Deep learning for signal authentication and security in massive internet-of-things systems. *IEEE Transactions on Communications*, 67(2):1371–1387, 2018.

[65] Chung-Sung Ferng and Hsuan-Tien Lin. Multi-label classification with error-correcting codes. In *Asian conference on machine learning*, pages 281–295, 2011.

[66] Jaime F Fisac, Anayo K Akametalu, Melanie N Zeilinger, Shahab Kaynama, Jeremy Gillula, and Claire J Tomlin. A general safety framework for learning-based control in uncertain robotic systems. *IEEE Transactions on Automatic Control*, 64(7):2737–2752, 2018.

[67] David B Flamholz, Anuradha M Annaswamy, and Eugene Lavretsky. Baiting for defense against stealthy attacks on cyber-physical systems. In *AIAA Scitech 2019 Forum*, page 2338, 2019.

[68] Massimo Franceschetti, Stefano Marano, and Vincenzo Matta. Chernoff test for strong-or-weak radar models. *IEEE Transactions on Signal Processing*, 65(2):289–302, 2016.

[69] Evrard Garcelon, Baptiste Roziere, Laurent Meunier, Jean Tarbouriech, Olivier Teytaud, Alessandro Lazaric, and Matteo Pirotta. Adversarial attacks on linear contextual bandits. *Advances in Neural Information Processing Systems*, 33, 2020.

[70] Claudio Gentile and Francesco Orabona. On multilabel classification and ranking with bandit feedback. *Journal of Machine Learning Research*, 15(1):2451–2487, 2014.

[71] Sascha Geulen, Berthold Vöcking, and Melanie Winkler. Regret minimization for online buffering problems using the weighted majority algorithm. In *COLT*, pages 132–143, 2010.

[72] Hoda Ghadeer. Cybersecurity issues in internet of things and countermeasures. In *2018 IEEE International Conference on Industrial Internet (ICII)*, pages 195–201. IEEE, 2018.

[73] Negin Golrezaei, Vahideh Manshadi, Jon Schneider, and Shreyas Sekar. Learning product rankings robust to fake users. *Available at SSRN*, 2020.

[74] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

[75] Nathan A Goodman, Phaneendra R Venkata, and Mark A Neifeld. Adaptive waveform design and sequential hypothesis testing for target recognition with active sensors. *IEEE Journal of Selected Topics in Signal Processing*, 1(1):105–113, 2007.

[76] Shivani Goyal and Rejo Mathew. Security issues in cloud computing. In *International conference on Computer Networks, Big data and IoT*, pages 363–373. Springer, 2019.

[77] Sudipto Guha and Kamesh Munagala. Approximation algorithms for budgeted learning problems. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 104–113. ACM, 2007.

[78] Anupam Gupta, Tomer Koren, and Kunal Talwar. Better algorithms for stochastic bandits with adversarial corruptions. In *Conference on Learning Theory*, pages 1562–1578, 2019.

[79] Andras Gyorgy and Gergely Neu. Near-optimal rates for limited-delay universal lossy source coding. *IEEE Transactions on Information Theory*, 60(5):2823–2834, 2014.

[80] Ralph VL Hartley. Transmission of information 1. *Bell System technical journal*, 7(3):535–563, 1928.

[81] Ernst Haselsteiner and Klemens Breitfuß. Security in near field communication (nfc). In *Workshop on RFID security*, pages 12–14. sn, 2006.

[82] Navid Hashemi and Justin Ruths. Gain design via LMIs to minimize the impact of stealthy attacks. In *2020 American Control Conference (ACC)*, pages 1274–1279. IEEE, 2020.

[83] Chien-Ju Ho, Shahin Jabbari, and Jennifer W Vaughan. Adaptive task assignment for crowdsourced classification. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 534–542, 2013.

[84] Chien-Ju Ho and Jennifer Wortman Vaughan. Online task assignment in crowdsourcing markets. In *AAAI*, volume 12, pages 45–51, 2012.

[85] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. In *The Collected Works of Wassily Hoeffding*, pages 409–426. Springer, 1994.

[86] Andreas Hoehn and Ping Zhang. Detection of covert attacks and zero dynamics attacks in cyber-physical systems. In *American Control Conference (ACC), 2016*, pages 302–307. IEEE, 2016.

[87] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1985.

[88] Yanling Hu, Mianxiong Dong, Kaoru Ota, Anfeng Liu, and Minyi Guo. Mobile target detection in wireless sensor networks with adjustable sensing frequency. *IEEE Systems Journal*, 10(3):1160–1171, 2016.

[89] Ling Huang, Anthony D Joseph, Blaine Nelson, Benjamin IP Rubinstein, and J Doug Tygar. Adversarial machine learning. In *Proceedings of the 4th ACM workshop on Security and artificial intelligence*, pages 43–58, 2011.

[90] M. Huang and J. H. Manton. Stochastic consensus seeking with noisy and directed inter-agent communication: Fixed and randomly varying topologies. *IEEE Transactions on Automatic Control*, 55(1):235–241, Jan 2010.

[91] Sandy Huang, Nicolas Papernot, Ian Goodfellow, Yan Duan, and Pieter Abbeel. Adversarial attacks on neural network policies. *arXiv preprint arXiv:1702.02284*, 2017.

[92] Yunhan Huang and Quanyan Zhu. Manipulating reinforcement learning: Poisoning attacks on cost signals. *arXiv preprint arXiv:2002.03827*, 2020.

[93] Ali Jadbabaie, Pooya Molavi, Alvaro Sandroni, and Alireza Tahbaz-Salehi. Non-Bayesian social learning. *Games and Economic Behavior*, 76(1):210–225, 2012.

[94] Ruixiang Jiang and Biao Chen. Fusion of censored decisions in wireless sensor networks. *IEEE Transactions on Wireless Communications*, 4(6):2668–2673, 2005.

[95] Yu Jiang and Zhong-Ping Jiang. Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics. *Automatica*, 48(10):2699–2704, 2012.

[96] Rong Jin and Zoubin Ghahramani. Learning with multiple labels. In *Advances in neural information processing systems*, pages 921–928, 2003.

[97] Kwang-Sung Jun, Lihong Li, Yuzhe Ma, and Jerry Zhu. Adversarial attacks on stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 3640–3649, 2018.

[98] Aris Kanellopoulos and Kyriakos G Vamvoudakis. A moving target defense control framework for cyber-physical systems. *IEEE Transactions on Automatic Control*, 65(3):1029–1043, 2019.

[99] S. Kar and J. M. Moura. Distributed consensus algorithms in sensor networks: quantized data and random link failures. *IEEE Transactions on Signal Processing*, 58(3):1383–1400, March 2010.

[100] S. Kar and J. M. F. Moura. Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise. *IEEE Transactions on Signal Processing*, 57(1):355–369, Jan 2009.

[101] Soummya Kar and José MF Moura. Sensor networks with random links: Topology design for distributed consensus. *IEEE Transactions on Signal Processing*, 56(7):3315–3326, 2008.

[102] David R Karger, Sewoong Oh, and Devavrat Shah. Budget-optimal crowdsourcing using low-rank matrix approximations. In *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, pages 284–291. IEEE, 2011.

[103] David R Karger, Sewoong Oh, and Devavrat Shah. Iterative learning for reliable crowdsourcing systems. In *Advances in neural information processing systems*, pages 1953–1961, 2011.

[104] Chris Karlof and David Wagner. Secure routing in wireless sensor networks: Attacks and countermeasures. *Ad hoc networks*, 1(2-3):293–315, 2003.

[105] Sumeet Katariya, Branislav Kveton, Csaba Szepesvari, and Zheng Wen. Dcm bandits: Learning to rank with multiple clicks. In *International Conference on Machine Learning*, pages 1215–1224, 2016.

[106] Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42, 2016.

[107] H. Kellerer, U. Pferschy, and D. Pisinger. *Knapsack Problems*. Springer, Berlin, Germany, 2004.

[108] David Kempe, Alin Dobra, and Johannes Gehrke. Gossip-based computation of aggregate information. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003.*, pages 482–491. IEEE, 2003.

[109] Ashish Khetan and Sewoong Oh. Achieving budget-optimality with adaptive schemes in crowdsourcing. In *Advances in Neural Information Processing Systems*, pages 4844–4852, 2016.

[110] Mohammad Javad Khojasteh, Vikas Dhiman, Massimo Franceschetti, and Nikolay Atanasov. Probabilistic safety constraints for learned high relative degree system dynamics. In *Learning for Dynamics and Control*, pages 781–792, 2020.

[111] Mohammad Javad Khojasteh, Anatoly Khina, Massimo Franceschetti, and Tara Javidi. Authentication of cyber-physical systems under learning-based attacks. *IFAC-PapersOnLine*, 52(20):369–374, 2019.

[112] Mohammad Javad Khojasteh, Anatoly Khina, Massimo Franceschetti, and Tara Javidi. Learning-based attacks in cyber-physical systems. *IEEE Transactions on Control of Network Systems*, 2020.

[113] Mohammad Javad Khojasteh, Anatoly Khina, Massimo Franceschetti, and Tara Javidi. Learning-based attacks in cyber-physical systems. *IEEE Transactions on Control of Network Systems*, 8(1):437–449, 2021.

[114] J Kiefer and J Sacks. Asymptotically optimum sequential inference and design. *The Annals of Mathematical Statistics*, pages 705–750, 1963.

[115] Kyoung-Dae Kim and Panganamala R Kumar. Cyber-physical systems: A perspective at the centennial. *Proceedings of the IEEE*, 100(Centennial Issue):1287–1308, 2012.

[116] Robert D Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Advances in Neural Information Processing Systems*, pages 697–704, 2005.

[117] George J Klir. *Uncertainty and Information. Foundations of Generalized Information Theory*. John Wiley, 2006.

[118] Tomáš Kocák, Gergely Neu, and Michal Valko. Online learning with erdős-rényi side-observation graphs. In *Uncertainty in Artificial Intelligence*, 2016.

[119] Andrey Nikolaevich Kolmogorov. Certain asymptotic characteristics of completely bounded metric spaces. *Doklady Akademii Nauk SSSR*, 108(3):385–388, 1956.

[120] Tomer Koren, Roi Livni, and Yishay Mansour. Bandits with movement costs and adaptive pricing. In *Conference on Learning Theory*, pages 1242–1268. PMLR, 2017.

[121] Tomer Koren, Roi Livni, and Yishay Mansour. Multi-armed bandits with metric movement costs. In *Advances in Neural Information Processing Systems*, pages 4122–4131, 2017.

[122] Jernej Kos and Dawn Song. Delving into adversarial attacks on deep policies. *arXiv preprint arXiv:1705.06452*, 2017.

[123] Purushottam Kulkarni, Deepak Ganesan, Prashant Shenoy, and Qifeng Lu. Senseye: a multi-tier camera sensor network. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 229–238. ACM, 2005.

[124] Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

[125] Anusha Lalitha, Tara Javidi, and Anand D Sarwate. Social learning and distributed hypothesis testing. *IEEE Transactions on Information Theory*, 64(9):6161–6179, 2018.

[126] John Langford and Tong Zhang. The epoch-greedy algorithm for multi-armed bandits with side information. In *Advances in neural information processing systems*, pages 817–824, 2008.

[127] Armin Lederer, Jonas Umlauft, and Sandra Hirche. Uniform error bounds for gaussian process regression with application to safe control. In *Advances in Neural Information Processing Systems*, pages 659–669, 2019.

[128] Chong Li and Meikang Qiu. *Reinforcement Learning for Cyber-Physical Systems: with Cybersecurity Case Studies*. CRC Press, 2019.

[129] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670, 2010.

[130] Shancang Li, Li Da Xu, and Shanshan Zhao. The internet of things: a survey. *Information Systems Frontiers*, 17(2):243–259, 2015.

[131] Shang Li and Xiaodong Wang. Fully distributed sequential hypothesis testing: Algorithms and asymptotic analyses. *IEEE Transactions on Information Theory*, 64(4):2742–2758, 2018.

[132] Taehyung J Lim and Massimo Franceschetti. Information without rolling dice. *IEEE Transactions on Information Theory*, 63(3):1349–1363, 2017.

[133] Yen-Chen Lin, Zhang-Wei Hong, Yuan-Hong Liao, Meng-Li Shih, Ming-Yu Liu, and Min Sun. Tactics of adversarial attack on deep reinforcement learning agents. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 3756–3762, 2017.

[134] Ying Lin, Biao Chen, and Pramod K Varshney. Decision fusion rules in multi-hop wireless sensor networks. *IEEE Transactions on Aerospace and Electronic Systems*, 41(2):475–488, 2005.

[135] Fang Liu and Ness Shroff. Data poisoning attacks on stochastic bandits. In *International Conference on Machine Learning*, pages 4042–4050, 2019.

[136] G. Liu and L. Lai. Action-manipulation attacks on stochastic bandits. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3112–3116, 2020.

[137] Guanlin Liu and Lifeng Lai. Action-manipulation attacks against stochastic bandits: Attacks and defense. *IEEE Transactions on Signal Processing*, 68:5152–5165, 2020.

[138] Yuxin Liu, Mianxiong Dong, Kaoru Ota, and Anfeng Liu. Activetrust: secure and trustable routing in wireless sensor networks. *IEEE Transactions on Information Forensics and Security*, 11(9):2013–2027, 2016.

[139] Thodoris Lykouris, Vahab Mirrokni, and Renato Paes Leme. Stochastic bandits robust to adversarial corruptions. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 114–122. ACM, 2018.

[140] Thodoris Lykouris, Max Simchowitz, Aleksandrs Slivkins, and Wen Sun. Corruption robust exploration in episodic reinforcement learning. *arXiv preprint arXiv:1911.08689*, 2019.

[141] Yuzhe Ma, Kwang-Sung Jun, Lihong Li, and Xiaojin Zhu. Data poisoning attacks in contextual bandits. In *International Conference on Decision and Game Theory for Security*, pages 186–204. Springer, 2018.

[142] Yuzhe Ma, Xuezhou Zhang, Wen Sun, and Jerry Zhu. Policy poisoning in batch reinforcement learning and control. In *Advances in Neural Information Processing Systems*, pages 14570–14580, 2019.

[143] Yuzhe Ma, Xiaojin Zhu, and Justin Hsu. Data poisoning against differentially-private learners: attacks and defenses. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 4732–4738. AAAI Press, 2019.

[144] Luca Mainetti, Luigi Patrono, and Antonio Vilei. Evolution of wireless sensor networks towards the internet of things: A survey. In *19th international conference on software, telecommunications and computer networks, SoftCOM 2011*, pages 1–6. IEEE, 2011.

[145] Shie Mannor and Ohad Shamir. From bandits to experts: On the value of side-observations. In *Advances in Neural Information Processing Systems*, pages 684–692, 2011.

[146] Yanbing Mao, Hamidreza Jafarnejadsani, Pan Zhao, Emrah Akyol, and Naira Hovakimyan. Novel stealthy attack and defense strategies for networked control systems. *IEEE Transactions on Automatic Control*, 2020.

[147] S. Marano and A. H. Sayed. Detection under one-bit messaging over adaptive networks. *IEEE Transactions on Information Theory*, 65(10):6519–6538, October 2019.

[148] Stefano Marano, Peter Willett, and Vincenzo Matta. Sequential testing of sorted and transformed data as an efficient way to implement long GLRTs. *IEEE Transactions on Signal Processing*, 51(2):325–337, 2003.

[149] Yajun Mei. Asymptotic optimality theory for decentralized sequential hypothesis testing in sensor networks. *IEEE Transactions on Information Theory*, 54(5):2072–2089, 2008.

[150] Chunyan Miao, Han Yu, Zhiqi Shen, and Cyril Leung. Balancing quality and budget considerations in mobile crowdsourcing. *Decision Support Systems*, 90:56–64, 2016.

[151] Fei Miao, Miroslav Pajic, and George J Pappas. Stochastic game approach for replay attack detection. In *Decision and control (CDC), 2013 IEEE 52nd annual conference on*, pages 1854–1859. IEEE, 2013.

[152] Yilin Mo, Rohan Chabukswar, and Bruno Sinopoli. Detecting integrity attacks on SCADA systems. *IEEE Transactions on Control Systems Technology*, 22(4):1396–1407, 2014.

[153] Yilin Mo, Sean Weerakkody, and Bruno Sinopoli. Physical authentication of control systems: designing watermarked control inputs to detect counterfeit sensor outputs. *IEEE Control Systems*, 35(1):93–109, 2015.

[154] Mohammad Naghshvar and Tara Javidi. Extrinsic Jensen-Shannon divergence with application in active hypothesis testing. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 2191–2195. IEEE, 2012.

[155] Mohammad Naghshvar and Tara Javidi. Sequentiality and adaptivity gains in active hypothesis testing. *IEEE Journal of Selected Topics in Signal Processing*, 7(5):768–782, 2013.

[156] Mohammad Naghshvar, Tara Javidi, et al. Active sequential hypothesis testing. *The Annals of Statistics*, 41(6):2703–2738, 2013.

[157] Girish N Nair. A nonstochastic information theory for feedback. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 1343–1348. IEEE, 2012.

[158] Girish N Nair. A nonstochastic information theory for communication and state estimation. *IEEE Transactions on automatic control*, 58(6):1497–1510, 2013.

[159] R. Nassif, C. Richard, A. Ferrari, and A. H. Sayed. Multitask diffusion adaptation over asynchronous networks. *IEEE Transactions on Signal Processing*, 64(11):2835–2850, June 2016.

[160] Angelia Nedic, Alex Olshevsky, Asuman Ozdaglar, and John N Tsitsiklis. On distributed averaging algorithms and quantization effects. *IEEE Transactions on Automatic Control*, 54(11):2506–2517, 2009.

[161] Angelia Nedić, Alex Olshevsky, and César A Uribe. Nonasymptotic convergence rates for cooperative learning over time-varying directed graphs. In *American Control Conference (ACC), 2015*, pages 5884–5889. IEEE, 2015.

[162] Angelia Nedić, Alex Olshevsky, and César A Uribe. Network independent rates in distributed learning. In *American Control Conference (ACC), 2016*, pages 1072–1077. IEEE, 2016.

[163] George L Nemhauser and Laurence A Wolsey. Best algorithms for approximating the maximum of a submodular set function. *Mathematics of operations research*, 3(3):177–188, 1978.

[164] Sirin Nitinawarat, George K Atia, and Venugopal V Veeravalli. Controlled sensing for multihypothesis testing. *IEEE Transactions on Automatic Control*, 58(10):2451–2464, 2013.

[165] Reza Olfati-Saber, J Alex Fax, and Richard M Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233, 2007.

[166] Rohit Parasnis, Massimo Franceschetti, and Behrouz Touri. Non-Bayesian social learning on random digraphs with aperiodically varying network connectivity. *arXiv preprint arXiv:2010.06695*, 2020.

[167] Fabio Pasqualetti, Florian Dörfler, and Francesco Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, 2013.

[168] S. Patterson, B. Bamieh, and A. El Abbadi. Convergence rates of distributed average consensus with stochastic link failures. *IEEE Transactions on Automatic Control*, 55(4):880–892, April 2010.

[169] H. V. Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, New York, 1988.

[170] Matthew Porter, Pedro Hespanhol, Anil Aswani, Matthew Johnson-Roberson, and Ramanarayan Vasudevan. Detecting generalized replay attacks via time-varying dynamic watermarking. *IEEE Transactions on Automatic Control*, 2020.

[171] Amin Rakhsha, Goran Radanovic, Rati Devidze, Xiaojin Zhu, and Adish Singla. Policy teaching in reinforcement learning via environment poisoning attacks. *arXiv preprint arXiv:2011.10824*, 2020.

[172] Amin Rakhsha, Goran Radanovic, Rati Devidze, Xiaojin Zhu, and Adish Singla. Policy teaching via environment poisoning: Training-time adversarial attacks against reinforcement learning. In *International Conference on Machine Learning*, pages 7974–7984. PMLR, 2020.

[173] Amin Rakhsha, Xuezhou Zhang, Xiaojin Zhu, and Adish Singla. Reward poisoning in reinforcement learning: Attacks against unknown learners in unknown environments. *arXiv preprint arXiv:2102.08492*, 2021.

[174] Anshuka Rangi and Massimo Franceschetti. Multi-armed bandit algorithms for crowdsourcing systems with online estimation of workers' ability. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1345–1352. International Foundation for Autonomous Agents and Multiagent Systems, 2018.

[175] Anshuka Rangi and Massimo Franceschetti. Online learning with feedback graphs and switching costs. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2435–2444, 2019.

[176] Anshuka Rangi and Massimo Franceschetti. Towards a non-stochastic information theory. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 997–1001. IEEE, 2019.

[177] Anshuka Rangi, Massimo Franceschetti, and Stefano Marano. Consensus-based chernoff test in sensor networks. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 6773–6778. IEEE, 2018.

[178] Anshuka Rangi, Massimo Franceschetti, and Stefano Marano. Decentralized Chernoff test in sensor networks. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 501–505. IEEE, 2018.

[179] Anshuka Rangi, Massimo Franceschetti, and Stefano Marano. Distributed chernoff test: Optimal decision systems over networks. *IEEE Transactions on Information Theory*, 2020.

[180] Anshuka Rangi, Massimo Franceschetti, and Long Tran-Thanh. Unifying the stochastic and the adversarial bandits with knapsack. *International Joint Conference on Artificial Intelligence*, 2019.

[181] Anshuka Rangi, Massimo Franceschetti, and Long Tran-Thanh. Unifying the stochastic and the adversarial bandits with knapsack. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 3311–3317. AAAI Press, 2019.

[182] Anshuka Rangi, Mohammad Javad Khojasteh, and Massimo Franceschetti. Learning-based attacks in cyber-physical systems: Exploration, detection, and control cost trade-offs. *arXiv preprint arXiv:2011.10718*, 2020.

[183] Anshuka Rangi, Long Tran-Thanh, Haifeng Xu, and Massimo Franceschetti. Saving stochastic bandits from poisoning attacks via limited data verification. *arXiv preprint arXiv:2102.07711*, 2021.

[184] Anders Rantzer. Concentration bounds for single parameter adaptive control. In *2018 Annual American Control Conference (ACC)*, pages 1862–1866. IEEE, 2018.

[185] Shansi Ren, Qun Li, Haining Wang, Xin Chen, and Xiaodong Zhang. Design and analysis of sensing scheduling algorithms under partial coverage for object detection in sensor networks. *IEEE Transactions on Parallel and Distributed Systems*, 18(3):334–350, 2007.

[186] Alfréd Rényi et al. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California, 1961.

[187] M Rosenfeld. On a problem of C.E. Shannon in graph theory. *Proceedings of the American Mathematical Society*, 18(2):315–319, 1967.

[188] Amir Saberi, Farhad Farokhi, and Girish Nair. Estimation and control over a nonstochastic binary erasure channel. *IFAC-PapersOnLine*, 51(23):265–270, 2018.

[189] Amir Saberi, Farhad Farokhi, and Girish N Nair. State estimation via worst-case erasure and symmetric channels with memory. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 3072–3076. IEEE, 2019.

[190] Karthik Abinav Sankararaman and Aleksandrs Slivkins. Combinatorial semi-bandits with knapsacks. In *International Conference on Artificial Intelligence and Statistics*, pages 1760–1770. PMLR, 2018.

[191] Tuhin Sarkar and Alexander Rakhlin. Near optimal finite time identification of arbitrary linear dynamical systems. In *Int. Conf. on Machine Learning (ICML)*, pages 5610–5618, 2019.

[192] Bharadwaj Satchidanandan and Panganamala R Kumar. Dynamic watermarking: Active defense of networked cyber–physical systems. *Proceedings of the IEEE*, 105(2):219–240, 2017.

[193] Bharadwaj Satchidanandan and PR Kumar. Control systems under attack: The securable and unsecurable subspaces of a linear stochastic system. In *Emerging Applications of Control and Systems Theory*, pages 217–228. Springer, 2018.

[194] Yevgeny Seldin and Gábor Lugosi. An improved parametrization and analysis of the exp3++ algorithm for stochastic and adversarial bandits. In *Conference on Learning Theory*, pages 1743–1759, 2017.

[195] Yevgeny Seldin and Aleksandrs Slivkins. One practical algorithm for both stochastic and adversarial bandits. In *International Conference on Machine Learning*, pages 1287–1295, 2014.

[196] Shahin Shahrampour and Ali Jadbabaie. Exponentially fast parameter estimation in networks using distributed dual averaging. In *52nd IEEE Conference on Decision and Control*, pages 6196–6201. IEEE, 2013.

[197] Shahin Shahrampour, Alexander Rakhlin, and Ali Jadbabaie. Distributed detection: Finite-time analysis and impact of network topology. *IEEE Transactions on Automatic Control*, 61(11):3256–3268, 2016.

[198] Claude Shannon. The zero error capacity of a noisy channel. *IRE Transactions on Information Theory*, 2(3):8–19, 1956.

[199] Claude Elwood Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

[200] Hidenori Shingin and Yoshito Ohta. Disturbance rejection with information constraints: Performance limitations of a scalar system for bounded and Gaussian disturbances. *Automatica*, 48(6):1111–1116, 2012.

[201] Yasser Shoukry, Michelle Chong, Masashi Wakaiki, Pierluigi Nuzzo, Alberto Sangiovanni-Vincentelli, Sanjit A Seshia, Joao P Hespanha, and Paulo Tabuada. Smt-based observer design for cyber-physical systems under sensor attacks. *ACM Transactions on Cyber-Physical Systems*, 2(1):5, 2018.

[202] Bartlomiej Sieka and Ajay D Kshemkalyani. Establishing authenticated channels and secure identifiers in ad-hoc networks. *IJ Network Security*, 5(1):51–61, 2007.

[203] Max Simchowitz, Horia Mania, Stephen Tu, Michael I Jordan, and Benjamin Recht. Learning without mixing: Towards a sharp analysis of linear system identification. *In Conference on Learning Theory*, 2018.

[204] Roy S Smith. Covert misappropriation of networked control systems: Presenting a feedback structure. *IEEE Control Systems*, 35(1):82–92, 2015.

[205] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013.

[206] S. Tantaratana. Some recent results on sequential detection. In *Advances in Statistical Signal Processing*, volume 2, pages 265–296. JAI, New York, 1993.

[207] S. Tantaratana and H. V. Poor. Asymptotic efficiencies of truncated sequential tests. *IEEE Transactions on Information Theory*, 28:911–923, 1982.

[208] André Teixeira, Iman Shames, Henrik Sandberg, and Karl Henrik Johansson. A secure control framework for resource-limited adversaries. *Automatica*, 51:135–148, 2015.

[209] V. M. Tikhomirov and A. N. Kolmogorov. ϵ-entropy and ϵ-capacity of sets in functional spaces. *Uspekhi Matematicheskikh Nauk*, 14(2):3–86, 1959.

[210] Long Tran-Thanh, Archie Chapman, Enrique Munoz de Cote, Alex Rogers, and Nicholas R Jennings. Epsilon–first policies for budget–limited multi-armed bandits. In *Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010.

[211] Long Tran-Thanh, Archie C Chapman, Alex Rogers, and Nicholas R Jennings. Knapsack based optimal policies for budget-limited multi-armed bandits. In *AAAI*, pages 1134–1140, 2012.

[212] Long Tran-Thanh, Alex Rogers, and Nicholas R Jennings. Long-term information collection with energy harvesting wireless sensors: a multi-armed bandit based approach. *Autonomous Agents and Multi-Agent Systems*, 25(2):352–394, 2012.

[213] Long Tran-Thanh, Sebastian Stein, Alex Rogers, and Nicholas R Jennings. Efficient crowdsourcing of unknown experts using bounded multi-armed bandits. *Artificial Intelligence*, 214:89–111, 2014.

[214] J. N. Tsitsiklis. Decentralized detection by a large number of sensors. *Math. Contr., Signals, Syst.*, 1:167–182, 1988.

[215] Pramod K Varshney. *Distributed detection and data fusion*. Springer Science & Business Media, 2012.

[216] Venugopal V Veeravalli, Tamer Basar, and H Vincent Poor. Decentralized sequential detection with a fusion center performing the sequential test. *IEEE Transactions on Information Theory*, 39(2):433–442, 1993.

[217] Sai Vemprala and Ashish Kapoor. Adversarial attacks on optimization based planners. *arXiv preprint arXiv:2011.00095*, 2020.

[218] Gunjan Verma and Ananthram Swami. Error correcting output codes improve probability estimation and adversarial robustness of deep neural networks. In *Advances in Neural Information Processing Systems*, pages 8646–8656, 2019.

[219] Ramanarayanan Viswanathan and Pramod K Varshney. Distributed detection with multiple sensors part I. Fundamentals. *Proceedings of the IEEE*, 85(1):54–63, 1997.

[220] Draguna Vrabie, O Pastravanu, Murad Abu-Khalaf, and Frank L Lewis. Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica*, 45(2):477–484, 2009.

[221] Abraham Wald. *Sequential analysis*. Courier Corporation, 1973.

[222] Abraham Wald and Jacob Wolfowitz. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, pages 326–339, 1948.

[223] J. Wang and N. Elia. Distributed averaging algorithms resilient to communication noise and dropouts. *IEEE Transactions on Signal Processing*, 61(9):2231–2242, May 2013.

[224] Shiqi Wang, Kexin Pei, Justin Whitehouse, Junfeng Yang, and Suman Jana. Efficient formal safety analysis of neural networks. In *Advances in Neural Information Processing Systems*, pages 6367–6377, 2018.

[225] Yan Wang and Yajun Mei. Asymptotic optimality theory for decentralized sequential multihypothesis testing problems. *IEEE Transactions on Information Theory*, 57(10):7068–7083, 2011.

[226] Sean Weerakkody, Omur Ozel, Yilin Mo, Bruno Sinopoli, et al. Resilient control in cyber-physical systems: Countering uncertainty, constraints, and adversarial behavior. *Foundations and Trends® in Systems and Control*, 7(1-2):1–252, 2019.

[227] Lily Weng, Huan Zhang, Hongge Chen, Zhao Song, Cho-Jui Hsieh, Luca Daniel, Duane Boning, and Inderjit Dhillon. Towards fast computation of certified robustness for relu networks. In *International Conference on Machine Learning*, pages 5276–5285. PMLR, 2018.

[228] Tsui-Wei Weng, Huan Zhang, Pin-Yu Chen, Jinfeng Yi, Dong Su, Yupeng Gao, Cho-Jui Hsieh, and Luca Daniel. Evaluating the robustness of neural networks: An extreme value theory approach. In *International Conference on Learning Representations*, 2018.

[229] Jacob Whitehill, Ting-fan Wu, Jacob Bergsma, Javier R Movellan, and Paul L Ruvolo. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Advances in neural information processing systems*, pages 2035–2043, 2009.

[230] Moritz Wiese, Karl Henrik Johansson, Tobias J Oechtering, Panos Papadimitratos, Henrik Sandberg, and Mikael Skoglund. Uncertain wiretap channels and secure estimation. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 2004–2008. IEEE, 2016.

[231] Yifan Wu, András György, and Csaba Szepesvári. Online learning with gaussian payoffs and side observations. In *Advances in Neural Information Processing Systems*, pages 1360–1368, 2015.

[232] Yingce Xia, Tao Qin, Weidong Ma, Nenghai Yu, and Tie-Yan Liu. Budgeted multi-armed bandits with multiple plays. In *International Joint Conference on Artificial Intelligence*, pages 2210–2216, 2016.

[233] Jin-Jun Xiao and Zhi-Quan Luo. Universal decentralized detection in a bandwidth-constrained sensor network. *IEEE Transactions on Signal Processing*, 53(8):2617–2624, August 2005.

[234] Lin Xiao and Stephen Boyd. Fast linear iterations for distributed averaging. *Systems & Control Letters*, 53(1):65–78, 2004.

[235] Pei Xie, Keyou You, and Cheng Wu. How to stop consensus algorithms, locally? In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 4544–4549. IEEE, 2017.

[236] Lin Yang, Mohammad Hajiesmaili, Mohammad Sadegh Talebi, John Lui, Wing Shing Wong, et al. Adversarial bandits with corruptions: Regret lower bound and no-regret algorithm. *Advances in Neural Information Processing Systems*, 33, 2020.

[237] Andrew Chi-Chin Yao. Probabilistic computations: Toward a unified measure of complexity. In *Foundations of Computer Science, 1977., 18th Annual Symposium on*, pages 222–227. IEEE, 1977.

[238] Jia Yuan Yu and Shie Mannor. Unimodal bandits. In *International Conference on Machine Learning*, 2011.

[239] Omar F Zaidan and Chris Callison-Burch. Crowdsourcing translation: Professional quality from non-professionals. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 1220–1229. Association for Computational Linguistics, 2011.

[240] Albert Zhan, Stas Tiomkin, and Pieter Abbeel. Preventing imitation learning with adversarial policy ensembles. *arXiv preprint arXiv:2002.01059*, 2020.

[241] Haoqi Zhang, David C Parkes, and Yiling Chen. Policy teaching through reward function learning. In *Proceedings of the 10th ACM conference on Electronic commerce*, pages 295–304, 2009.

[242] Xuezhou Zhang, Yuzhe Ma, Adish Singla, and Xiaojin Zhu. Adaptive reward-poisoning attacks against reinforcement learning. In *International Conference on Machine Learning*, pages 11225–11234. PMLR, 2020.

[243] Xuezhou Zhang, Xiaojin Zhu, and Laurent Lessard. Online data poisoning attacks. In *Learning for Dynamics and Control*, pages 201–210, 2020.

[244] Z. Zhang, E. K. P. Chong, A. Pezeshki, W. Moran, and S. D. Howard. Detection performance in balanced binary relay trees with node and link failures. *IEEE Transactions on Signal Processing*, 61(9):2165–2177, May 2013.

[245] Xiangyu Zhao, Long Xia, Liang Zhang, Zhuoye Ding, Dawei Yin, and Jiliang Tang. Deep reinforcement learning for page-wise recommendations. In *Proceedings of the 12th ACM Conference on Recommender Systems*, pages 95–103, 2018.

[246] Yaling Zheng, Stephen Scott, and Kun Deng. Active learning from multiple noisy labelers with varied costs. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 639–648. IEEE, 2010.

[247] Datong P Zhou and Claire J Tomlin. Budget-constrained multi-armed bandits with multiple plays. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[248] Dengyong Zhou, Qiang Liu, John C Platt, Christopher Meek, and Nihar B Shah. Regularized minimax conditional entropy for crowdsourcing. *arXiv preprint arXiv:1503.07240*, 2015.

[249] Minghui Zhu and Sonia Martínez. On the performance analysis of resilient networked control systems under replay attacks. *IEEE Transactions on Automatic Control*, 59(3):804–808, 2014.

[250] Xiaojin Zhu. An optimal control view of adversarial machine learning. *arXiv preprint arXiv:1811.04422*, 2018.

[251] Ingvar Ziemann and Henrik Sandberg. Parameter privacy versus control performance: Fisher information regularized control. In *2020 American Control Conference (ACC)*, pages 1259–1265. IEEE, 2020.

[252] Shi Zong, Hao Ni, Kenny Sung, Nan Rosemary Ke, Zheng Wen, and Branislav Kveton. Cascading bandits for large-scale recommendation problems. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*, pages 835–844, 2016.

[253] Shiliang Zuo. Near optimal adversarial attack on UCB bandits. *arXiv preprint arXiv:2008.09312*, 2020.