ABSTRACT

Title of dissertation: PRIVATE INFORMATION RETRIEVAL WITH SIDE INFORMATION AND CODING FOR SECURITY

Yi-Peng Wei, Doctor of Philosophy, 2018

Dissertation directed by: Professor Şennur Ulukuş
Department of Electrical and Computer Engineering

This dissertation studies privacy and security problems from an information-theoretic point of view. We study the privacy problem via the private information retrieval (PIR) problem with a focus on its interactions with available side information. We study the security problem via the wiretap channel with a focus on the design of practical coding schemes to achieve information-theoretically achievable random-coding based secrecy rates.

First, we consider the problem of PIR from N non-colluding and replicated databases when the user is equipped with a cache that holds an uncoded fraction r from each of the K stored messages in the databases. We consider the case where the databases are unaware of the cache content. We investigate $D^*(r)$ the optimal download cost normalized with the message size as a function of K, N, r. For a fixed K, N, we develop converses and achievability schemes for the $D^*(r)$ curve. The largest additive gap between our achievability and the converse bounds is $\frac{1}{6}$. Our results show that the download cost can be reduced beyond memory-sharing if the databases are unaware of the cached content.

Second, we consider the same setting under a more restricted model where the databases know the user cache content partially. The user receives an uncoded fraction r from each of the K stored messages, with the $\frac{r}{N}$ fraction of it coming from the nth database. The side information obtained from the nth database is known by the nth database and is unknown by the remaining databases. We investigate the optimal normalized download cost $D^*(r)$, and develop converses and achievability schemes for $D^*(r)$. The largest additive gap between our achievability and the converse bounds is $\frac{5}{32}$ for this case. We observe that the achievable download cost here is larger than that in the previous case due to the partial knowledge of the databases regarding the cache content.

Third, we consider the problem of PIR with private side information (PSI) when the cache content is partially known by the databases. Here, a cache-enabled user of cache-size M possesses side information in the form of full messages that are partially known by the databases. The user wishes to download a desired message privately while keeping the identities of the side information messages that the user did not prefetch from a database private against that database. We characterize the exact capacity of PIR with PSI under partially known PSI condition. We show that the capacity of PIR with partially known PSI is the same as the capacity of PIR with fully unknown PSI.

Fourth, we consider PIR with PSI under storage constraints where a cache-enabled user of cache-size S possesses side information in the form M messages that are unknown to the databases, where $M > S$. We address the problem of which uncoded parts of M messages the user should keep in its constrained cache of size S

in order to minimize the download cost during PIR subject to PSI. We characterize the exact capacity of this PIR-PSI problem under the storage constraint S. We show that a uniform caching scheme which caches equal amounts from all available M messages achieves the lowest normalized download cost.

Fifth, we consider the PIR problem from decentralized uncoded caching databases. Here, the contents of the databases are not fixed a priori, and we design the probability distribution adopted by each database in the decentralized caching phase in order to minimize the expected normalized download cost in the retrieval phase. We characterize the exact capacity of this problem, and show that uniform and random caching results in the lowest normalized download cost.

Next, we focus on security of communication by designing practical coding schemes to achieve the information-theoretically achievable random-coding based secrecy rates. By applying two recently developed techniques for polar codes, namely, universal polar coding and polar coding for asymmetric channels, we propose a polar coding scheme to achieve the secrecy capacity of the general (non-degraded) wiretap channel. We then apply this coding scheme to achieve the best-known secrecy rates for the multiple access wiretap channel, and the broadcast and interference channels with confidential messages.

PRIVATE INFORMATION RETRIEVAL WITH SIDE INFORMATION AND CODING FOR SECURITY

by

Yi-Peng Wei

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Professor Şennur Ulukuş, Chair/Advisor
Professor Alexander Barg
Professor Behtash Babadi
Professor Adrian Papamarcou
Professor Lawrence C. Washington

Dedication

To my family.

Acknowledgments

First of all, I would like to thank my advisor Prof. Sennur Ulukus, whose door is always open for me. She gives me endless advices and support during my Ph.D. studies. She is a role model of hardworking and perseverance, which now I know these are two most important characteristics for doing high quality research. I am also impressed by her carefulness and pursuit of perfection, which especially reflects on all of our papers and presentations. I also want to thank her for being patient for letting me explore different research topics. Personally, I enjoy her humor and optimism a lot. I can always hear her laughter and obtain positive energy after our meeting. I really appreciate her help; without her, I cannot come to this stage.

Second, I would like to thank my committee members Professors Alexander Barg, Behtash Babadi, Adrian Papamarcou, and Lawrence Washington, for their valuable feedback. I am thankful to all the professors I have interacted with over the past years at UMD. I especially want to thank Professors Patrick Fitzpatrick and Leonid Koralov who disclose the beauty of theoretical math to me.

Third, I am thankful to my lab mates at UMD. I would like to thank Karim Banawan, Pritam Mukherjee, Praneeth Boda, Abdulrahman Baknina, Ajaykrishnan Nageswaran, Berk Gurakan, Ahmed Arafa, Jianwei Xie, Omur Ozel, Melih Bastopcu, Baturalp Buyukates, Brian Kim and Batuhan Arasli. I would like to specially thank to Karim Banawan, Pritam Mukherjee, Praneeth Boda and Abdulrahman Baknina. I have a lot of cooperation with Karim Banawan, which gives me so much fun and huge rewards. I really enjoy the time working with him. I also

feel thankful to Praneeth and Pritam for their kindness and friendship. I cherish the special bonding between me, Karim and Abdulrahman, since we started at the same time and went through this process altogether.

Next, I am thankful to my friends at UMD. I would like to thank Min Ye, Itzhak Tamo, Bo Miao, Chao Gao, Xizheng Wang, Ye Chen, Ziyun Zhu and Yu Jin. Thanks for their accompanies and friendships, which makes my life more colorful. I especially want to thank Min Ye. It is always interesting and pleasant to talk to him, and I am really impressed by his intelligence.

Finally, my deepest gratitude goes to my family who constantly love me without asking payback, and always support me without any conditions. I especially want to thank my grandfather, Tsao-Chi Wei, who emphasized the importance of education to the whole family, my father, Chu-Hsien Wei, who shows me the spirit of perseverance, my mother, Chin-Hsiu Hsu, who loves me unconditionally, my brother, Yi-Hung Wei, who redefines the brotherhood and my wife Hung-Hsin Lin, who is my source of happiness in my life. This dissertation is dedicated to all of them.

Table of Contents

List of Figures ix

List of Tables xi

1 Introduction 1

2 Fundamental Limits of Cache-Aided Private Information Retrieval with Unknown and Uncoded Prefetching 20
- 2.1 Introduction . . . 20
- 2.2 System Model . . . 21
- 2.3 Main Results and Discussions . . . 24
- 2.4 Achievability Proof . . . 33
  - 2.4.1 Motivating Example: The Optimal Tradeoff Curve for $K = 3$ Messages and $N = 2$ Databases . . . 34
    - 2.4.1.1 Caching Ratio $r_1 = \frac{1}{7}$ . . . 34
    - 2.4.1.2 Caching Ratio $r_2 = \frac{1}{3}$ . . . 36
    - 2.4.1.3 Caching Ratio $r = \frac{1}{5}$ . . . 37
  - 2.4.2 Achievable Scheme for the Corner Points for Arbitrary K, N . . . 38
    - 2.4.2.1 Decodability, Privacy, and the Achievable Normalized Download Cost . . . 40
  - 2.4.3 Achievable Scheme for Non-Corner Points for Arbitrary K, N 42
- 2.5 Converse Proof . . . 43
- 2.6 Further Examples . . . 50
  - 2.6.1 $K = 4$ Messages, $N = 2$ Databases . . . 50
  - 2.6.2 $K = 4$ Messages, $N = 3$ Databases . . . 51
  - 2.6.3 $K = 5$, $K = 10$ and $K = 100$ Messages, $N = 2$ Databases . . . 51
  - 2.6.4 $K = 5$, $K = 10$ and $K = 100$ Messages, $N = 3$ Databases . . . 52
- 2.7 Gap Analysis . . . 53
- 2.8 Conclusion . . . 68

3 Cache-Aided Private Information Retrieval with Partially Known Uncoded Prefetching: Fundamental Limits 70
3.1 Introduction . . . 70
3.2 System Model . . . 71
3.3 Main Results . . . 74
3.4 Achievable Scheme . . . 81
3.4.1 Motivating Example: $K = 3$ Messages and $N = 2$ Databases . 81
3.4.1.1 Caching Ratio $r_1 = \frac{1}{4}$ . . . 81
3.4.1.2 Caching Ratio $r_2 = \frac{1}{2}$ . . . 83
3.4.1.3 Caching Ratio $r = \frac{1}{3}$ . . . 83
3.4.2 Achievable Scheme . . . 84
3.4.2.1 Achievable Scheme for the Caching Ratio r_s . . . 84
3.4.2.2 Achievable Scheme for the Caching Ratios not Equal to r_s . . . 87
3.5 Converse Proof . . . 88
3.6 Further Examples . . . 93
3.6.1 $K = 4$ Messages, $N = 2$ Databases . . . 93
3.6.2 $K = 5$, $K = 10$ and $K = 100$ Messages, $N = 2$ Databases . . . 93
3.7 Gap Analysis . . . 94
3.8 Comparisons with Other Cache-Aided PIR Models . . . 97
3.9 Conclusion . . . 101
3.10 Appendix . . . 102

4 The Capacity of Private Information Retrieval with Partially Known Private Side Information 108
4.1 Introduction . . . 108
4.2 System Model . . . 110
4.3 Main Results . . . 113
4.4 Converse Proof . . . 115
4.5 Achievability Proof . . . 121
4.5.1 Motivating Examples . . . 124
4.5.1.1 $N = 2$ Databases, $K = 4$ Messages, and $M = 2$ Cached Messages . . . 124
4.5.1.2 $N = 2$ Databases, $K = 5$ Messages, and $M = 2$ Cached Messages . . . 125
4.5.2 General Achievable Scheme for $\frac{M}{N} \in \mathbb{N}$ . . . 126
4.5.3 Normalized Download Cost . . . 128
4.6 Conclusion . . . 131

5 The Capacity of Private Information Retrieval with Private Side Information Under Storage Constraints 132
5.1 Introduction . . . 132
5.2 System Model . . . 133
5.3 Main Results and Discussions . . . 137
5.4 Converse Proof . . . 142

5.5 Achievability Proof . . . 147
5.5.1 Motivating Examples . . . 148
5.5.1.1 $N = 2$ Databases, $K = 5$ Messages, $M = 2$ Accessed Messages, and $S = 1$ with Uniform Caching . . . 148
5.5.1.2 $N = 2$ Databases, $K = 5$ Messages, $S = 1$, $M = 3$ with $r_1 = \frac{1}{2}$, and $r_2 = r_3 = \frac{1}{4}$ . . . 149
5.5.2 General Achievable Scheme . . . 152
5.6 Conclusion . . . 153

6 The Capacity of Private Information Retrieval from Decentralized Uncoded Caching Databases 154
6.1 Introduction . . . 154
6.2 System Model . . . 155
6.3 Main Results and Discussions . . . 160
6.3.1 Motivating Example: $K = 3$ and $N = 2$ . . . 160
6.3.1.1 Achievability Scheme . . . 161
6.3.1.2 Converse Proof . . . 163
6.3.2 Further Examples and Numerical Results . . . 166
6.3.3 Remarks . . . 167
6.4 Achievability Scheme . . . 170
6.5 Converse Proof . . . 172
6.6 Conclusion . . . 176

7 Polar Coding for the General Wiretap Channel with Extensions to Multiuser Scenarios 177
7.1 Introduction . . . 177
7.2 System Model . . . 178
7.2.1 Wiretap Channel Model . . . 178
7.2.2 Multiple Access Wiretap Channel . . . 179
7.2.3 Broadcast Channel With Confidential Messages . . . 180
7.2.4 Interference Channel With Confidential Messages . . . 180
7.3 Existing Polar Coding Techniques . . . 181
7.3.1 Polar Codes for Asymmetric Channels . . . 181
7.3.2 Universal Polar Coding . . . 185
7.3.3 Polar Coding for MAC Based on Monotone Chain Rules . . . 187
7.4 Polar Coding for the General Wiretap Channel . . . 188
7.4.1 The Scheme . . . 189
7.4.2 Reliability . . . 194
7.4.3 Equivocation Calculation . . . 194
7.5 Polar Coding for the Multiple Access Wiretap Channel . . . 197
7.5.1 The Scheme . . . 197
7.5.2 Equivocation Calculation . . . 200
7.6 Polar Coding for the Broadcast Channel with Confidential Messages . 202
7.6.1 Polar Coding for the Binning Region . . . 202
7.6.2 The Scheme . . . 204

7.6.3 Reliability 208
7.6.4 Equivocation Calculation 209
7.7 Polar Coding for the Interference Channel with Confidential Messages 211
7.7.1 The Scheme 211
7.7.2 Equivocation Calculation 214
7.8 Conclusion 215

8 Conclusions 217

Bibliography 221

List of Figures

2.1 Cache-aided PIR with unknown and uncoded prefetching for $N = 3$, $K = 4$ and $r = \frac{1}{4}$. 23
2.2 Comparison between the optimal download cost for known prefetching and the achievable download cost for unknown prefetching in (2.13) for $K = 5$ and $N = 2$. 26
2.3 Inner and outer bounds for $K = 4$ and $N = 2$. For the (x, y) points in this figure, x denotes the caching ratio r and y denotes the normalized download cost $\frac{D}{L}$. 32
2.4 Optimal download cost caching ratio tradeoff for the case of $K = 3$ messages. 33
2.5 Inner and outer bounds for $K = 4$ and $N = 3$. 54
2.6 Inner and outer bounds for $K = 5$ and $N = 2$. 55
2.7 Inner and outer bounds for $K = 10$ and $N = 2$. 56
2.8 Inner and outer bounds for $K = 100$ and $N = 2$. 57
2.9 Inner and outer bounds for $K = 5$ and $N = 3$. 58
2.10 Inner and outer bounds for $K = 10$ and $N = 3$. 59
2.11 Inner and outer bounds for $K = 100$ and $N = 3$. 60
2.12 Outer bounds for $N = 2$, $K = 4$, $K = 5$ and $K = 6$. 61

3.1 Cache-aided PIR with partially known and uncoded prefetching for $N = 2$, $K = 4$ and $r = \frac{1}{4}$. 71
3.2 Inner and outer bounds for $K = 4$, $N = 2$. 80
3.3 Inner and outer bounds for $K = 5$, $N = 2$. 95
3.4 Inner and outer bounds for $K = 10$, $N = 2$. 96
3.5 Inner and outer bounds for $K = 100$, $N = 2$. 97
3.6 Outer bounds for $N = 2$, $K = 3$, $K = 4$ and $K = 5$. 98
3.7 Outer bounds for $N = 2$, $K = 12$ for different cache-aided PIR models. 99
3.8 Comparison between this work and Chapter 2 for $N = 3$ and $K = 6$. 100

4.1 PIR with partially known PSI for $N = 2$, $K = 4$ and $M = 2$. 109

5.1 PIR-PSI under a storage constraint. Here $N = 3$, $K = 5$, $S = 1$, and $M = 3$. 138

5.2 Achievable scheme: $K = 5$, $S = 1$, and $M = 3$ with $r_1 = \frac{1}{2}$, and $r_2 = r_3 = \frac{1}{4}$. 150

6.1 PIR from decentralized caching databases with $K = 3$, $N = 2$, and $\mu = \frac{1}{3}$. 159
6.2 PIR from different number of available databases in the retrieval phase with $K = 10$ and $\mu = \frac{1}{2}$. 167
6.3 PIR from $N = 5$ databases with different storage constraint μ with $K = 10$. 168
6.4 PIR from centralized caching databases and decentralized caching databases. 170

7.1 Chaining construction for the general wiretap channel. 191
7.2 General MAC regions. 198
7.3 Chaining construction for the MAC-WTC for user 1. 200
7.4 Chaining construction for the second user to achieve the binning region in a broadcast channel. 203
7.5 Chaining construction for the BC-CM for user 1. 207
7.6 Chaining construction for the BC-CM for user 2. 208
7.7 Chaining construction for the IC-CM for user 1. 213

List of Tables

2.1 Query table for $K = 3$, $N = 2$ and $r_1 = \frac{1}{7}$ 36
2.2 Query table for $K = 3$, $N = 2$ and $r_2 = \frac{1}{3}$ 37
2.3 Query table for $K = 3$, $N = 2$ and $r = \frac{1}{5}$ 38
2.4 Query table for $K = 4$, $N = 2$ and $r_1 = \frac{1}{15}$ 50
2.5 Query table for $K = 4$, $N = 2$ and $r_2 = \frac{1}{5}$ 51
2.6 Query table for $K = 4$, $N = 2$ and $r_3 = \frac{1}{3}$ 51
2.7 Query table for $K = 4$, $N = 3$ and $r_1 = \frac{1}{40}$ 52
2.8 Query table for $K = 4$, $N = 3$ and $r_2 = \frac{2}{17}$ 53
2.9 Query table for $K = 4$, $N = 3$ and $r_3 = \frac{1}{4}$ 53

3.1 Query table for $K = 3$, $N = 2$, $r_1 = \frac{1}{4}$. 82
3.2 Query table for $K = 3$, $N = 2$, $r_2 = \frac{1}{2}$. 83
3.3 Query table for $K = 3$, $N = 2$, $r = \frac{1}{3}$. 84
3.4 Query table for $K = 4$, $N = 2$ and $r_1 = \frac{1}{8}$. 93
3.5 Query table for $K = 4$, $N = 2$, $r_2 = \frac{1}{3}$. 94
3.6 Query table for $K = 4$, $N = 2$, $r_3 = \frac{1}{2}$. 94

4.1 Query table for $K = 4$, $N = 2$, $M = 2$. 125
4.2 Query table for $K = 5$, $N = 2$, $M = 2$. 127

CHAPTER 1

Introduction

In today's communication networks, the end-users are equipped with large memories, and the data transmitted in the network has shifted from real-time generated data like voice to pre-generated content like movies. These two factors together have enabled caching techniques, which store data in user cache a priori in order to reduce the peak-hour network traffic load. In the meanwhile, privacy has become an important consideration for users, who wish to download data from publicly accessable databases as privately and as efficiently as possible. This is studied under the subject of private information retrieval (PIR). From an information-theoretic point of view, cached data can be regarded as a form of *side information*. In the first part of this thesis, we study privacy of users in caching networks, by studying the PIR problem in the presence of side information (SI).

The problem of PIR was introduced by Chor et al. [1] to investigate the privacy of users while downloading data from public databases. The PIR problem has become a major research area within the computer science literature subsequently, see e.g., [2–5]. In the classical form of the problem [1], a user requests to download

a message (or a file) from N non-communicating and replicated databases where each database contains the same set of K messages such that no database can distinguish individually which message has been retrieved. The user performs this task by preparing N queries, one for each database, such that the queries do not reveal the user's interest in the desired message. Each database responds truthfully to the received query by an answer string. The user reconstructs the desired message from the collected answer strings. A naive PIR scheme is to download all of the K messages from a database. However, this trivial PIR scheme is quite inefficient from the retrieval rate perspective, which is defined as the number of desired bits per bit of downloaded data. Consequently, the aim of the PIR problem is to retrieve the desired message correctly by downloading as few bits as possible from the N databases under the privacy constraint.

Recently, the PIR problem is revisited by information theorists with early examples [6–11]. In the information-theoretic re-formulation of the problem, the length of the message L is assumed to be arbitrarily large to conform with the traditional Shannon-theoretic arguments, and the upload cost is neglected as it does not scale with the message length. This formulation provides an absolute privacy guarantee by ensuring statistical independence between the queries and the identity of the desired message. In the influential paper by Sun and Jafar [12], the notion of PIR capacity is introduced, which is the supremum of PIR rates over all achievable retrieval schemes. In [12], the authors characterize the capacity of classical PIR. In [12], a greedy iterative algorithm is proposed for the achievability scheme and an induction based converse is provided to obtain an exact result. The

achievable scheme is based on an interesting correspondence between PIR and blind interference alignment [13] as observed earlier in [14]. Sun and Jafar show that in order to privately retrieve a message, the optimal total downloaded bits normalized with the message size is $\frac{D}{L} = 1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}}$. Consequently, the PIR capacity is the reciprocal of this optimal normalized download cost, i.e., $C = (1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}})^{-1}$.

Following the work of [12], the fundamental limits of many interesting variants of the classical PIR problem have been considered [15–52], such as: PIR with T colluding databases (TPIR) [15, 23], where any T of N databases might collude; robust PIR (RPIR) [15,26,30], where some databases may fail to respond; symmetric PIR (SPIR) [16], which adds the constraint that the user should only learn the desired message; MDS-coded PIR (CPIR) [17], where the contents of the databases are not replicated, but coded via an MDS code; PIR under message size constraint L (LPIR) [18]; multi-round PIR, where the queries are permitted to be a function of the answer strings collected in previous rounds [20]; multi-message PIR (MPIR) [24], where the user wishes to jointly retrieve P messages; PIR from Byzantine databases (BPIR), where B databases are outdated or worse adversarial [27].

In Chapter 2, we study cache-aided PIR with unknown and uncoded prefetching. Recently, reference [28] has considered cache-aided PIR, where the user has local cache memory of size rKL bits and it can store any function of the K messages subject to this memory size constraint. With the assumption that the cache content is known by all the N databases, reference [28] characterizes the optimal download cost. The achievability scheme is based on memory-sharing[1] and the con-

[1]Memory-sharing, introduced in [28], is an achievability concept similar to the classical achiev-

verse bound is obtained with the aid of Han's inequality. To privately retrieve a message, the optimal total downloaded bits normalized with the message size is $\frac{D(r)}{L} = (1-r)(1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}})$. The result is quite pessimistic as it implies that the cached bits cannot be used as side information within the retrieval scheme and the user must download the uncached portion of the file (the remaining $L(1-r)$ bits) using the original PIR scheme in [12]. The reason is that the databases are fully knowledgeable about the cached bits and can infer which message is desired if the user exploits these cached bits as side information in any form.

The above discussion motivates us to investigate the other extreme where the databases are fully unaware of the cache content, i.e., when the prefetched bits are unknown to all of the N databases (in contrast to having the cache content as public knowledge at all the N databases as in [28]). In this case, the user can leverage the cached bits as side information without sacrificing the privacy constraint as the databases are unaware of the cached bits. This poses an interesting question: What is the optimal way to exploit the cached bits as side information in order to minimize the normalized download cost, and what is the corresponding gain beyond memory-sharing if any? The assumption of unknown prefetching can be interpreted in practice as the prefetching phase is performed via an external database which does

ability concept of time-sharing. Reference [28, Lemma 1] first shows that the download cost $D(S)$ is a convex function of the cache memory size S. That is, for two different cache sizes S_1 and S_2, we have $D(\alpha S_1 + (1-\alpha)S_2) \leq \alpha D(S_1) + (1-\alpha)D(S_2)$. Reference [28, Lemma 1] shows this by dividing the messages into two independent parts of sizes αL and $(1-\alpha)L$ and correspondingly scaling the cache memory sizes with α and $(1-\alpha)$, and applying two different PIR schemes to the two independent parts of the message. This implies that memory-sharing between zero caching (and requiring the download cost in [12]) and caching all the messages (and requiring zero download cost), a normalized download cost of $(1-r)\left(1+\frac{1}{N}+\cdots,\frac{1}{N^{K-1}}\right)$ is achievable with a caching ratio of r, which is linear in r.

not participate in the retrieval phase. We further assume that the cache content is uncoded, which is a common assumption in the caching literature [53–55].

In this chapter, we consider PIR with unknown and uncoded prefetching, i.e., we assume that the cache content is unknown to all databases, and the cache supports only direct (uncoded) portions of all messages (smaller subfiles). We aim to characterize the optimal tradeoff between the normalized download cost $\frac{D(r)}{L}$ and the caching ratio r. For the outer bound, we explicitly determine the achievable download rates for specific $K+1$ caching ratios. Download rates for any other caching ratio can be achieved by proper memory-sharing between the nearest two explicit points. This implies that the outer bound is a piece-wise linear curve which consists of K line segments. For the inner bound, we extend the techniques of [12,28] to obtain a piece-wise linear curve which also consists of K line segments. We show that the inner and the outer bounds match exactly at three of the K line segments for any number of messages K. This means that we characterize the optimal tradeoff for the very low ($r \leq \frac{1}{1+N+N^2+\cdots+N^{K-1}}$) and the very high ($r \geq \frac{K-2}{(N+1)K+N^2-2N-2}$) caching ratios. As a direct corollary, we fully characterize the optimal download cost caching ratio tradeoff for $K=3$ messages. For general K, N and r, we show that for fixed N, the outer bound monotonically increases as K increases. To characterize the worst-case gap between the inner and the outer bounds, we determine the asymptotic achievability bound as $K \to \infty$ for fixed N, r. We then show that the asymptotic gap monotonically decreases in N. Therefore, the worst-case gap happens at $N=2$ and $K \to \infty$. By maximizing this over r, we show that the largest gap between the achievability and the converse bounds is $\frac{1}{6}$. Our results

show the benefits of the cached content when the databases are unaware of it over the scenario in [28] where the databases are fully aware of the cached content.

In Chapter 3, we study cache-aided PIR with partially known uncoded prefetching, which is closely related to our formulation in Chapter 2. In Chapter 2, the databases are assumed to be completely unaware of the side information. However, this may be practically challenging to implement. In Chapter 3, we consider a more natural model which uses the same set of databases for both prefetching and retrieval phases. Therefore, different from Chapter 2, here each database gains partial knowledge about the side information, that is the part it provides during the prefetching phase. Our aim is to determine if there is a rate loss due to this partial knowledge with respect to the fully unknown case in Chapter 2, and characterize this rate loss as a function of K, N and r.

We consider the PIR problem with a two-phase scheme, namely, prefetching phase and retrieval phase. In the prefetching phase, the user caches an uncoded $\frac{r}{N}$ fraction of each message from the nth database. The nth database is aware of these $\frac{KLr}{N}$ bit side information, while it has no knowledge about the cached bits from the other $(N-1)$ databases. We aim at characterizing the optimal tradeoff between the normalized download cost $\frac{D(r)}{L}$ and the caching ratio r. For the outer bound, we explicitly determine the achievable download rates for specific $K+1$ caching ratios. Download rates for any other caching ratio can be achieved by memory-sharing between the nearest explicit points. Hence, the outer bound is a piece-wise linear curve which consists of K line segments. For the inner bound, we extend the techniques of [12] and Chapter 2 to obtain a piece-wise linear curve which also

consists of K line segments. We show that the inner and the outer bounds match exactly at three line segments for any K. Consequently, we characterize the optimal tradeoff for the very low ($r \leq \frac{1}{N^{K-1}}$) and the very high ($r \geq \frac{K-2}{N^2-3N+KN}$) caching ratios. As a direct corollary, we fully characterize the optimal download cost caching ratio tradeoff for $K = 3$ messages. For general K, N and r, we show that the worst-case additive gap between the inner and the outer bounds is $\frac{5}{32}$.

In Chapter 4, we study PIR with partially known private side information (PSI). This chapter is most closely related to [28, 32, 33, 50] and Chapters 2 and 3 here. These works investigate the PIR problem when the user (retriever) possesses some form of side information about the contents of the databases. However, the models of [28, 32, 33, 50] and Chapters 2 and 3 differ in three important aspects, namely, 1) the structure of the side information, 2) the presence or absence of privacy constraints on the side information, and 3) the databases' awareness of the side information at its initial acquisition. Here, structure of the side information refers to whether the side information is in the form of full messages or parts of messages or whether messages are mixed through functions (coded/uncoded); privacy of the side information refers to whether the user further aims to keep the side information private from the databases; and databases' awareness of the side information refers to whether the databases knew the initially prefetched side information.

Specifically, reference [28] studies the capacity of the cache-aided PIR where the user caches rLK bits in the form of any arbitrary function of the K messages, where L is the message size, and $0 \leq r \leq 1$ is the caching ratio. Reference [32] considers the case where $\lfloor rK \rfloor$ full messages are cached, and Chapters 2 and 3

consider the case where a random r fraction of the symbols of each of K messages is cached. Reference [28] assumes that the cache content is perfectly known by all the databases, and hence there is no need to protect the privacy of the cached content. Reference [28] motivates [32] and Chapter 2 to study the other extreme when the databases are completely unaware of the side information at its initial acquisition.

Reference [32] further introduces another model where the cached content (in the form of full messages) which is unknown to the databases at the time of initial prefetching, must remain unknown throughout the PIR, i.e., the queries of the user should not leak any information about the cached content to the databases. The exact capacity for this problem is settled in [33] to be $C = \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-M-1}}\right)^{-1}$. The optimal achievable scheme in this case starts from the traditional achievable scheme without side information in [12] and reduces the download cost by utilizing the reconstruction property of MDS codes. Reference [32] also considers the case of no privacy constraint on the cached content. The exact capacity for this problem is settled in [50] to be $C = \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{\lceil \frac{K}{M+1} \rceil - 1}}\right)^{-1}$. We note that there is no privacy constraint on the cached content in Chapters 2 and 3 here.

In Chapter 4, we take a deeper look at the issue of *awareness* or otherwise *unawareness* of the databases about the cached content *at its initial acquisition*. We first note that it is practically challenging to make the side information completely unknown to the databases at its initial acquisition as assumed in [32, 33, 50] and Chapter 2. One way to do this could be to employ one of the databases for prefetching the side information and exclude it from the retrieval process. Therefore, for the remaining $N - 1$ databases, the side information is completely unknown. This

solution is strictly sub-optimal as the capacity expression in [33] (shown as C in the previous paragraph) is monotonically decreasing in N. We also note that the other extreme of the problem, where the databases are fully aware of the cached content [28], is discouraging as the user cannot benefit from the cached side information. Therefore, a natural model is to use the databases for both prefetching and retrieval phases, such that the databases gain partial knowledge about the side information available to the user, which makes it possible for the user to exploit the remaining side information that is unknown to each individual database to reduce the download cost during the retrieval process. This poses the following questions: Can we propose efficient joint prefetching-retrieval strategies that exploit the limited knowledge of each database to drive down the download cost? How much is the loss from the fully unknown case in [32, 33]?

In this chapter, we investigate the PIR problem when the user and the databases engage in a two-phase scheme. In the prefetching phase, the user caches m_n full messages out of the K messages from the nth database under a total cache memory size constraint $\sum_{n=1}^{N} m_n \leq M$. Hence, each database has a *partial knowledge* about the side information possessed by the user, namely, the part of the side information that this database has provided during the prefetching phase. In the retrieval phase, the user wants to retrieve a message (which is not present in its memory) without leaking any information to any individual database about the desired message or the remaining side information messages that are unknown to each database. The goal of this work is to design a joint prefetching-retrieval scheme that minimizes the download cost in the retrieval phase.

To that end, we first derive a general lower bound for the normalized download cost that is independent of the prefetching strategy. Then, we prove that this bound is attainable using two achievable schemes. The first achievable scheme, which is proposed in [33] for completely unknown side information, is a valid achievable scheme for our problem with partially known side information for any prefetching strategy.[2] We provide a second achievable scheme for the case of uniform prefetching, i.e., $m_n = \frac{M}{N} \in \mathbb{N}$, which requires smaller sub-packetization and smaller field size for realizing MDS codes. While the first achievable scheme [33] requires a message size of $L = N^K$, the second achievable scheme proposed here requires a message size of $L = N^{K-\frac{M}{N}}$, which scales down the message size by an exponential factor $N^{\frac{M}{N}}$, which in turn simplifies the achievable scheme and minimizes the total number of downloaded bits without sacrificing from the capacity. We prove that the exact capacity of this problem is $C = \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-M-1}}\right)^{-1}$. Surprisingly, this is the same capacity expression for the PIR problem when the databases are completely unaware of the side information possessed by the user as found in [33] recently. Therefore, our result implies that there is no loss in the capacity if the same databases are employed in both prefetching and retrieval phases.

In Chapter 5, we consider the PIR problem with PSI for a cache-enabled user under a cache storage size constraint. The goal of the PIR-PSI problem is to devise the most efficient retrieval scheme under the joint desired message and side information privacy constraints. The system operates in two phases, a prefetching phase and a retrieval phase. In the prefetching phase, the user can access M messages,

[2]We thank Dr. Hua Sun for pointing this out.

and has a local cache storage of S messages (SL symbols), where $S \leq M$. For each of these M messages, the user caches the first Lr_i symbols out of the total L symbols for $i = 1, \ldots, M$. The caching scheme is subject to a memory size constraint, i.e., $\sum_{i=1}^{M} r_i = S$. Note that in [36] and Chapters 2 and 3, for each message, the user randomly chooses Lr symbols out of the total L symbols to cache. In [36] and Chapters 2 and 3, to reliably reconstruct the desired message, the user should record the indices of the cached symbols within each message. In contrast, here, we consider the case where the user caches the first Lr_i symbols of each message instead of random Lr_i symbols; this saves the user extra storage overhead. The databases are aware of the caching scheme, but do not know the identities of the cached messages, i.e., the databases know M and r_i for $i = 1, \ldots, M$, but do not know the identities of the cached messages. In the retrieval phase, the user wishes to jointly keep the identities of the cached messages and the desired message private. We call this model as PIR-PSI under a storage constraint.

For any given caching scheme, i.e., for given M and $(r_1, r_2, \ldots, r_M)$, we characterize the optimal normalized download cost to be $D^* = 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1-M}} + \frac{1-r_M}{N^{K-M}} + \frac{1-r_{M-1}}{N^{K-M+1}} + \cdots + \frac{1-r_1}{N^{K-1}}$, where without loss of generality $r_1 \geq r_2 \geq \cdots \geq r_M$. Based on this capacity result, we prove two important facts: First, for a fixed memory size S and fixed number of accessible messages M, uniform caching achieves the lowest normalized download cost, where uniform caching means $r_i = \frac{S}{M}$, for $i = 1, \ldots, M$. Second, for a fixed memory size S, among all the $K - \lceil S \rceil + 1$ uniform caching schemes, the uniform caching scheme which caches K messages achieves the lowest normalized download cost. That is, in order to optimally utilize the limited

user cache memory, if the user has access to M files, it should keep SL/M bits (equal amounts) from each message in its cache memory; and second, if possible, the user should aim to have access to all K messages, i.e., $M = K$ yields the lowest download cost.

In Chapter 6, we study PIR from decentralized uncoded caching databases. Currently, most of the previous works on the PIR problem consider the case where the contents of the databases are fixed a priori in an uncontrollable manner, and a vast majority of them consider the case of replicated databases where each database stores the same set of K files. In this chapter, we modify these two assumptions.

Coded caching refers to the problem of placing files in users' local storage caches ahead of time properly and designing efficient delivery schemes at the time of specific user requests in such a way to minimize the traffic during the delivery phase. In the original setup [53], a server with K files connects to N users through an error-free shared link, where each user has a local memory which can store up to M files. The system operates in two phases. In the placement phase, the server places the files into each user's local memory. In the delivery phase, each user requests a file from the server, and the server aims to satisfy all the requests with the lowest traffic load. If the set of users in the two phases are identical, the server can arrange the content in each user's local memory in an optimized manner, which is called *centralized coded caching*. Reference [53] proposes a symmetric batch caching scheme, which is shown to be optimal for the case of centralized uncoded placement in [55]. If the set of users in the two phases varies, the server cannot arrange the files in user caches in a centralized manner. Instead, the server treats each user identically

and independently which is called *decentralized coded caching* [54]. Reference [54] proposes a uniform and random caching scheme, which is shown to be optimal for the case of decentralized uncoded placement in [55]. Many interesting variants of coded caching problem have been investigated in [56–70].

The references that are most closely related to this chapter are [38, 43]. References [38, 43] formulate a new type of PIR problem where the content of each database is not fixed a priori, but can be optimized to minimize the download cost. These papers bring PIR and coded caching problems together in a practically relevant and theoretically interesting manner. In their problem setting, there is a data center (server) containing all the K files where each file is of size L bits, and the system operates in two phases. In the caching phase, there are N databases in the system with a common storage size constraint μ, i.e., each database can at most store μKL bits, $\frac{1}{N} \leq \mu \leq 1$. In the retrieval phase, a user accesses the N databases, and wishes to download a desired file privately. They consider the problem of optimally storing content from the data center to the databases in the caching phase in such a way that the normalized download cost during the retrieval phase is minimized. They focus on the *centralized uncoded* caching case, i.e., the set of users in the two phases are identical so that the data center can assign the files to each database in a *centralized* manner, and caching is *uncoded* in that each database stores a subset of the bits from the data center (no coding), i.e., each database stores μKL bits out of the total KL bits. Surprisingly, they show that the symmetric batch caching scheme proposed in [53] results in the lowest normalized download cost in the retrieval phase.

In this chapter, we consider the PIR problem from *decentralized uncoded* caching databases. In our problem setting, the system also operates in two phases as in [38, 43]. However, the set of databases active in the two phases are different, and we do not know in advance which databases the user can access in the retrieval phase. Therefore, we consider a *decentralized* setting for the caching phase, i.e., the data center treats each database identically and independently, or equivalently, each database chooses a subset of bits to store independently according to the same probability distribution. Here, we aim at designing the optimal probability distribution in the caching phase and PIR scheme in the retrieval phase such that the normalized download cost in the retrieval phase is minimized. Another main difference between our work and references [38, 43] is that, in the caching phase, references [38, 43] require that the N databases altogether can reconstruct the entire K files, i.e., when the user connects to the N databases, their collective content is equivalent to the content in the data center, so the user can download any desired file. While this can be guaranteed in the centralized setting, in the decentralized setting, where cache placement is probabilistic, we cannot guarantee that any given N databases contain all the bits that exist in the data center. Thus, in order to formulate a meaningful PIR problem, we allow the user access the data center as well as the databases in the retrieval phase. Finally, we remark about the relationship of the problem considered in this chapter to another sub-branch of PIR literature that considers caching as in [32, 33, 50] and in Chapters 2 and 3. There the user itself has a cache memory where it stores a subset of the bits available in the databases. That problem is unrelated to the setting considered in this chapter even though it is also referred to

as PIR with caching; in essence, it is PIR with side information.

In this chapter, for PIR from decentralized caching databases, we show that uniform and random caching scheme, originally proposed in [54] for decentralized coded caching, results in the lowest expected normalized download cost in the retrieval phase. For the achievability, we apply the PIR scheme in [12] successively for all resulting subfile parts. For the converse, we first apply the lower bound derived in [43], which replaces the random variables for queries and answering strings by the content of the distributed databases in a novel manner extending the lower bounding techniques in [12, Lemma 5 and Lemma 6]. To compare different probability distributions in the caching phase, we focus on the marginal distributions on each separate bit. Then, by using the nature of decentralization and uncoded caching, we further lower bound the normalized download cost. Finally, we show the matching converse for the expected normalized download cost to be $\frac{D}{L} = \sum_{n=1}^{N+1} \binom{N}{n-1} \mu^{n-1} (1-\mu)^{N+1-n} \left(1 + \frac{1}{n} + \cdots + \frac{1}{n^{K-1}}\right)$, which yields an exact capacity result for the problem.

Next, in the second part of this thesis, we study the wiretap channel with a focus on the design of practical coding schemes to achieve the information-theoretically achievable random-coding based secrecy rates. The wiretap channel was first introduced by Wyner [71], in which a legitimate transmitter (Alice) wishes to send messages to a legitimate receiver (Bob) secretly in the presence of an eavesdropper (Eve). Wyner [71] characterized the capacity equivocation region for the degraded wiretap channel, in which the received signal at Eve is a degraded version of the received signal at Bob. Later, Csiszár and Körner [72] characterized the capacity

equivocation region for general, not necessarily degraded, wiretap channels. These works are based on information-theoretic random-coding schemes.

Polar coding, invented by Arıkan [73], is the first code that provably achieves the capacity of the binary-input discrete symmetric output channels (B-DMC). The idea of polar coding has been extended to lossless source coding [74], lossy source coding [75], and to multi-user scenarios, such as, multiple access channel [76–78], broadcast channel [79, 80], interference channel [81], and Slepian-Wolf coding problem [82].

On a B-DMC, polarization results in two kinds of sub-channels [73]. The first kind is good sub-channels. The capacity for these sub-channels approaches 1 bit per channel use. The second kind is bad sub-channels. The channel output for these sub-channels is independent of the channel input; therefore the capacity for these sub-channels approaches 0. In particular, if a B-DMC A is degraded with respect to a B-DMC B, then the good sub-channels of A must be a subset of the good sub-channels of B [83]. We call this the *subset property*.

Polar coding schemes for *degraded* wiretap channels with *symmetric* main and eavesdropper channels are developed using the subset property in [84–87]. For degraded wiretap channels, the good sub-channels of Eve is a subset of the good sub-channels of Bob. The polar coding scheme is designed to transmit the confusion messages (random bits) on the sub-channels simultaneously good for Bob and Eve, and to transmit the secret messages on the sub-channels only good for Bob. However, for non-degraded wiretap channels, the subset property no longer holds [88–92], i.e., the good sub-channels of Eve is not necessary a subset of the good sub-channels

of Bob. Moreover, the secrecy capacity achieving input distribution is not necessarily a uniform distribution. Therefore, the polar coding schemes in [84–87] cannot directly extend to the non-degraded wiretap channel.

By applying two recently developed techniques for polar codes, we can achieve the secrecy capacity of the general wiretap channel. The first technique is *universal polar codes* [91,92]. Universal polar coding allows us to align the good sub-channels of Bob and Eve together. Therefore, we can artificially construct the subset property for the non-degraded wiretap channel. Then, Alice transmits the random bits on the sub-channels simultaneously good for Bob and Eve, and the secret message on the sub-channels only good for Bob. The second technique is *polar coding for asymmetric models* [93], which allows us to deal with the non-uniform input distribution. Different from B-DMC, polarization for asymmetric channels results in three different kinds of sub-channels.

Another polar coding scheme for the general wiretap channel is provided in [94], which uses a concatenated code consisting of two polar codes. The inner layer ensures that the transmitted message can be reliably decoded by Bob, and the outer layer guarantees that the message is kept secret from Eve. Our work jointly handles these two goals in one shot. Hence, the decoding error probability of our scheme is approximately $O(2^{-n^{1/2}})$, whereas it is $O(\sqrt{n}2^{-n^{1/4}})$ in [94]. Although the scheme in [94] does not require to share randomness, for practical code construction, there is still no efficient way to characterize the outer index set [94, Sec. III. C.], while our coding scheme can be efficiently constructed by [89].

Next, we extend our coding scheme to several multiuser scenarios: multiple

access wiretap channel (MAC-WTC) [95, 96], broadcast channel with confidential messages (BC-CM) [97], and interference channel with confidential messages (IC-CM) [97]. In the MAC-WTC, two transmitters wish to send independent messages to the legitimate receiver in the presence of an eavesdropper. In the BC-CM[3], the transmitter wishes to send independent messages to two receivers, while keeping the messages secret from the unintended receiver. In the IC-CM, two transmitters wish to send independent messages to their respective receivers, and keep the messages confidential from the other receiver.

To the best of our knowledge, there are no practical coding schemes for these multiuser scenarios. For the MAC-WTC, we achieve the entire dominant face of the best-known achievable region by combining the coding scheme for the general wiretap channel we introduce here with the *monotone chain rule* [82]. For the BC-CM, we introduce a *double chaining* construction to achieve the best-known inner bound. Finally, we extend the coding scheme for the general wiretap channel to the setting of IC-CM.

We acknowledge independent and concurrent papers which present similar results on polar coding for general wiretap channels at the same conference; see [98,99]. Reference [98] generalizes the polar coding scheme for strong secrecy in [100], while in our work, we artificially construct the subset property to extend the polar coding scheme in [84–87]. Interestingly, these two points of view lead to the same

[3]Although the naming of BC-CM is similar to [72], these two channel models are different. In particular, [72] is a "single-user" wiretap channel, in the sense that there is only one message to be secured; it is a generalization of [71] to non-degraded channels, together with the introduction of a common message to be sent (insecurely) to both Bob and Eve. BC-CM [97], on the other hand, has two messages each to be secured from the unintended receiver.

chaining construction method [99]. However, the remaining parts of these three works are different. References [98, 99] mainly deal with broadcast channel with a confidential component [72]. However, we not only achieve the secrecy capacity of [72] but also propose coding schemes to achieve the best-known inner bounds of the multiuser models of MAC-WTC, BC-CM and IC-CM, which require different constructions.

In Chapter 8, we conclude this dissertation.

CHAPTER 2

Fundamental Limits of Cache-Aided Private Information Retrieval with Unknown and Uncoded Prefetching

2.1 Introduction

We consider the problem of private information retrieval (PIR) from N non-colluding and replicated databases when the user is equipped with a cache that holds an uncoded fraction r from each of the K stored messages in the databases. We assume that the databases are unaware of the cache content. We investigate $D^*(r)$ the optimal download cost normalized with the message size as a function of K, N, r. For a fixed K, N, we develop an inner bound (converse bound) for the $D^*(r)$ curve. The inner bound is a piece-wise linear function in r that consists of K line segments. For the achievability, we develop explicit schemes that exploit the cached bits as side information to achieve $K-1$ non-degenerate corner points. These corner points differ in the number of cached bits that are used to generate one side information equation. We obtain an outer bound (achievability) for any caching ratio by memory-sharing between these corner points. Thus, the outer bound is also a piece-wise linear function in r that consists of K line segments. The inner

and the outer bounds match in general for the cases of very low caching ratio and very high caching ratio. As a corollary, we fully characterize the optimal download cost caching ratio tradeoff for $K = 3$. For general K, N, and r, we show that the largest gap between the achievability and the converse bounds is $\frac{1}{6}$. Our results show that the download cost can be reduced beyond memory-sharing if the databases are unaware of the cached content.

2.2 System Model

We consider a classic PIR problem with K independent messages $W_1, \ldots, W_K$. Each message is of size L bits,

$$H(W_1) = \cdots = H(W_K) = L, \tag{2.1}$$

$$H(W_1, \ldots, W_K) = H(W_1) + \cdots + H(W_K). \tag{2.2}$$

There are N non-communicating databases, and each database stores all the K messages, i.e., the messages are coded via $(N, 1)$ repetition code [17]. The user (retriever) has a local cache memory whose content is denoted by a random variable Z. For each message W_k of size L bits, the user randomly and independently caches Lr bits out of the L bits to Z, where $0 \leq r \leq 1$, and r is called the *caching ratio* (See Fig. 2.1). Therefore,

$$H(Z) = KLr. \tag{2.3}$$

The caching ratio r is known to the databases. Since the user caches a subset of the bits from each message, this is called *uncoded prefetching*. We denote the indices of the cached bits by random variable $\mathbb{H}$. For each message W_k, we have

$$H(W_k|Z,\mathbb{H}) = L(1-r). \tag{2.4}$$

Here, different from [28], we consider the case where none of the databases knows the prefetched cache content.

After the uncoded prefetching phase, the user privately generates an index $\theta \in [K]$, where $[K] = \{1,\ldots,K\}$, and wishes to retrieve message W_θ such that no database knows which message is retrieved. Note that during the prefetching phase, the desired message is unknown a priori. Note further that the cached bit indices $\mathbb{H}$ are independent of the message contents and the desired message index θ. Therefore, for random variables θ, $\mathbb{H}$, and $W_1,\ldots,W_K$, we have

$$H\left(\theta,\mathbb{H},W_1,\ldots,W_K\right) = H\left(\theta\right) + H\left(\mathbb{H}\right) + H(W_1) + \cdots + H(W_K). \tag{2.5}$$

Suppose $\theta = k$. The user sends N queries $Q_1^{[k]},\ldots,Q_N^{[k]}$ to the N databases, where $Q_n^{[k]}$ is the query sent to the nth database for message W_k. The queries are generated according to $\mathbb{H}$ and Z, but are independent of the realizations of the uncached messages. Therefore,

$$I(W_1,\ldots,W_K;Q_1^{[k]},\ldots,Q_N^{[k]}|Z,\mathbb{H}) = 0. \tag{2.6}$$

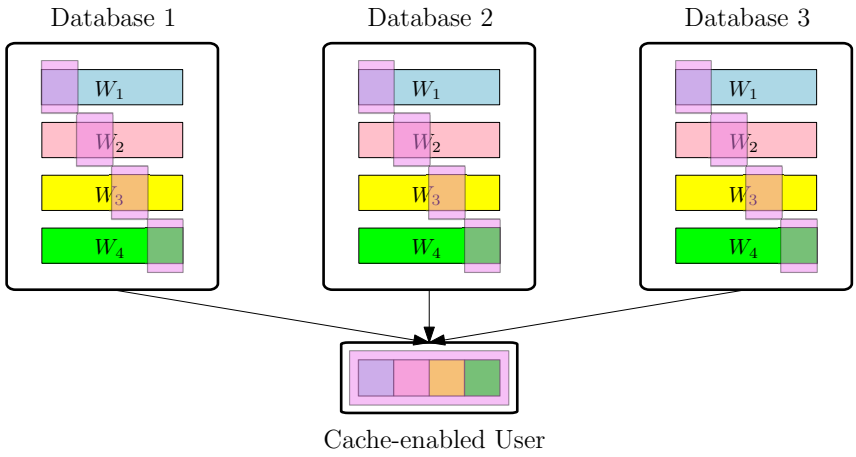

Figure 2.1: Cache-aided PIR with unknown and uncoded prefetching for $N = 3$, $K = 4$ and $r = \frac{1}{4}$.

To ensure that individual databases do not know which message is retrieved, we need to satisfy the following privacy constraint, $\forall n \in [N]$, $\forall k \in [K]$,

$$(Q_n^{[1]}, A_n^{[1]}, W_1, \ldots, W_K) \sim (Q_n^{[k]}, A_n^{[k]}, W_1, \ldots, W_K). \tag{2.7}$$

Upon receiving the query $Q_n^{[k]}$, the nth database replies with an answering string $A_n^{[k]}$, which is a function of $Q_n^{[k]}$ and all the K messages. Therefore, $\forall k \in [K], \forall n \in [N]$,

$$H(A_n^{[k]}|Q_n^{[k]}, W_1, \ldots, W_K) = 0. \tag{2.8}$$

After receiving the answering strings $A_1^{[k]}, \ldots, A_N^{[k]}$ from all the N databases, the user needs to decode the desired message W_k reliably. By using Fano's inequality, we have the following reliability constraint

$$H\left(W_k|Z, \mathbb{H}, Q_1^{[k]}, \ldots, Q_N^{[k]}, A_1^{[k]}, \ldots, A_N^{[k]}\right) = o(L), \tag{2.9}$$

where $o(L)$ denotes a function such that $\frac{o(L)}{L} \to 0$ as $L \to \infty$.

For a fixed N, K, and caching ratio r, a pair $(D(r), L)$ is achievable if there exists a PIR scheme for message of size L bits with unknown and uncoded prefetching satisfying the privacy constraint (2.7) and the reliability constraint (2.9), where $D(r)$ represents the expected number of downloaded bits (over all the queries) from the N databases via the answering strings $A_{1:N}^{[k]}$, i.e.,

$$D(r) = \sum_{n=1}^{N} H\left(A_n^{[k]}\right). \tag{2.10}$$

In this work, we aim to characterize the optimal normalized download cost $D^*(r)$ corresponding to every caching ratio $0 \leq r \leq 1$, where

$$D^*(r) = \inf\left\{\frac{D(r)}{L} : (D(r), L) \text{ is achievable}\right\}, \tag{2.11}$$

which is a function of the caching ratio r.

2.3 Main Results and Discussions

Our first result characterizes an outer bound (achievable rate) for the normalized download cost $D^*(r)$ for general K, N and r.

Theorem 2.1 (Outer bound) *In the cache-aided PIR with uncoded and unknown prefetching, for the caching ratios*

$$r_s = \frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N}, \tag{2.12}$$

where $s \in \{1, 2, \cdots, K-1\}$*, the optimal normalized download cost* $D^*(r_s)$ *is upper bounded by,*

$$D^*(r_s) \leq \bar{D}(r_s) = \frac{\sum_{i=0}^{K-1-s} \binom{K}{s+1+i}(N-1)^i N}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N}. \tag{2.13}$$

Moreover, if $r_s < r < r_{s+1}$*, and* $\alpha \in (0,1)$ *such that* $r = \alpha r_s + (1-\alpha) r_{s+1}$*, then*

$$D^*(r) \leq \bar{D}(r) = \alpha \bar{D}(r_s) + (1-\alpha)\bar{D}(r_{s+1}). \tag{2.14}$$

The proof of Theorem 2.1 can be found in Section 2.4. Theorem 2.1 implies that there exist $K+1$ *interesting* caching ratios denoted by r_s, where $s \in \{1, 2, \cdots, K-1\}$ in addition to $r = 0$ point (no caching) and $r = 1$ point (everything cached). The index s, which characterizes r_s for these points, represents the number of cached bits that can be used within one bit of the download (if this downloaded bit uses cached bits as side information). For example, if $s = 2$, this means that the user should use two of the cached bits as side information in the form of mixture of two bits if the caching ratio is r_2. The achievability scheme for any other caching ratio r can be obtained by memory-sharing between the most adjacent interesting caching ratios that include r. Consequently, the outer bound is a piece-wise linear convex curve that connects the $K+1$ interesting caching ratio points including the $(0, \frac{1}{C})$ point, where C is the PIR capacity without caching found in [12], and $(1, 0)$ where everything is cached; here, in (x, y), x denotes the caching ratio and y denotes the normalized download cost.

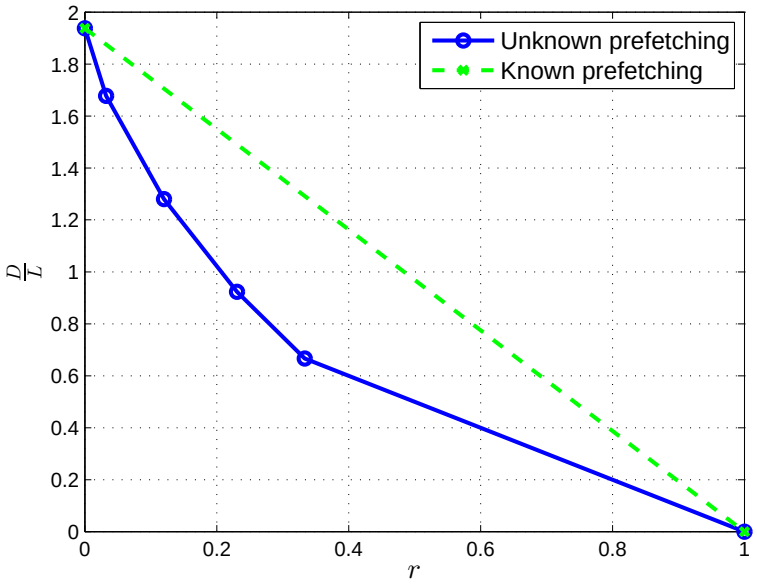

Figure 2.2: Comparison between the optimal download cost for known prefetching and the achievable download cost for unknown prefetching in (2.13) for $K = 5$ and $N = 2$.

As a direct corollary for Theorem 2.1, we note that since the databases do not know the cached bits, the download cost is strictly smaller than the case when the databases have the full knowledge about the cached bits in [28]. We state and prove this in the following corollary. As a concrete example, Figure 2.2 shows the gain that can be achieved due to the unawareness of the databases about the cached bits.

Corollary 2.1 (Unawareness gain) *The achievable normalized download cost* $\hat{D}(r)$ *in the cache-aided PIR with known prefetching [28]*

$$\hat{D}(r) = (1-r)\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}\right) \tag{2.15}$$

is strictly larger than the achievable normalized download cost $\bar{D}(r)$ *in* (2.13) *for*

$0 < r < 1$, *i.e., the databases' unawareness contributes to reducing the download cost beyond the memory-sharing scheme in [28].*

Proof: For $r = 0$, the achievable download cost $\bar{D}(r)$ in (2.13) is $\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}\right)$, which is the same as (2.15). For $r = 1$, the achievable download cost $\bar{D}(r)$ in (2.13) is 0, which is the same as (2.15). To show that $\hat{D}(r)$ in (2.15) is larger than $\bar{D}(r)$ in (2.13) for $0 < r < 1$, it suffices to show that there exists a caching ratio r such that $\bar{D}(r) < \hat{D}(r)$, since the other caching ratios can be achieved by the memory-sharing scheme. Taking $s = K - 1$ in (2.12), we have $r_{K-1} = \frac{1}{1+N}$. For $r = \frac{1}{1+N}$, we have $\bar{D}(r) = \frac{N}{1+N}$, and $\hat{D}(r) = \frac{N}{1+N}\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}\right)$. Therefore, for r_{K-1}, we have $\bar{D}(r) < \hat{D}(r)$, which shows the sub-optimality of $\hat{D}(r)$ in (2.15) for the case of known prefetching. ■

Our second result characterizes an inner bound (converse bound) for the normalized download cost $D^*(r)$ for general K, N, r.

Theorem 2.2 (Inner bound) *In the cache-aided PIR with uncoded and unknown prefetching, the normalized download cost is lower bounded as,*

$$D^*(r) \geq \tilde{D}(r) = \max_{i \in \{2,\cdots,K+1\}} \quad (1-r) \sum_{j=0}^{K+1-i} \frac{1}{N^j} - r \sum_{j=0}^{K-i} \frac{K+1-i-j}{N^j}, \qquad (2.16)$$

The proof of Theorem 2.2 can be found in Section 2.5. Theorem 2.2 implies that the inner bound is also a piece-wise linear curve, which consists of K line segments with decreasing slope as r increases. The points at which the curve changes

its slope are given by,

$$\tilde{r}_i = \frac{1}{1+N+N^2+\cdots+N^{K-i}}, \quad i = 1, \cdots, K-1. \tag{2.17}$$

We note that r_i in (2.12) and $\tilde{r}_i$ in (2.17) are the same for $i = 1$ and $i = K-1$.

As a consequence of Theorem 2.1 and Theorem 2.2, we characterize the optimal download cost caching ratio tradeoff for very low and very high caching ratios in the following corollary. Here, by very low caching ratios we mean $0 \leq r \leq r_1 = \tilde{r}_1 = \frac{1}{1+N+N^2+\cdots+N^{K-1}}$, and by very high caching ratios we mean $r_{K-2} = \frac{K-2}{(N+1)K+N^2-2N-2} \leq r \leq 1$. Note that, in the very high caching ratios, we have two segments, one in $r_{K-2} \leq r \leq r_{K-1}$ and the other in $r_{K-1} \leq r \leq 1$. Therefore, in the inner and outer bounds, each composed of K line segments, the first (very low r) and the last two (very high r) segments match giving exact result. This is stated and proved in the next corollary.

Corollary 2.2 (Optimal tradeoff for very low and very high caching ratios) *In the cache-aided PIR with uncoded and unknown prefetching, for very low caching ratios, i.e., for $r \leq \frac{1}{1+N+N^2+\cdots+N^{K-1}}$, the optimal normalized download cost is given by,*

$$D^*(r) = (1-r)\left(1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}}\right) - r\left(K-1+\frac{K-2}{N}+\cdots+\frac{1}{N^{K-2}}\right) \tag{2.18}$$

On the other hand, for very high caching ratios, i.e., for $r \geq \frac{K-2}{(N+1)K+N^2-2N-2}$, the

optimal normalized download cost is given by,

$$D^*(r) = \begin{cases} (1-r)\left(1+\frac{1}{N}\right) - r, & \frac{K-2}{(N+1)K+N^2-2N-2} \leq r \leq \frac{1}{1+N} \\ 1-r, & \frac{1}{1+N} \leq r \leq 1 \end{cases} . \tag{2.19}$$

Proof: First, from (2.12) and (2.17), let us note that

$$r_1 = \tilde{r}_1 = \frac{1}{1+N+N^2+\cdots+N^{K-1}}, \tag{2.20}$$

$$r_{K-2} = \frac{K-2}{(N+1)K+N^2-2N-2}, \tag{2.21}$$

$$r_{K-1} = \tilde{r}_{K-1} = \frac{1}{1+N}. \tag{2.22}$$

Then, we note from (2.13) that

$$\bar{D}(r_1) = \frac{\sum_{i=0}^{K-2} \binom{K}{2+i}(N-1)^i N}{\binom{K-2}{0} + \sum_{i=0}^{K-2} \binom{K-1}{1+i}(N-1)^i N} \tag{2.23}$$

$$= \frac{\frac{N}{(N-1)^2}\left[N^K - \sum_{i=0}^{1}\binom{K}{i}(N-1)^i\right]}{\binom{K-2}{0} + \frac{N}{(N-1)^1}\left[N^{K-1} - \sum_{i=0}^{0}\binom{K-1}{i}(N-1)^i\right]} \tag{2.24}$$

$$= \frac{N\left[N^K - 1 - K(N-1)\right]}{(N-1)^2 + N(N-1)\left[N^{K-1}-1\right]} \tag{2.25}$$

$$= \frac{N^{K+1} - KN^2 + (K-1)N}{N^{K+1} - N^K - N + 1} \tag{2.26}$$

Further, we note from (2.16), by choosing $i = 2$ and using $r = r_1$, that

$$\tilde{D}(r_1) \geq (1-r_1) \sum_{j=0}^{K+1-2} \frac{1}{N^j} - r_1 \sum_{j=0}^{K-2} \frac{K-1-j}{N^j} \tag{2.27}$$

$$= \left(1 - \frac{N-1}{N^K-1}\right)\frac{N^K-1}{N^K-N^{K-1}} - \frac{N-1}{N^K-1}\frac{N}{1-N}\left(-K + \frac{N^K-1}{N^K-N^{K-1}}\right) \tag{2.28}$$

$$= \frac{N^K-N}{N^K-1}\frac{N^K-1}{N^K-N^{K-1}} + \frac{N}{N^K-1}\left(-K + \frac{N^K-1}{N^K-N^{K-1}}\right) \tag{2.29}$$

$$= \frac{N^K-N}{N^K-N^{K-1}} + N\left(\frac{-K}{N^K-1} + \frac{1}{N^K-N^{K-1}}\right) \tag{2.30}$$

$$= \frac{N^{K+1} - KN^2 + (K-1)N}{N^{K+1} - N^K - N + 1} \tag{2.31}$$

$$= \bar{D}(r_1) \tag{2.32}$$

Thus, since $\tilde{D}(r_1) \leq \bar{D}(r_1)$ by definition, (2.32) implies $\tilde{D}(r_1) = \bar{D}(r_1)$.

Similarly, from (2.13),

$$\bar{D}(r_{K-2}) = \frac{\sum_{i=0}^{1}\binom{K}{K-1+i}(N-1)^i N}{\binom{K-2}{K-3} + \sum_{i=0}^{1}\binom{K-1}{K-2+i}(N-1)^i N} \tag{2.33}$$

$$= \frac{N^2 + (K-1)N}{N^2 + (K-2)N + (K-2)}, \tag{2.34}$$

and from (2.16) by choosing $i = K$ and using $r = r_{K-2}$,

$$\tilde{D}(r_{K-2}) \geq (1 - r_{K-2})\sum_{j=0}^{1}\frac{1}{N^j} - r_{K-2}\sum_{j=0}^{0}\frac{1-j}{N^j} \tag{2.35}$$

$$= \left(\frac{N^2 + (K-2)N}{N^2 + (K-2)N + (K-2)}\right)\left(1 + \frac{1}{N}\right) - \frac{K-2}{N^2 + (K-2)N + (K-2)} \tag{2.36}$$

$$= \frac{N^2 + (K-1)N}{N^2 + (K-2)N + (K-2)} \tag{2.37}$$

$$= \bar{D}(r_{K-2}) \tag{2.38}$$

implying $\tilde{D}(r_{K-2}) = \bar{D}(r_{K-2})$.

Finally, from (2.13),

$$\bar{D}(r_{K-1}) = \frac{N}{1+N}, \tag{2.39}$$

and from (2.16) by choosing $i = K+1$ and using $r = r_{K-1}$,

$$\tilde{D}(r_{K-1}) \geq \frac{N}{1+N} = \bar{D}(r_{K-1}) \tag{2.40}$$

implying $\tilde{D}(r_{K-1}) = \bar{D}(r_{K-1})$.

Therefore, $\tilde{D}(r) = \bar{D}(r)$ at $r = r_1$, $r = r_{K-2}$ and $r = r_{K-1}$. We also note that $\tilde{D}(0) = \bar{D}(0)$ and $\tilde{D}(1) = \bar{D}(1)$. Since both $\bar{D}(r)$ and $\tilde{D}(r)$ are linear functions of r, and since $\tilde{D}(0) = \bar{D}(0)$ and $\tilde{D}(r_1) = \bar{D}(r_1)$, we have $\tilde{D}(r) = \bar{D}(r) = D^*(r)$ for $0 \leq r \leq r_1$. This is the very low caching ratio region. In addition, since $\tilde{D}(r_{K-2}) = \bar{D}(r_{K-2})$, $\tilde{D}(r_{K-1}) = \bar{D}(r_{K-1})$ and $\tilde{D}(1) = \bar{D}(1)$, we have $\tilde{D}(r) = \bar{D}(r) = D^*(r)$ for $r_{K-2} \leq r \leq 1$. This is the very high caching ratio region. ■

As an example, the case of $K = 4$ and $N = 2$ is shown in Figure 2.3. In this case, $r_1 = \tilde{r}_1 = \frac{1}{15}$, $r_{K-2} = \frac{1}{5}$, and $r_{K-1} = \tilde{r}_{K-1} = \frac{1}{3}$. Therefore, we have exact results for $0 \leq r \leq \frac{1}{15}$ (very low caching ratios) and $\frac{1}{5} \leq r \leq 1$ (very high caching ratios). We have a gap between the achievability and the converse for medium caching ratios in $\frac{1}{15} \leq r \leq \frac{1}{5}$. More specifically, line segments connecting $(0, \frac{15}{8})$ and $(\frac{1}{15}, \frac{22}{15})$; connecting $(\frac{1}{5}, 1)$ and $(\frac{1}{3}, \frac{2}{3})$; and connecting $(\frac{1}{3}, \frac{2}{3})$ and $(1, 0)$ are tight.

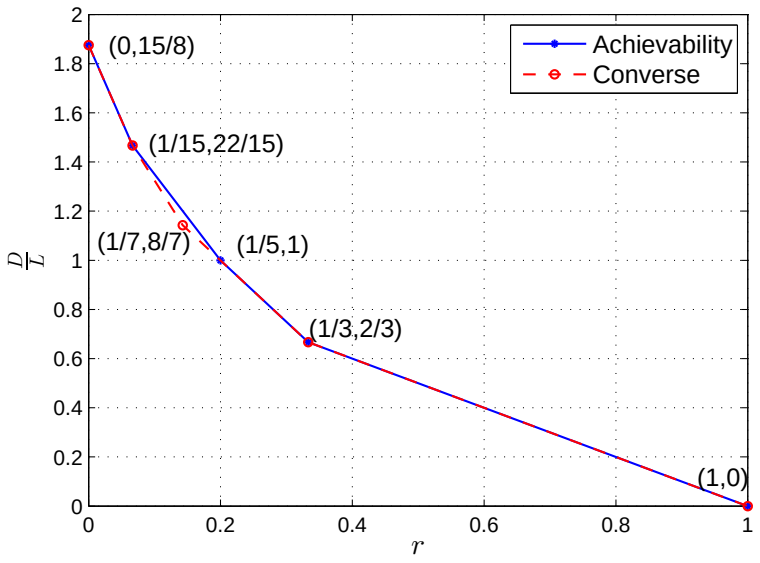

Figure 2.3: Inner and outer bounds for $K = 4$ and $N = 2$. For the (x, y) points in this figure, x denotes the caching ratio r and y denotes the normalized download cost $\frac{D}{L}$.

Finally, we characterize the exact tradeoff curve for any N, r for the special case of $K = 3$ in the following corollary.

Corollary 2.3 (Optimal tradeoff for $K = 3$) *In the cache-aided PIR with uncoded and unknown prefetching with $K = 3$ messages, the optimal download cost caching ratio tradeoff is given explicitly as (see Figure 2.4),*

$$D^*(r) = \begin{cases} (1-r)\left(1+\frac{1}{N}+\frac{1}{N^2}\right) - r\left(2+\frac{1}{N}\right), & 0 \leq r \leq \frac{1}{1+N+N^2} \\ (1-r)\left(1+\frac{1}{N}\right) - r, & \frac{1}{1+N+N^2} \leq r \leq \frac{1}{1+N} \\ 1-r, & \frac{1}{1+N} \leq r \leq 1 \end{cases} \tag{2.41}$$

Proof: The proof follows from the proof of Corollary 2.2. Note that in this case, from (2.20) and (2.21), $r_1 = r_{K-2} = \frac{1}{1+N+N^2}$; and from (2.22), $r_2 = r_{K-1} = \frac{1}{1+N}$. Thus, we have a tight result for $0 \leq r \leq r_1 = \frac{1}{1+N+N^2}$ (very low caching ratios) and

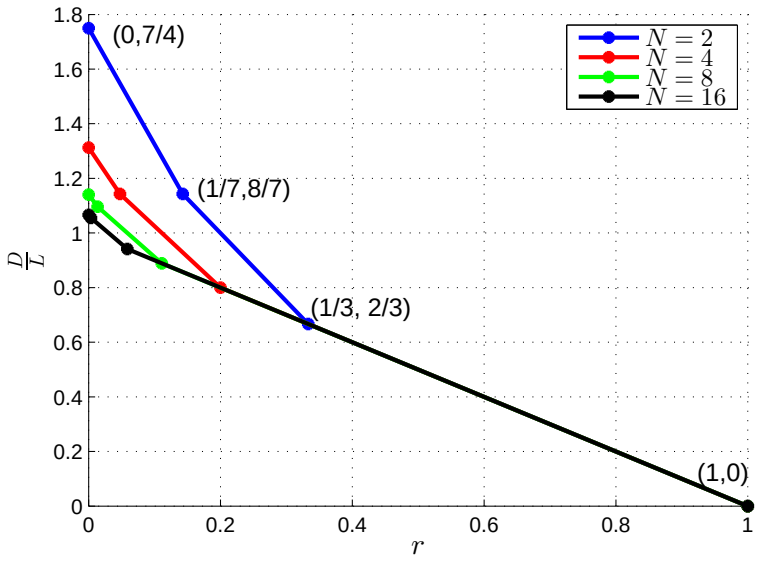

Figure 2.4: Optimal download cost caching ratio tradeoff for the case of $K = 3$ messages.

a tight result for $r_{K-2} = r_1 = \frac{1}{1+N+N^2} \leq r \leq 1$, i.e., a tight result for all $0 \leq r \leq 1$. We have three segments in this case: $[0, r_1]$, $[r_1, r_2]$ and $[r_2, 1]$ with three different line expressions for the exact result as given in (2.12)-(2.13) and written explicitly in (2.41). ■

2.4 Achievability Proof

Our achievability scheme is based on the PIR schemes in [12,28]. Similar to [12], we apply the following three principles recursively: 1) database symmetry, 2) message symmetry within each database, and 3) exploiting undesired messages as side information. Different from [12], we start the PIR scheme from the third principle due to the availability of pre-existing side information as a result of uncoded prefetching. These cached bits can be exploited right away as side information without compro-

mising the privacy constraint as the databases do not know them. We begin the discussion by presenting the case of $K = 3$ and $N = 2$ as a motivating example to illustrate the main ideas of our achievability scheme.

2.4.1 Motivating Example: The Optimal Tradeoff Curve for $K = 3$ Messages and $N = 2$ Databases

In this example, we show the achievability for $K = 3$ and $N = 2$. We know from Corollary 2.3 that the inner and the outer bounds match for this case. The optimal download cost caching ratio tradeoff is shown in Figure 2.4. We note that there are 4 corner points. Two of them are degenerate, corresponding to $r = 0$, $r = 1$ caching ratios. For $r = 0$, the user has no cached bits and is forced to apply the achievable scheme in [12] that achieves $\bar{D}(0) = \frac{7}{4} = \frac{1}{C}$. For $r = 1$, the user has already cached the entire desired file and does not download any extra bits from the databases, i.e., $\bar{D}(1) = 0$. We have two other corner points, corresponding to $r_1 = \frac{1}{1+N+N^2} = \frac{K-2}{(N+1)K+N^2-2N-2} = \frac{1}{7}$, and $r_2 = \frac{1}{1+N} = \frac{1}{3}$. In the sequel, we show the achievability of these two corner points.

2.4.1.1 Caching Ratio $r_1 = \frac{1}{7}$

Let s be the number of cached bits that are mixed together to form side information equation. The first corner point corresponds to $s = 1$. This means that the user exploits every bit in the cache individually as a side information. Using the notation in [24], we can say that the user starts downloading from round 2 that sums bits

from every two messages together. We next show how $s = 1$ suffices to achieve $r_1 = \frac{1}{7}$, $\bar{D}(\frac{1}{7}) = \frac{8}{7}$ for $K = 3$ and $N = 2$; see Figure 2.4.

We use a_i, b_i, and c_i to denote the bits of messages W_1, W_2 and W_3, respectively. We assume that the user wants to retrieve message W_1 privately without loss of generality. We initialize the process by permuting the indices of messages W_1, W_2, W_3 randomly and independently. The steps of the retrieval can be followed in Table 2.1. The user has already cached one bit from each message, i.e., a_1, b_1, c_1 as denoted by Z in Table 2.1. We start from the third principle by exploiting each bit in the cache as an individual side information. The user downloads $a_2 + b_1$ and $a_3 + c_1$ from the first database (DB1). Then, we apply the first principle, and the user downloads $a_4 + b_1$ and $a_5 + c_1$ from the second database (DB2) to satisfy the database symmetry. Next, we apply the second principle to ensure the message symmetry within the queries. The user downloads $b_2 + c_2$ from DB1, and $b_3 + c_3$ from DB2. At this point, all side information corresponding to the cached bits have been exploited. Next, we apply the third principle, since undesired message mixes are available in the form of $b_2 + c_2$ and $b_3 + c_3$. The user downloads $a_6 + b_3 + c_3$ from DB1. Finally, we apply the first principle of database symmetry, and the user downloads $a_7 + b_2 + c_2$ from DB2. Now, the iterations stop, since all the undesired side information is used and the symmetry across databases and symmetry within the queries is attained. We summarize the process in the query table in Table 2.1.

Since the databases do not know the local cache memory Z, and for each database, the user's queries are symmetric across messages, the privacy constraint (2.7) is satisfied. The decodability can be easily checked as the user can cancel out

Table 2.1: Query table for $K = 3$, $N = 2$ and $r_1 = \frac{1}{7}$

s	DB1	DB2
$s = 1$	$a_2 + b_1$	$a_4 + b_1$
	$a_3 + c_1$	$a_5 + c_1$
	$b_2 + c_2$	$b_3 + c_3$
	$a_6 + b_3 + c_3$	$a_7 + b_2 + c_2$

$Z = (a_1, b_1, c_1)$

b_1, c_1 which it has previously cached, and also cancel $b_2 + c_2$ and $b_3 + c_3$ which are previously downloaded, to obtain $a_2, \cdots, a_7$. Since a_1 is already cached, the user has $a_1, \cdots, a_7$. Here, $L = 7$ and the user has cached 1 bit from each message. There are total of 8 downloads. Hence $r = \frac{1}{7}$, and $\bar{D}(\frac{1}{7}) = \frac{8}{7}$.

2.4.1.2 Caching Ratio $r_2 = \frac{1}{3}$

For the second non-degenerate corner point, we have $s = 2$. This means that each 2 bits from the cache are mixed together to form a side information equation. We next show how $s = 2$ suffices to achieve $r_2 = \frac{1}{3}$, $\bar{D}(\frac{1}{3}) = \frac{2}{3}$ for $K = 3$ and $N = 2$; see Figure 2.4.

Let $[a_1, a_2, a_3]$, $[b_1, b_2, b_3]$, and $[c_1, c_2, c_3]$ denote a random permutation of the 3 bits of messages W_1, W_2 and W_3, respectively. Suppose the user caches a_1, b_1, c_1 in advance and wants to retrieve message W_1 privately. We start from the third principle. The user downloads $a_2 + b_1 + c_1$ from the first database (DB1). Then, we apply the first principle, and the user downloads $a_3 + b_1 + c_1$ from the second database (DB2). Now, the iterations stop, since all the undesired side information is used and the symmetry across databases and messages is attained. We summarize

the process in the query table in Table 2.2. In this case $L = 3$, hence $r = \frac{1}{3}$, and the normalized download cost is $\bar{D}(\frac{1}{3}) = \frac{2}{3}$.

Table 2.2: Query table for $K = 3$, $N = 2$ and $r_2 = \frac{1}{3}$

s	DB1	DB2
$s = 2$	$a_2 + b_1 + c_1$	$a_3 + b_1 + c_1$

$Z = (a_1, b_1, c_1)$

2.4.1.3 Caching Ratio $r = \frac{1}{5}$

So far, we have characterized all the corner points by varying $s = 1, 2$ and achieved the points corresponding to caching ratios r_s in addition to the degenerate caching ratios $r = 0$ and $r = 1$; see Figure 2.4. An achievable scheme for any other caching ratio can be obtained by memory-sharing between the two nearest corner points. As an example, we next consider the caching ratio $r = \frac{1}{5}$.

The achievability scheme for this case is a combination of the achievability schemes in Sections 2.4.1.1 and 2.4.1.2. Observe that by choosing $L = 10$, the achievable schemes in Sections 2.4.1.1 and 2.4.1.2 can be concatenated to achieve the caching ratio $r = \frac{1}{5}$. In this case, the user caches $a_1, a_2, b_1, b_2, c_1, c_2$ and wants to retrieve message W_1 privately. For cached bits a_1, b_1, c_1, we apply the same process as in Section 2.4.1.1, i.e., we use $s = 1$ and use every cached bit as individual side information equation. For cached bits a_2, b_2, c_2, we apply the same process as in Section 2.4.1.2, and choose $s = 2$, which implies that we use the mixture of two cached bits as a side information equation. We summarize the process in the query table in Table 2.3.

Table 2.3: Query table for $K = 3$, $N = 2$ and $r = \frac{1}{5}$

s	DB1	DB2
$s = 1$	$a_3 + b_1$	$a_5 + b_1$
	$a_4 + c_1$	$a_6 + c_1$
	$b_3 + c_3$	$b_4 + c_4$
	$a_7 + b_4 + c_4$	$a_8 + b_3 + c_3$
$s = 2$	$a_9 + b_2 + c_2$	$a_{10} + b_2 + c_2$

$Z = (a_1, a_2, b_1, b_2, c_1, c_2)$

Here, we have $L = 10$, therefore $r = \frac{1}{5}$, and $\bar{D}(\frac{1}{5}) = \frac{10}{10} = 1$. In fact, by applying [28, Lemma 1] and taking $\alpha = \frac{7}{10}$, we can show that the normalized download cost of this example can be obtained from the download costs obtained in Sections 2.4.1.1 and 2.4.1.2, as $\bar{D}(\frac{1}{5}) = \bar{D}(\frac{1}{7} \cdot \frac{7}{10} + \frac{1}{3} \cdot \frac{3}{10}) = \frac{7}{10}\bar{D}(\frac{1}{7}) + \frac{3}{10}\bar{D}(\frac{1}{3}) = \frac{7}{10} \cdot \frac{8}{7} + \frac{3}{10} \cdot \frac{2}{3} = 1$.

2.4.2 Achievable Scheme for the Corner Points for Arbitrary K, N

For fixed N and K, there are $K - 1$ non-degenerate corner points (in addition to degenerate caching ratios $r = 0$, $r = 1$). The caching ratios corresponding to these non-degenerate corner points are indexed by s, which enumerate the number of cached bits that are involved in the side information mixture. Hence, r_s is given by

$$r_s = \frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N}, \tag{2.42}$$

where $s \in \{1, 2, \ldots, K-1\}$. We choose the length of the message to be $L(s)$ for the corner point indexed by s, where

$$L(s) = \binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N \tag{2.43}$$

bits per message. The details of the achievable scheme are as follows:

1. *Initialization:* The user permutes each message randomly and independently. The user caches randomly and privately $\binom{K-2}{s-1}$ bits from each message. Set the round index to $i = s + 1$, where the ith round involves downloading sums of every i combinations of the K messages.

2. *Exploiting side information:* If $i = s + 1$, the user mixes s bits from the cache bits to form one side information equation. Each side information equation is added to one bit from the uncached portion of the desired message. Therefore, the user downloads $\binom{K-1}{s}$ equations in the form of a desired bit added to a mixture of s cached bits from other messages. On the other hand, if $i > s+1$, the user exploits the $\binom{K-1}{i-1}(N-1)^{i-s-1}$ side information equations generated from the remaining $(N-1)$ databases in the $(i-1)$th round.

3. *Symmetry across databases:* The user downloads the same number of equations with the same structure as in step 2 from every database. Consequently, the user downloads $\binom{K-1}{i-1}(N-1)^{i-s-1}$ bits from every database, which are done either using the cached bits as side information if $i = s + 1$, or the side information generated in the $(i-1)$th round if $i > s + 1$.

4. *Message symmetry:* To satisfy the privacy constraint, the user should download equal amount of bits from all other messages. Therefore, the user downloads $\binom{K-1}{i}(N-1)^{i-s-1}$ undesired equations from each database in the form of sum of i bits from the uncached portion of the undesired messages.

5. *Repeat* steps 2, 3, 4 after setting $i = i + 1$ until $i = K$.

6. *Shuffling the order of queries:* By shuffling the order of queries uniformly, all possible queries can be made equally likely regardless of the message index. This guarantees the privacy.

2.4.2.1 Decodability, Privacy, and the Achievable Normalized Download Cost

Decodability: It is clear that the side information in each round is either constructed from the cached bits (if $i = s + 1$) or obtained from the remaining $(N - 1)$ databases in the $(i - 1)$th round. Consequently, the user can cancel out these side information bits in order to decode the uncached portion of the desired message (the remaining $L(1 - r)$ bits).

Privacy: The randomized mapping of the cached and the uncached portions of the messages and the randomization of the order of queries guarantees privacy as in [12].

Normalized Download Cost: We now calculate the total number of downloaded bits for the caching ratio r in (2.42). First, we exploit s bits of side information. Therefore, each download is a sum of $s + 1$ bits. Since the second principle enforces symmetry across K messages, we download $\binom{K}{s+1}$ bits from a database. Due to the first principle enforcing symmetry across databases, in total, we download $\binom{K}{s+1}N$ bits. Since we utilize s bits of side information of undesired messages for each download, for each undesired message we use $\frac{\binom{K-1}{s}s}{K-1} = \binom{K-2}{s-1}$ bits, which is the

amount of bits we cached in advance for each message. Next, each download is a sum of $s+2$ bits since the available side information is in the form of sums of $s+1$ bits. Due to message symmetry and $(N-1)$ available side information from other $(N-1)$ databases, we download $\binom{K}{s+2}(N-1)$ bits from each database. Due to the first principle enforcing symmetry across databases, in total, we download $\binom{K}{s+2}(N-1)N$ bits. Next, each download is the sum of $s+3$ bits since the available side information is in the form of sums of $s+2$ bits. Note that in the previous iteration, each database provides $(N-1)$ sets of side information, and each database exploits the side information from the other $(N-1)$ databases. Therefore, we download $\binom{K}{s+3}(N-1)^2$ bits from each database. Due to the first principle enforcing symmetry across databases, in total, we download $\binom{K}{s+3}(N-1)^2N$ bits. By continuing in this manner, the total number of downloaded bits is,

$$D(r_s) = \sum_{i=0}^{K-1-s} \binom{K}{s+1+i}(N-1)^i N. \tag{2.44}$$

Now, we calculate the number of desired bits we have downloaded in this process. At the beginning of the iteration, each download is a sum of $s+1$ bits. If the download includes a desired bit, the other s bits are from the local cache memory. Therefore, we download $\binom{K-1}{s}$ desired bits from each database, and thus we download a total of $\binom{K-1}{s}N$ desired bits. Next, each download is a sum of $s+2$ bits. If the download includes a desired bit, the other $s+1$ bits are from the side information of undesired bits. For each database, there are $(N-1)$ sets of side information obtained from the previous iteration with one set from each

database. Therefore, we download $\binom{K-1}{s+1}(N-1)$ bits from each database, and thus we download a total of $\binom{K-1}{s+1}(N-1)N$ desired bits. Next, each download is a sum of $s+3$ bits. If the download includes a desired bit, the other $s+2$ bits are from the side information of undesired bits. For each database, there are $(N-1)^2$ sets of side information obtained from the previous iteration with $(N-1)$ sets from one database. Therefore, we download $\binom{K-1}{s+2}(N-1)^2N$ desired bits from this iteration. In the end, the number of desired bits we downloaded is $L(s) - \binom{K-2}{s-1}$, where $L(s)$ is given in (2.43). Finally, the normalized download cost is,

$$\bar{D}(r_s) = \frac{D(r_s)}{L(s)} = \frac{\sum_{i=0}^{K-1-s} \binom{K}{s+1+i}(N-1)^i N}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N}. \tag{2.45}$$

2.4.3 Achievable Scheme for Non-Corner Points for Arbitrary K, N

For caching ratios r which are not exactly equal to (2.42) for some s, we first find an s such that $r_s < r < r_{s+1}$, and combine the achievability schemes of r_s and r_{s+1}. Then, we can write the achievable normalized download cost as a convex combination of $\bar{D}(r_s)$ and $\bar{D}(r_{s+1})$ using [28, Lemma 1] as follows,

$$\bar{D}(r) = \alpha \bar{D}(r_s) + (1-\alpha)\bar{D}(r_{s+1}), \tag{2.46}$$

where $r = \alpha r_s + (1-\alpha) r_{s+1}$ and r_s is defined in (2.42), and $\bar{D}(r)$ is given in (2.45).

2.5 Converse Proof

In this section, we derive an inner bound for the cache-aided PIR with uncoded and unknown prefetching. The inner bound is tight in general for very high and very low caching ratios, and in particular, the inner bound is tight everywhere for $K = 3$. We extend the techniques presented in [12, 28] to our problem. We first need the following lemma, which characterizes a lower bound on the length of the undesired portion of the answer strings as a consequence of the privacy constraint.

Lemma 2.1 (Interference lower bound) *For the cache-aided PIR with unknown and uncoded prefetching, the interference from undesired messages within the answer strings $D(r) - L(1-r)$ is lower bounded by,*

$$D(r) - L(1-r) + o(L) \geq I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right) \tag{2.47}$$

for all $k \in \{2, \ldots, K\}$.

If the privacy constraint is absent, the user downloads only $L(1-r)$ bits in order to decode the desired message, however, when the privacy constraint is present, it should download $D(r)$. The difference $D(r) - L(1-r)$ corresponds to the undesired portion of the answer strings. Lemma 2.1 shows that this portion is lower bounded by the mutual information between the answer strings and the messages $W_{k:K}$ after knowing the first $W_{1:k-1}$ messages and the cached bits. Lemma 2.1 provides $K-1$ lower bounds on $D(r) - L(1-r)$ by changing the index k from 2 to K. Each of

these $K-1$ bounds contributes a different line segment for the final inner bound. Note that Lemma 2.1 is an extension to [12, Lemma 5] if $k=2$, $r=0$.

Proof: We start with the right hand side of (2.47),

$$I\left(W_{k:K};Q_{1:N}^{[k-1]},A_{1:N}^{[k-1]}|W_{1:k-1},Z,\mathbb{H}\right)$$
$$=I\left(W_{k:K};Q_{1:N}^{[k-1]},A_{1:N}^{[k-1]},W_{k-1}|W_{1:k-2},Z,\mathbb{H}\right)-I\left(W_{k:K};W_{k-1}|W_{1:k-2},Z,\mathbb{H}\right) \tag{2.48}$$

For the first term on the right hand side of (2.48), we have

$$I\left(W_{k:K};Q_{1:N}^{[k-1]},A_{1:N}^{[k-1]},W_{k-1}|W_{1:k-2},Z,\mathbb{H}\right)$$
$$=I\left(W_{k:K};Q_{1:N}^{[k-1]},A_{1:N}^{[k-1]}|W_{1:k-2},Z,\mathbb{H}\right)$$
$$+I\left(W_{k:K};W_{k-1}|Q_{1:N}^{[k-1]},A_{1:N}^{[k-1]},W_{1:k-2},Z,\mathbb{H}\right) \tag{2.49}$$
$$\overset{(2.9)}{=}I\left(W_{k:K};Q_{1:N}^{[k-1]},A_{1:N}^{[k-1]}|W_{1:k-2},Z,\mathbb{H}\right)+o(L) \tag{2.50}$$
$$\overset{(2.5),(2.6)}{=}I\left(W_{k:K};A_{1:N}^{[k-1]}|W_{1:k-2},Z,\mathbb{H},Q_{1:N}^{[k-1]}\right)+o(L) \tag{2.51}$$
$$=H\left(A_{1:N}^{[k-1]}|W_{1:k-2},Z,\mathbb{H},Q_{1:N}^{[k-1]}\right)-H\left(A_{1:N}^{[k-1]}|W_{1:k-2},Z,\mathbb{H},Q_{1:N}^{[k-1]},W_{k:K}\right)+o(L) \tag{2.52}$$
$$\overset{(2.9)}{=}H\left(A_{1:N}^{[k-1]}|W_{1:k-2},Z,\mathbb{H},Q_{1:N}^{[k-1]}\right)$$
$$-H\left(W_{k-1},A_{1:N}^{[k-1]}|W_{1:k-2},Z,\mathbb{H},Q_{1:N}^{[k-1]},W_{k:K}\right)+o(L) \tag{2.53}$$
$$\leq H\left(A_{1:N}^{[k-1]}|W_{1:k-2},Z,\mathbb{H},Q_{1:N}^{[k-1]}\right)-H\left(W_{k-1}|W_{1:k-2},Z,\mathbb{H},Q_{1:N}^{[k-1]},W_{k:K}\right)+o(L) \tag{2.54}$$
$$\overset{(2.5),(2.6)}{=}H\left(A_{1:N}^{[k-1]}|W_{1:k-2},Z,\mathbb{H},Q_{1:N}^{[k-1]}\right)-H\left(W_{k-1}|Z,\mathbb{H}\right)+o(L) \tag{2.55}$$

$$= H\left(A_{1:N}^{[k-1]}|W_{1:k-2}, Z, \mathbb{H}, Q_{1:N}^{[k-1]}\right) - L(1-r) + o(L) \tag{2.56}$$

$$\leq D(r) - L(1-r) + o(L) \tag{2.57}$$

where (2.50), (2.53) follow from the reliability constraint of W_{k-1}, (2.51) follows from the independence of the queries $Q_{1:N}^{[k-1]}$ and the messages $W_{k:K}$ given Z and $\mathbb{H}$, (2.54) follows from the chain rule and the non-negativity of the entropy function, (2.55) is due to the fact that given Z and $\mathbb{H}$, W_{k-1} is statistically independent of $(W_{1:k-2}, W_{k:K}, Q_{1:N}^{[k-1]})$, (2.56) follows from the uncoded nature of the cache, and (2.57) follows from conditioning reduces entropy.

For the second term on the right hand side of (2.48), we have

$$I\left(W_{k:K}; W_{k-1}|W_{1:k-2}, Z, \mathbb{H}\right)$$

$$= H\left(W_{k-1}|W_{1:k-2}, Z, \mathbb{H}\right) - H\left(W_{k-1}|W_{1:k-2}, W_{k:K}, Z, \mathbb{H}\right) \tag{2.58}$$

$$= (L - Lr) - (L - Lr) \tag{2.59}$$

$$= 0 \tag{2.60}$$

Combining (2.48), (2.57), and (2.60) yields (2.47). ■

In the following lemma, we prove an inductive relation for the mutual information term on the right hand side of (2.47).

Lemma 2.2 (Induction lemma) *For all $k \in \{2, \ldots, K\}$, the mutual information*

term in Lemma 2.1 can be inductively lower bounded as,

$$I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right) \geq \frac{1}{N} I\left(W_{k+1:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k}, Z, \mathbb{H}\right) + \frac{L(1-r) - o(L)}{N} - (K-k+1)Lr. \tag{2.61}$$

Lemma 2.2 relates the mutual information between $W_{k:K}$ and the answer strings to the same mutual information term with $W_{k+1:K}$, i.e., it shifts the term by one message. Since the two terms have the same structure, Lemma 2.2 constructs an inductive relation.

We obtain an explicit lower bound for $I(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H})$ by applying this lemma $K-k+1$ times, and therefore characterize an explicit lower bound on $D(r) - L(1-r)$. We do this in Lemma 2.3 by combining Lemma 2.1 and Lemma 2.2. Lemma 2.2 reduces to [12, Lemma 6] if $r = 0$.

Proof: We start with the left hand side of (2.61),

$$\begin{aligned}
&I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right) \\
&= I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}, Z, \mathbb{H} | W_{1:k-1}\right) - I(W_{k:K}; Z, \mathbb{H} | W_{1:k-1}) && (2.62) \\
&= I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}\right) + I\left(W_{k:K}; Z, \mathbb{H} | W_{1:k-1}, Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}\right) \\
&\quad - I(W_{k:K}; Z, \mathbb{H} | W_{1:k-1}) && (2.63) \\
&\geq I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}\right) - I(W_{k:K}; Z, \mathbb{H} | W_{1:k-1}) && (2.64)
\end{aligned}$$

where (2.64) follows from the non-negativity of mutual information.

For the first term in (2.64), we have

$$
\begin{aligned}
& NI\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}\right) \\
& \geq \sum_{n=1}^{N} I\left(W_{k:K}; Q_n^{[k-1]}, A_n^{[k-1]} | W_{1:k-1}\right) && (2.65) \\
& \overset{(2.7)}{=} \sum_{n=1}^{N} I\left(W_{k:K}; Q_n^{[k]}, A_n^{[k]} | W_{1:k-1}\right) && (2.66) \\
& \geq \sum_{n=1}^{N} I\left(W_{k:K}; A_n^{[k]} | W_{1:k-1}, Q_n^{[k]}\right) && (2.67) \\
& \overset{(2.8)}{=} \sum_{n=1}^{N} H\left(A_n^{[k]} | W_{1:k-1}, Q_n^{[k]}\right) && (2.68) \\
& \geq \sum_{n=1}^{N} H\left(A_n^{[k]} | W_{1:k-1}, \mathbb{H}, Q_{1:N}^{[k]}, A_{1:n-1}^{[k]}, Z\right) && (2.69) \\
& \overset{(2.8)}{=} \sum_{n=1}^{N} I\left(W_{k:K}; A_n^{[k]} | W_{1:k-1}, \mathbb{H}, Q_{1:N}^{[k]}, A_{1:n-1}^{[k]}, Z\right) && (2.70) \\
& = I\left(W_{k:K}; A_{1:N}^{[k]} | W_{1:k-1}, \mathbb{H}, Q_{1:N}^{[k]}, Z\right) && (2.71) \\
& \overset{(2.5),(2.6)}{=} I\left(W_{k:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k-1}, Z, \mathbb{H}\right) && (2.72) \\
& \overset{(2.9)}{=} I\left(W_{k:K}; W_k, Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k-1}, Z, \mathbb{H}\right) - o(L) && (2.73) \\
& = I\left(W_{k:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k}, Z, \mathbb{H}\right) + I\left(W_{k:K}; W_k | W_{1:k-1}, Z, \mathbb{H}\right) - o(L) && (2.74) \\
& = I\left(W_{k:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k}, Z, \mathbb{H}\right) + L(1-r) - o(L) && (2.75) \\
& = I\left(W_{k+1:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k}, Z, \mathbb{H}\right) + L(1-r) - o(L) && (2.76)
\end{aligned}
$$

where (2.65), (2.69) follow from the non-negativity of mutual information, (2.66) follows from the privacy constraint, (2.67) follows from the chain rule and the non-negativity of the mutual information, (2.68), (2.70) follow from the fact that the answer string $A_n^{[k]}$ is a deterministic function of $(Q_n^{[k]}, W_{1:K})$, (2.71) follows from the

chain rule, (2.72) follows from the statistical independence of $(Q_{1:N}^{[k]}, W_{k:K})$ given $(Z, \mathbb{H})$, (2.73) is consequence of the decodability of W_k from $(Q_{1:N}^{[k]}, A_{1:N}^{[k]})$, and (2.75) is due to the uncoded assumption of the cached bits.

For the second term in (2.64), we have

$$
\begin{aligned}
& I(W_{k:K}; Z, \mathbb{H} | W_{1:k-1}) \\
& \quad = H\left(W_{k:K} | W_{1:k-1}\right) - H(W_{k:K} | W_{1:k-1}, Z, \mathbb{H}) && (2.77) \\
& \quad = (K-k+1)\, L - (K-k+1)\, L(1-r) && (2.78) \\
& \quad = (K-k+1)\, Lr && (2.79)
\end{aligned}
$$

where (2.79) follows from the uncoded nature of the cached bits.

Combining (2.64), (2.76), and (2.79) yields (2.61). ■

Now we are ready to derive the general inner bound for arbitrary K, N, r. To obtain this bound, we use Lemma 2.1 to find K lower bounds on the length of the undesired portion of the answer strings $D(r) - L(1-r)$. Each lower bound is obtained by varying the index k in the lemma from $k = 2$ to $k = K$. Next, we inductively lower bound each result of Lemma 2.1 by using Lemma 2.2, precisely $(K-k+1)$ times, to get K explicit lower bounds. This is stated in the following lemma.

Lemma 2.3 *For N and K, we have*

$$
D(r) \geq L(1-r) \sum_{j=0}^{K+1-k} \frac{1}{N^j} - Lr \sum_{j=0}^{K-k} \frac{K+1-k-j}{N^j} - o(L), \tag{2.80}
$$

where $k = 2, \ldots, K+1$.

Proof: We have

$$
\begin{aligned}
& D(r) + o(L) \\
& \overset{(2.47)}{\geq} L(1-r) + I\left(W_{k:K};, Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right) && (2.81) \\
& \overset{(2.61)}{\geq} L(1-r) + \frac{L(1-r) - o(L)}{N} - (K-k+1)Lr \\
& \quad + \frac{1}{N} I\left(W_{k+1:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k}, Z, \mathbb{H}\right) && (2.82) \\
& \overset{(2.61)}{\geq} L(1-r)\left[1 + \frac{1}{N} + \frac{1}{N^2} + o(L)\right] - Lr\left[(K-k+1) + \frac{(K-k)}{N}\right] \\
& \quad + \frac{1}{N^2} I\left(W_{k+2:K}; Q_{1:N}^{[k+1]}, A_{1:N}^{[k+1]} | W_{1:k+1}, Z, \mathbb{H}\right) && (2.83) \\
& \overset{(2.61)}{\geq} \cdots && (2.84) \\
& \overset{(2.61)}{\geq} L(1-r) \sum_{j=0}^{K+1-k} \frac{1}{N^j} - Lr \sum_{j=0}^{K-k} \frac{K+1-k-j}{N^j} \\
& \quad + o(L), && (2.85)
\end{aligned}
$$

where (2.81) follows from Lemma 2.1 starting from general index k, and the remaining bounding steps correspond to successive application of Lemma 2.2. ■

We conclude the converse proof by dividing by L and taking the limit as $L \to \infty$, then for $k = 2, \cdots, K+1$, we have

$$
D^*(r) \geq (1-r) \sum_{j=0}^{K+1-k} \frac{1}{N^j} - r \sum_{j=0}^{K-k} \frac{K+1-k-j}{N^j} \tag{2.86}
$$

Finally, (2.86) gives K intersecting line segments, therefore, the normalized download cost is lower bounded by their maximum value

$$D^*(r) \geq \max_{i\in\{2,\cdots,K+1\}} (1-r) \sum_{j=0}^{K+1-i} \frac{1}{N^j} - r \sum_{j=0}^{K-i} \frac{K+1-i-j}{N^j}. \tag{2.87}$$

2.6 Further Examples

2.6.1 $K = 4$ Messages, $N = 2$ Databases

For $K = 4$ and $N = 2$, we show the achievable PIR schemes for caching ratios $r_1 = \frac{1}{15}$ in Table 2.4, $r_2 = \frac{1}{5}$ in Table 2.5, and $r_3 = \frac{1}{3}$ in Table 2.6. The achievable normalized download costs for these caching ratios are $\frac{22}{15}$, 1 and $\frac{2}{3}$, respectively. We show the normalized download cost and caching ratio trade off curve in Figure 2.3.

Table 2.4: Query table for $K = 4$, $N = 2$ and $r_1 = \frac{1}{15}$

s	DB1	DB2
$s = 1$	$a_2 + b_1$	$a_5 + b_1$
	$a_3 + c_1$	$a_6 + c_1$
	$a_4 + d_1$	$a_7 + d_1$
	$b_2 + c_2$	$b_4 + c_4$
	$b_3 + d_2$	$b_5 + d_4$
	$c_3 + d_3$	$c_5 + d_5$
	$a_8 + b_4 + c_4$	$a_{11} + b_2 + c_2$
	$a_9 + b_5 + d_4$	$a_{12} + b_3 + d_2$
	$a_{10} + c_5 + d_5$	$a_{13} + c_3 + d_3$
	$b_6 + c_6 + d_6$	$b_7 + c_7 + d_7$
	$a_{14} + b_7 + c_7 + d_7$	$a_{15} + b_6 + c_6 + d_6$

$Z = (a_1, b_1, c_1, d_1)$

Table 2.5: Query table for $K = 4$, $N = 2$ and $r_2 = \frac{1}{5}$

s	DB1	DB2
$s = 2$	$a_3 + b_1 + c_1$	$a_6 + b_1 + c_1$
	$a_4 + d_1 + b_2$	$a_7 + d_1 + b_2$
	$a_5 + c_2 + d_2$	$a_8 + c_2 + d_2$
	$b_3 + c_3 + d_3$	$b_4 + c_4 + d_4$
	$a_9 + b_4 + c_4 + d_4$	$a_{10} + b_3 + c_3 + d_3$

$Z = (a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2)$

Table 2.6: Query table for $K = 4$, $N = 2$ and $r_3 = \frac{1}{3}$

s	DB1	DB2
$s = 3$	$a_2 + b_1 + c_1 + d_1$	$a_3 + b_1 + c_1 + d_1$

$Z = (a_1, b_1, c_1, d_1)$

2.6.2 $K = 4$ Messages, $N = 3$ Databases

For $K = 4$ and $N = 3$, we show the achievable PIR schemes for caching ratios $r_1 = \frac{1}{40}$ in Table 2.7, $r_2 = \frac{2}{17}$ in Table 2.8, and $r_3 = \frac{1}{4}$ in Table 2.9. We show the normalized download cost and caching ratio trade off in Figure 2.5. The achievable normalized download costs for these caching ratios are $\frac{27}{20}$, $\frac{18}{17}$ and $\frac{3}{4}$, respectively. By comparing Figure 2.5 with Figure 2.3, we observe that, for fixed K, as N grows, the gap between the achievable bound and the converse bound shrinks. This observation will be specified in Section 2.7.

2.6.3 $K = 5$, $K = 10$ and $K = 100$ Messages, $N = 2$ Databases

For $N = 2$, we show the numerical results for the inner and outer bounds for $K = 5$, $K = 10$ and $K = 100$ in Figures 2.6, 2.7 and 2.8. For fixed N as K grows, the gap between the achievable bound and converse bound increases. This observation will

Table 2.7: Query table for $K = 4$, $N = 3$ and $r_1 = \frac{1}{40}$

s	DB1	DB2	DB3
$s = 1$	$a_2 + b_1$	$a_5 + b_1$	$a_8 + b_1$
	$a_3 + c_1$	$a_6 + c_1$	$a_9 + c_1$
	$a_4 + d_1$	$a_7 + d_1$	$a_{10} + d_1$
	$b_2 + c_2$	$b_4 + c_4$	$b_6 + c_6$
	$b_3 + d_2$	$b_5 + d_4$	$b_7 + d_6$
	$c_3 + d_3$	$c_5 + d_5$	$c_7 + d_7$
	$a_{11} + b_4 + d_4$	$a_{17} + b_2 + c_2$	$a_{23} + b_2 + c_2$
	$a_{12} + b_5 + d_4$	$a_{18} + b_3 + d_2$	$a_{24} + b_3 + d_2$
	$a_{13} + c_5 + d_5$	$a_{19} + c_3 + d_3$	$a_{25} + c_3 + d_3$
	$a_{14} + b_6 + c_6$	$a_{20} + b_6 + c_6$	$a_{26} + b_4 + c_4$
	$a_{15} + b_7 + d_6$	$a_{21} + b_7 + d_6$	$a_{27} + b_5 + d_4$
	$a_{16} + c_7 + d_7$	$a_{22} + c_7 + d_7$	$a_{28} + c_5 + d_5$
	$b_8 + c_8 + d_8$	$b_{10} + c_{10} + d_{10}$	$b_{12} + c_{12} + d_{12}$
	$b_9 + c_9 + d_9$	$b_{11} + c_{11} + d_{11}$	$b_{13} + c_{13} + d_{13}$
	$a_{29} + b_{10} + c_{10} + d_{10}$	$a_{33} + b_8 + c_8 + d_8$	$a_{37} + b_8 + c_8 + d_8$
	$a_{30} + b_{11} + c_{11} + d_{11}$	$a_{34} + b_9 + c_9 + d_9$	$a_{38} + b_9 + c_9 + d_9$
	$a_{31} + b_{12} + c_{12} + d_{12}$	$a_{35} + b_{12} + c_{12} + d_{12}$	$a_{39} + b_{10} + c_{10} + d_{10}$
	$a_{32} + b_{13} + c_{13} + d_{13}$	$a_{36} + b_{13} + c_{13} + d_{13}$	$a_{40} + b_{11} + c_{11} + d_{11}$

$Z = (a_1, b_1, c_1, d_1)$

be elaborated in Section 2.7.

2.6.4 $K = 5$, $K = 10$ and $K = 100$ Messages, $N = 3$ Databases

For $N = 3$, we show the numerical results for the inner and outer bounds for $K = 5$, $K = 10$ and $K = 100$ in Figures 2.9, 2.10 and 2.11. For fixed N as K grows, the gap between the achievable bound and converse bound increases. This observation will be further clarified in Section 2.7.

Table 2.8: Query table for $K = 4$, $N = 3$ and $r_2 = \frac{2}{17}$

s	DB1	DB2	DB3
$s = 2$	$a_3 + b_1 + c_1$	$a_6 + b_1 + c_1$	$a_9 + b_1 + c_1$
	$a_4 + d_1 + b_2$	$a_7 + d_1 + b_2$	$a_{10} + d_1 + b_2$
	$a_5 + c_2 + d_2$	$a_8 + c_2 + d_2$	$a_{11} + c_2 + d_2$
	$b_3 + c_3 + d_3$	$b_4 + c_4 + d_4$	$b_5 + c_5 + d_5$
	$a_{12} + b_4 + c_4 + d_4$	$a_{14} + b_3 + c_3 + d_3$	$a_{16} + b_3 + c_3 + d_3$
	$a_{13} + b_5 + c_5 + d_5$	$a_{15} + b_5 + c_5 + d_5$	$a_{17} + b_4 + c_4 + d_4$

$$Z = (a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2)$$

Table 2.9: Query table for $K = 4$, $N = 3$ and $r_3 = \frac{1}{4}$

s	DB1	DB2	DB3
$s = 3$	$a_2 + b_1 + c_1 + d_1$	$a_3 + b_1 + c_1 + d_1$	$a_4 + b_1 + c_1 + d_1$

$$Z = (a_1, b_1, c_1, d_1)$$

2.7 Gap Analysis

In this section, we analyze the gap between the achievability and converse bounds for general N, K, and r, and show that the worst-case gap, which happens when $N = 2$ and $K \to \infty$, is at most $\frac{1}{6}$. We start this section with an interesting property for the monotonicity of the achievable bounds. We first see an example. For $N = 2$, $K = 4$, $K = 5$ and $K = 6$, the achievable bounds are shown in Figure 2.12. The achievable bound for $K = 6$ is above the achievable bound for $K = 5$, and the achievable bound for $K = 5$ is above the achievable bound for $K = 4$. By denoting $r_s^{(K)}$ as the caching ratio with total K messages and parameter s (see (2.12)), we observe that $(r_1^{(5)}, \bar{D}(r_1^{(5)}))$ falls on the line connecting $(r_0^{(4)}, \bar{D}(r_0^{(4)}))$ and $(r_1^{(4)}, \bar{D}(r_1^{(4)}))$. This observation is general, $(r_s^{(K+1)}, \bar{D}(r_s^{(K+1)}))$ falls on the line connecting $(r_{s-1}^{(K)}, \bar{D}(r_{s-1}^{(K)}))$ and $(r_s^{(K)}, \bar{D}(r_s^{(K)}))$. We state and prove this observation in the following lemma.

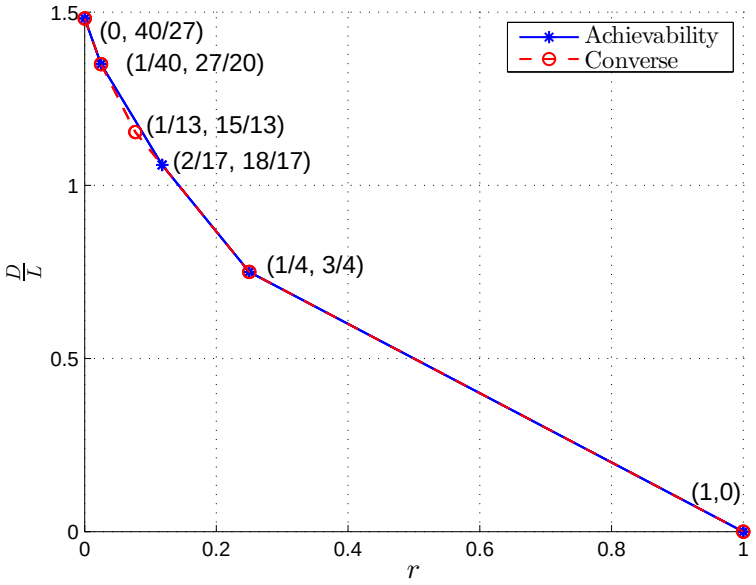

Figure 2.5: Inner and outer bounds for $K = 4$ and $N = 3$.

Lemma 2.4 (Monotonicity of the achievable bounds) *In cache-aided PIR with uncoded and unknown prefetching, for fixed number of databases N, if the number of messages K increases, then the achievable normalized download cost increases. Furthermore, we have*

$$r_s^{(K+1)} = \alpha r_{s-1}^{(K)} + (1-\alpha) r_s^{(K)}, \tag{2.88}$$

$$\bar{D}(r_s^{(K+1)}) = \alpha \bar{D}(r_{s-1}^{(K)}) + (1-\alpha) \bar{D}(r_s^{(K)}), \tag{2.89}$$

where $0 \leq \alpha \leq 1$.

Proof: To show (2.89) is equivalent to show

$$\bar{D}(r_s^{(K+1)}) - \bar{D}(r_s^{(K)}) = \alpha \left(\bar{D}(r_{s-1}^{(K)}) - \bar{D}(r_s^{(K)}) \right), \tag{2.90}$$

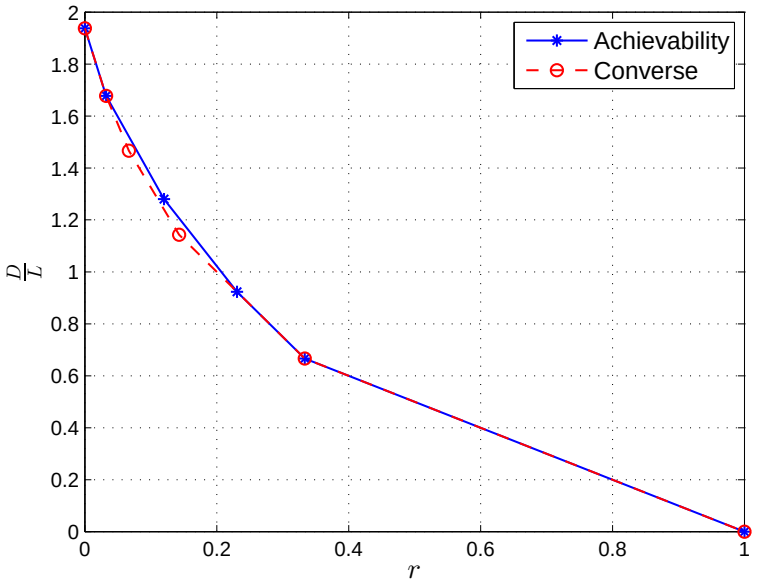

Figure 2.6: Inner and outer bounds for $K = 5$ and $N = 2$.

where $\bar{D}\left(r_{s-1}^{(K)}\right) > \bar{D}\left(r_s^{(K)}\right)$. From (2.88), we have

$$\alpha = \frac{r_s^{(K)} - r_s^{(K+1)}}{r_s^{(K)} - r_{s-1}^{(K)}}. \tag{2.91}$$

Therefore, to show (2.90) is equivalent to show

$$\left(r_s^{(K)} - r_{s-1}^{(K)}\right)\left(\bar{D}(r_s^{(K+1)}) - \bar{D}(r_s^{(K)})\right) = \left(r_s^{(K)} - r_s^{(K+1)}\right)\left(\bar{D}(r_{s-1}^{(K)}) - \bar{D}(r_s^{(K)})\right). \tag{2.92}$$

Let $\bar{D}(r_s^{(K)}) = \frac{D_s^{(K)}}{L_s^{(K)}}$, where

$$L_s^{(K)} = \binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N, \tag{2.93}$$

$$D_s^{(K)} = \sum_{i=0}^{K-1-s} \binom{K}{s+1+i}(N-1)^i N. \tag{2.94}$$

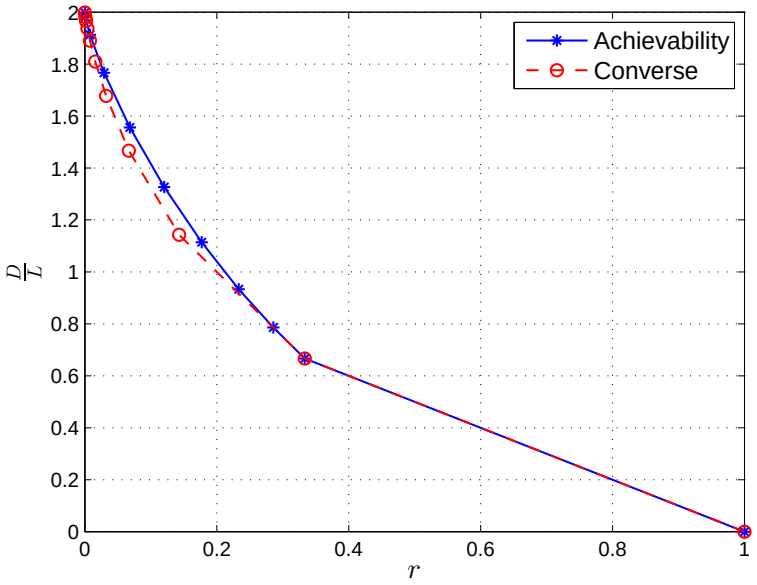

Figure 2.7: Inner and outer bounds for $K = 10$ and $N = 2$.

To show (2.92) is equivalent to show

$$\left[\frac{\binom{K-2}{s-1}}{L_s^{(K)}} - \frac{\binom{K-2}{s-2}}{L_{s-1}^{(K)}}\right]\left[\frac{D_s^{(K+1)}}{L_s^{(K+1)}} - \frac{D_s^{(K)}}{L_s^{(K)}}\right] = \left[\frac{\binom{K-2}{s-1}}{L_s^{(K)}} - \frac{\binom{K-1}{s-1}}{L_s^{(K+1)}}\right]\left[\frac{D_{s-1}^{(K)}}{L_{s-1}^{(K)}} - \frac{D_s^{(K)}}{L_s^{(K)}}\right], \tag{2.95}$$

which is obtained by using (2.12), (2.13), (2.93) and (2.94). Expanding (2.95), we have

$$\begin{aligned}&\frac{\binom{K-2}{s-1}}{L_s^{(K)}}\frac{D_s^{(K+1)}}{L_s^{(K+1)}} - \frac{\binom{K-2}{s-2}}{L_{s-1}^{(K)}}\frac{D_s^{(K+1)}}{L_s^{(K+1)}} + \frac{\binom{K-2}{s-2}}{L_{s-1}^{(K)}}\frac{D_s^{(K)}}{L_s^{(K)}} \\ &\quad = \frac{\binom{K-2}{s-1}}{L_s^{(K)}}\frac{D_{s-1}^{(K)}}{L_{s-1}^{(K)}} - \frac{\binom{K-1}{s-1}}{L_s^{(K+1)}}\frac{D_{s-1}^{(K)}}{L_{s-1}^{(K)}} + \frac{\binom{K-1}{s-1}}{L_s^{(K+1)}}\frac{D_s^{(K)}}{L_s^{(K)}}.\end{aligned} \tag{2.96}$$

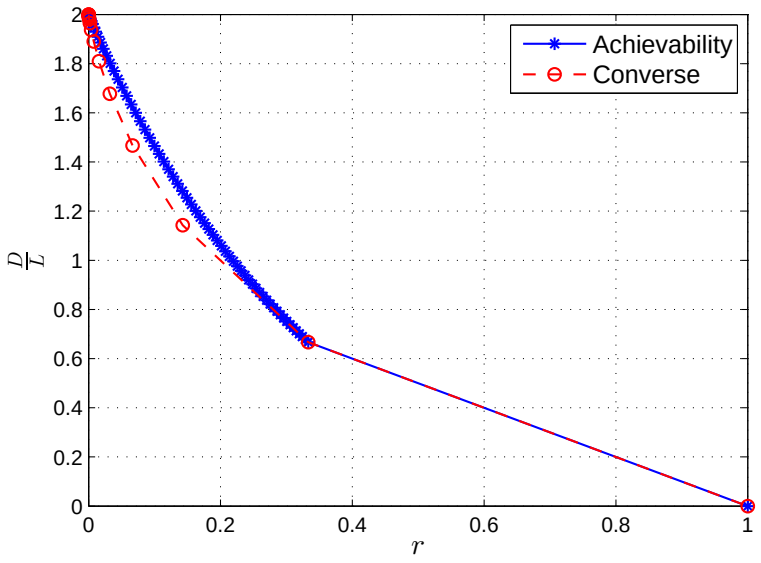

Figure 2.8: Inner and outer bounds for $K = 100$ and $N = 2$.

Multiplying $L_s^{(K)} L_{s-1}^{(K)} L_s^{(K+1)}$ to both side of (2.96), we have

$$\begin{aligned}
&\binom{K-2}{s-1} D_s^{(K+1)} L_{s-1}^{(K)} + \binom{K-1}{s-1} D_{s-1}^{(K)} L_s^{(K)} + \binom{K-2}{s-2} D_s^{(K)} L_s^{(K+1)} \\
&= \binom{K-2}{s-1} D_{s-1}^{(K)} L_s^{(K+1)} + \binom{K-2}{s-2} D_s^{(K+1)} L_s^{(K)} + \binom{K-1}{s-1} D_s^{(K)} L_{s-1}^{(K)}. \qquad (2.97)
\end{aligned}$$

By using (2.93) and (2.94), we further have

$$\begin{aligned}
&\binom{K-2}{s-1} \left[\sum_{i=0}^{K-s} \binom{K+1}{s+1+i} (N-1)^i N \right] \\
&\quad \times \left[\binom{K-2}{s-2} + \sum_{i=0}^{K-s} \binom{K-1}{s-1+i} (N-1)^i N \right] \\
&+ \binom{K-1}{s-1} \left[\sum_{i=0}^{K-s} \binom{K}{s+i} (N-1)^i N \right] \\
&\quad \times \left[\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i} (N-1)^i N \right]
\end{aligned}$$

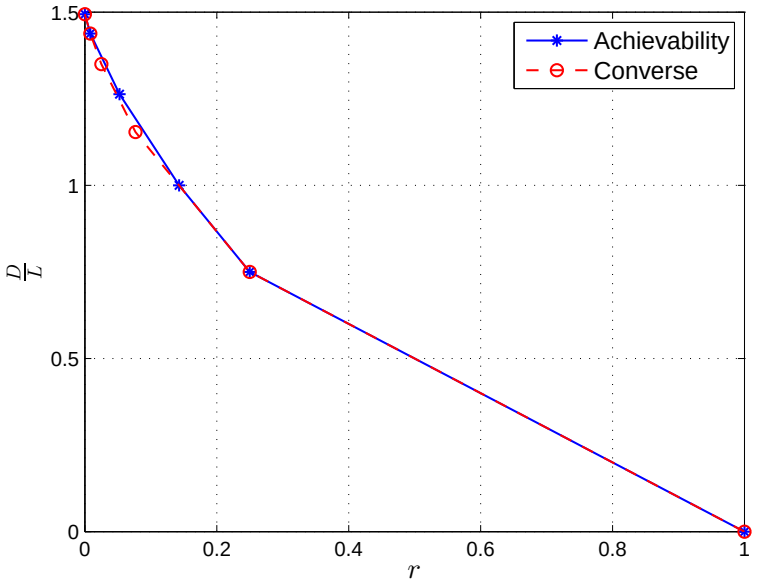

Figure 2.9: Inner and outer bounds for $K = 5$ and $N = 3$.

$$
\begin{aligned}
&+ \binom{K-2}{s-2} \left[\sum_{i=0}^{K-1-s} \binom{K}{s+1+i} (N-1)^i N \right] \\
&\quad \times \left[\binom{K-1}{s-1} + \sum_{i=0}^{K-s} \binom{K}{s+i} (N-1)^i N \right] \\
&= \binom{K-2}{s-1} \left[\sum_{i=0}^{K-s} \binom{K}{s+i} (N-1)^i N \right] \\
&\quad \times \left[\binom{K-1}{s-1} + \sum_{i=0}^{K-s} \binom{K}{s+i} (N-1)^i N \right] \\
&\quad + \binom{K-2}{s-2} \left[\sum_{i=0}^{K-s} \binom{K+1}{s+1+i} (N-1)^i N \right] \\
&\quad \quad \times \left[\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i} (N-1)^i N \right] \\
&\quad + \binom{K-1}{s-1} \left[\sum_{i=0}^{K-1-s} \binom{K}{s+1+i} (N-1)^i N \right] \\
&\quad \quad \times \left[\binom{K-2}{s-2} + \sum_{i=0}^{K-s} \binom{K-1}{s-1+i} (N-1)^i N \right]. \qquad (2.98)
\end{aligned}
$$

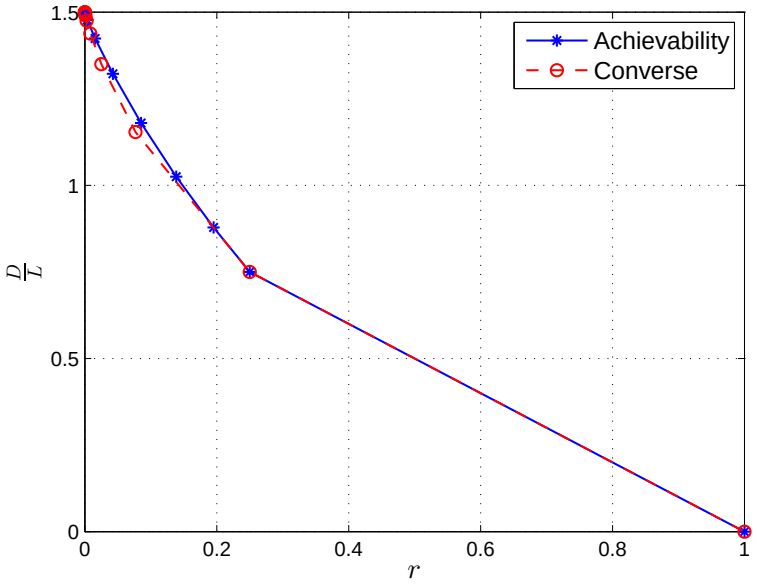

Figure 2.10: Inner and outer bounds for $K = 10$ and $N = 3$.

By canceling same terms on both sides, we have

$$
\begin{aligned}
&\binom{K-2}{s-1}\left[\sum_{i=0}^{K-s}\binom{K+1}{s+1+i}(N-1)^i\right]\left[\sum_{i=0}^{K-s}\binom{K-1}{s-1+i}(N-1)^i\right] \\
&\quad+\binom{K-1}{s-1}\left[\sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i\right]\left[\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i\right] \\
&\quad+\binom{K-2}{s-2}\left[\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^i\right]\left[\sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i\right] \\
&=\binom{K-2}{s-1}\left[\sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i\right]\left[\sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i\right] \\
&\quad+\binom{K-2}{s-2}\left[\sum_{i=0}^{K-s}\binom{K+1}{s+1+i}(N-1)^i\right]\left[\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i\right] \\
&\quad+\binom{K-1}{s-1}\left[\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^i\right]\left[\sum_{i=0}^{K-s}\binom{K-1}{s-1+i}(N-1)^i\right]. \qquad (2.99)
\end{aligned}
$$

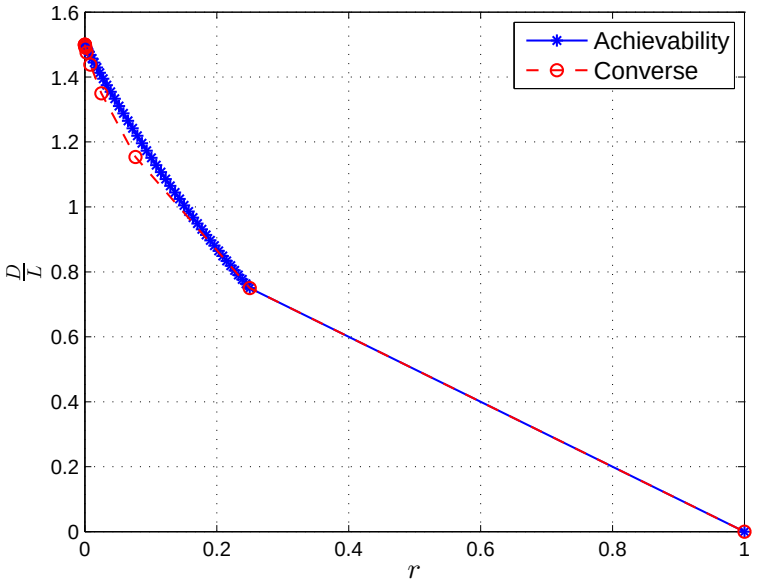

Figure 2.11: Inner and outer bounds for $K = 100$ and $N = 3$.

By using the fact that $\binom{K}{s} = \binom{K-1}{s} + \binom{K-1}{s-1}$, we have

$$
\begin{aligned}
&\binom{K-2}{s-1}\left[\sum_{i=0}^{K-s}\left(\binom{K}{s+1+i}+\binom{K}{s+i}\right)(N-1)^i\right] \\
&\quad\times\left[\sum_{i=0}^{K-s}\binom{K-1}{s-1+i}(N-1)^i\right] \\
&\quad+\left(\binom{K-2}{s-1}+\binom{K-2}{s-2}\right)\left[\sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i\right] \\
&\quad\times\left[\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i\right] \\
&\quad+\binom{K-2}{s-2}\left[\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^i\right] \\
&\quad\times\left[\sum_{i=0}^{K-s}\left(\binom{K-1}{s+i}+\binom{K-1}{s+i-1}\right)(N-1)^i\right] \\
&=\binom{K-2}{s-1}\left[\sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i\right] \\
&\quad\times\left[\sum_{i=0}^{K-s}\left(\binom{K-1}{s+i}+\binom{K-1}{s+i-1}\right)(N-1)^i\right]
\end{aligned}
$$

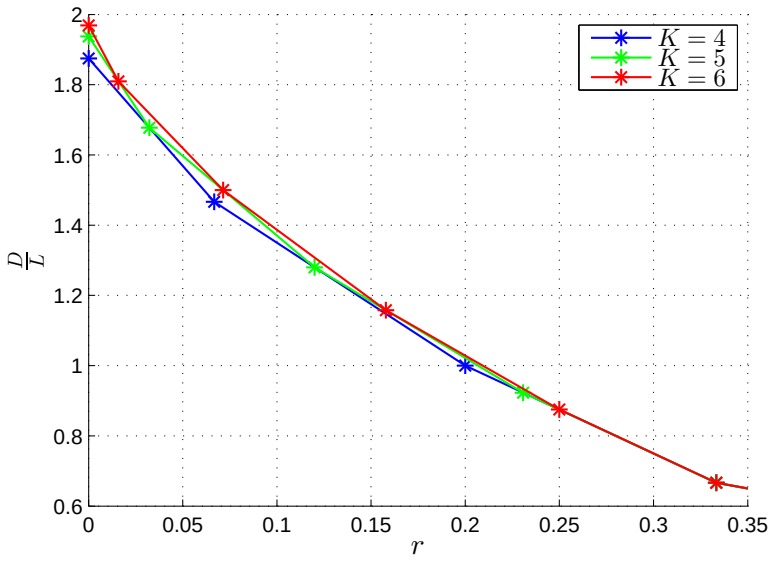

Figure 2.12: Outer bounds for $N = 2$, $K = 4$, $K = 5$ and $K = 6$.

$$
\begin{aligned}
&+ \binom{K-2}{s-2} \left[\sum_{i=0}^{K-s} \left(\binom{K}{s+1+i} + \binom{K}{s+i} \right) (N-1)^i \right] \\
&\quad \times \left[\sum_{i=0}^{K-1-s} \binom{K-1}{s+i} (N-1)^i \right] \\
&+ \left(\binom{K-2}{s-1} + \binom{K-2}{s-2} \right) \\
&\quad \times \left[\sum_{i=0}^{K-1-s} \binom{K}{s+1+i} (N-1)^i \right] \left[\sum_{i=0}^{K-s} \binom{K-1}{s-1+i} (N-1)^i \right]. \qquad (2.100)
\end{aligned}
$$

Since the left hand side of (2.100) is equal to the right hand side of (2.100), (2.89) holds.

To show $\alpha \geq 0$, since $r_s^{(K)} > r_{s-1}^{(K)}$ in (2.91), it suffices to show that $r_s^{(K)} \geq r_s^{(K+1)}$. From (2.12), it is equivalent to show that

$$
\frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i} (N-1)^i N} \geq \frac{\binom{K-1}{s-1}}{\binom{K-1}{s-1} + \sum_{i=0}^{K-s} \binom{K}{s+i} (N-1)^i N}. \qquad (2.101)
$$

By using the fact that $\binom{K}{s} = \binom{K-1}{s} + \binom{K-1}{s-1}$, we have

$$\begin{aligned}&\frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N} \\ &\geq \frac{\binom{K-2}{s-1} + \binom{K-2}{s-2}}{\binom{K-2}{s-1} + \binom{K-2}{s-2} + \sum_{i=0}^{K-s} \left[\binom{K-1}{s+i} + \binom{K-1}{s+i-1}\right](N-1)^i N}\end{aligned} \tag{2.102}$$

which is equivalent to

$$\frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N} \geq \frac{\binom{K-2}{s-2}}{\binom{K-2}{s-2} + \sum_{i=0}^{K-s} \binom{K-1}{s+i-1}(N-1)^i N}. \tag{2.103}$$

By using (2.12), (2.103) is equivalent to

$$r_s^{(K)} \geq r_{s-1}^{(K)}. \tag{2.104}$$

Since (2.104) holds, we have $\alpha \geq 0$. Furthermore, $\alpha \leq 1$ can be proved similarly. For fixed N, since $\bar{D}(r_0^{(K+1)}) > \bar{D}(r_0^{(K)})$, the achievable normalized download cost monotonically increases. ■

The following lemma provides an asymptotic upper bound for the achievable normalized download cost as a smooth function in (r, N). From this expression, we characterize the worst-case gap between the outer and the inner bounds to be $\frac{1}{6}$.

Lemma 2.5 (Asymptotics and the worst-case gap) *In cache-aided PIR with uncoded and unknown prefetching, as $K \to \infty$, the outer bound is tightly upper*

bounded by,

$$\bar{D}(r) \leq \frac{N(1-r)^2}{(N-1)+r} \tag{2.105}$$

Hence, the worst-case gap is $\frac{1}{6}$. *The asymptotic unawareness multiplicative gain over memory-sharing in [28] is* $\frac{1-r}{1+\frac{r}{N-1}} \leq 1$.

Proof: We write the outer bound $\bar{D}(r_s)$ as

$$\bar{D}(r_s) = \frac{\sum_{i=0}^{K-1-s} \binom{K}{s+1+i}(N-1)^i N}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N} \tag{2.106}$$

$$= \frac{\frac{\sum_{i=0}^{K-1-s} \binom{K}{s+1+i}(N-1)^i}{\sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i}}{\frac{\binom{K-2}{s-1}}{\sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N} + 1} \tag{2.107}$$

$$= \frac{\psi_1(N,K,s)}{\psi_2(N,K,s)+1}. \tag{2.108}$$

Denote $\lambda = \frac{s}{K}$. To upper bound $\psi_1(N,K,s)$,

$$\psi_1(N,K,s) = \frac{\sum_{i=0}^{K-1-s} \binom{K}{s+1+i}(N-1)^i}{\sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i} \tag{2.109}$$

$$= \frac{\sum_{i=0}^{K-1-s} \frac{K}{s+1+i}\binom{K-1}{s+i}(N-1)^i}{\sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i} \tag{2.110}$$

$$\leq \frac{\sum_{i=0}^{K-1-s} \frac{K}{s}\binom{K-1}{s+i}(N-1)^i}{\sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i} = \frac{1}{\lambda}. \tag{2.111}$$

We upper bound the reciprocal of $\psi_2(N,K,s)$ as,

$$\frac{1}{\psi_2(N,K,s)}$$

$$= \sum_{i=0}^{K-1-s} \frac{\binom{K-1}{s+i}(N-1)^i}{\binom{K-2}{s-1}} N \tag{2.112}$$

$$= \sum_{i=0}^{K-1-s} \frac{(K-1)(K-1-s)(K-2-s)\cdots(K-i-s)}{s(s+1)(s+2)\cdots(s+i)} N(N-1)^i \tag{2.113}$$

$$\leq \sum_{i=0}^{K-1-s} \frac{K(K-s)^i}{s^{i+1}} N(N-1)^i \tag{2.114}$$

$$= \sum_{i=0}^{(1-\lambda)K-1} \frac{(1-\lambda)^i}{\lambda^{i+1}} N(N-1)^i \tag{2.115}$$

$$= \frac{N}{\lambda} \sum_{i=0}^{(1-\lambda)K-1} \left(\frac{(1-\lambda)(N-1)}{\lambda} \right)^i. \tag{2.116}$$

Now, if $\lambda > 1 - \frac{1}{N}$, then $\frac{(1-\lambda)(N-1)}{\lambda} < 1$. Hence, as $K \to \infty$, $\frac{1}{\psi_2(N,K,s)}$ converges to

$$\lim_{K\to\infty} \frac{1}{\psi_2(N,K,s)} \leq \frac{N}{\lambda} \sum_{i=0}^{\infty} \left(\frac{(1-\lambda)(N-1)}{\lambda} \right)^i \tag{2.117}$$

$$= \frac{N}{\lambda} \cdot \frac{1}{1 - \frac{(1-\lambda)(N-1)}{\lambda}} \tag{2.118}$$

$$= \frac{N}{N\lambda - (N-1)}. \tag{2.119}$$

Moreover, (2.113) can be lower bounded by keeping the first ϵK terms in the sum for any ϵ such that $0 < \epsilon < 1 - \lambda$,

$$\frac{1}{\psi_2(N,K,s)}$$

$$\geq \sum_{i=0}^{\epsilon K} \frac{(K-1)(K-1-s)(K-2-s)\cdots(K-i-s)}{s(s+1)(s+2)\cdots(s+i)} N(N-1)^i \tag{2.120}$$

$$\geq \sum_{i=0}^{\epsilon K} \frac{(K-1)(K-\epsilon K-s)^i}{(s+\epsilon K)^{i+1}} N(N-1)^i \tag{2.121}$$

$$= \sum_{i=0}^{\epsilon K} \frac{(1-\frac{1}{K})((1-(\lambda+\epsilon))^i}{(\lambda+\epsilon)^{i+1}} N(N-1)^i. \tag{2.122}$$

Similarly, by taking $K \to \infty$, for any $0 < \epsilon < 1 - \lambda$, we have

$$\lim_{K\to\infty} \frac{1}{\psi_2(N,K,s)} \geq \frac{N}{\lambda+\epsilon} \sum_{i=0}^{\infty} \left(\frac{(1-(\lambda+\epsilon))(N-1)}{\lambda+\epsilon} \right)^i \tag{2.123}$$

$$= \frac{N}{N(\lambda+\epsilon)-(N-1)}. \tag{2.124}$$

Since ϵ is arbitrarily chosen, then as $K \to \infty$, $\epsilon \to 0$, we have $\psi_2(N,K,s) \to \frac{N\lambda-(N-1)}{N}$.

Consequently, as $K \to \infty$, r_s converges to

$$r_s \to r = \lim_{K\to\infty} \frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N} \tag{2.125}$$

$$= \lim_{K\to\infty} \frac{\psi_2(N,K,s)}{\psi_2(N,K,s)+1} \tag{2.126}$$

$$= \frac{N\lambda-(N-1)}{N\lambda+1}. \tag{2.127}$$

Note that if $\lambda = 1 - \frac{1}{N}$, then $r = 0$, while if $\lambda = 1$, then $r = \frac{1}{1+N}$. This means that the restriction in the limit to have $\lambda > 1 - \frac{1}{N}$ is without loss of generality as $\lambda > 1 - \frac{1}{N}$ corresponds to the entire range of r other than the $1-r$ matching bound. We can write λ as

$$\lambda = \frac{r+(N-1)}{N(1-r)}. \tag{2.128}$$

Substituting in (2.108), we have the following upper bound on $\bar{D}(r)$

$$\bar{D}(r) \leq \frac{\frac{1}{\lambda}}{\frac{N\lambda-(N-1)}{N}+1} \tag{2.129}$$

$$= \frac{N}{\lambda(N\lambda+1)} \tag{2.130}$$

$$= \frac{N}{\frac{r+(N-1)}{N(1-r)}\left(\frac{r+(N-1)}{(1-r)}+1\right)} \tag{2.131}$$

$$= \frac{N^2(1-r)^2}{(r+(N-1))^2+(1-r)(r+(N-1))} \tag{2.132}$$

$$= \frac{N^2(1-r)^2}{Nr+N(N-1)} \tag{2.133}$$

$$= \frac{N(1-r)^2}{(N-1)+r}. \tag{2.134}$$

The memory-sharing scheme in [28] achieves $\frac{N}{N-1}(1-r)$ if $K \to \infty$, hence the asymptotic unawareness gain is given by the multiplicative factor $\frac{1-r}{1+\frac{r}{N-1}} \leq 1$.

For the inner bound, we note that the ith corner point is given by,

$$\tilde{r}_i = \frac{1}{1+N+\cdots+N^i}, \quad i = 1, \cdots, K-1. \tag{2.135}$$

Therefore, although there exist K linear bounds, it suffices to consider only a small number of them, as the remaining bounds are concentrated around $r = 0$. Denote the gap between the inner and the outer bounds by $\Delta(N, K, r)$. We note that the gap $\Delta(N, \infty, r)$ is a piece-wise convex function for $0 \leq r \leq 1$ since it is the difference between a convex function $\bar{D}(r)$ and a piece-wise linear function. Hence, the maximizing caching ratio for the gap exists exactly at the corner points $\tilde{r}_i$ and it suffices to examine the gap at these corner points.

For the outer bound, we have

$$\bar{D}(\tilde{r}_i) \leq \frac{N\left(1-\frac{1}{1+N+\cdots+N^i}\right)^2}{(N-1)+\frac{1}{1+N+\cdots+N^i}} \tag{2.136}$$

$$= \frac{N(1+N+N^2+\cdots+N^i-1)^2}{(N-1)(1+N+\cdots+N^i)^2+(1+N+\cdots+N^i)} \tag{2.137}$$

$$= \frac{N^2(1+N+\cdots+N^{i-1})^2}{N^i(1+N+\cdots+N^i)}. \tag{2.138}$$

Furthermore, for the inner bound, we have

$$\tilde{D}(\tilde{r}_i) = \left(1+\frac{1}{N}+\cdots+\frac{1}{N^i}\right) - \frac{1}{1+N+\cdots+N^i}\left(i+1+\frac{i}{N}+\cdots+\frac{1}{N^i}\right) \tag{2.139}$$

$$= \frac{1+N+\cdots+N^i}{N^i} - \frac{(i+1)N^i+iN^{i-1}+\cdots+1}{N^i(1+N+\cdots+N^i)} \tag{2.140}$$

$$= \frac{(1+N+\cdots+N^i)^2}{N^i(1+N+\cdots+N^i)} - \frac{(1+2N+3N^2+\cdots+(i+1)N^i)}{N^i(1+N+\cdots+N^i)} \tag{2.141}$$

Consequently, we can upper bound the asymptotic gap at the corner point $\tilde{r}_i$ as

$$\Delta(N,\infty,\tilde{r}_i) = \bar{D}(\tilde{r}_i) - \tilde{D}(\tilde{r}_i) \tag{2.142}$$

$$\leq \frac{N^2(1+N+\cdots+N^{i-1})^2-(1+N+\cdots+N^i)^2}{N^i(1+N+\cdots+N^i)} + \frac{(1+2N+3N^2+\cdots+(i+1)N^i)}{N^i(1+N+\cdots+N^i)} \tag{2.143}$$

$$= \frac{-1-2N(1+N+\cdots+N^{i-1}))}{N^i(1+N+\cdots+N^i)} + \frac{(1+2N+3N^2+\cdots+(i+1)N^i)}{N^i(1+N+\cdots+N^i)} \tag{2.144}$$

$$= \frac{N^2 + 2N^3 + \cdots + (i-1)N^i}{N^i(1 + N + \cdots + N^i)} \tag{2.145}$$

$$= \frac{\frac{1}{N^{i-2}} + \frac{2}{N^{i-3}} + \cdots + (i-1)}{1 + N + \cdots + N^i} \tag{2.146}$$

Hence, $\Delta(N, \infty, \tilde{r}_i)$ is monotonically decreasing in N. Therefore,

$$\begin{aligned} \Delta(N, K, r) &\leq \Delta(2, \infty, r) \\ &\leq \max_i \ \frac{(2)^2 + 2(2)^3 + \cdots + (i-1)(2)^i}{2^i(1 + 2 + \cdots + 2^i)} \end{aligned} \tag{2.147}$$

For the case $N = 2$, we note that all the inner bounds after the 6th corner point are concentrated around $r = 0$ since $\tilde{r}_i \leq \frac{1}{127}$ for $i \geq 6$. Therefore, it suffices to characterize the gap only for the first 6 corner points. Considering the 6th corner point which corresponds to $\tilde{r}_6 = \frac{1}{127} = 0.0078$, and $\bar{D}(r) \leq 2$ trivially for all r, and $\tilde{D}(\frac{1}{127}) = 1.8898$. Hence, $\Delta(2, \infty, r) \leq 0.11$, for $r \leq \frac{1}{127}$. Now, we focus on calculating the gap at $\tilde{r}_i$, $i = 1, \cdots, 6$. Examining all the corner points, we see that $r = \frac{1}{15}$ is the maximizing caching ratio for the gap (corresponding to $i = 3$), and $\Delta(2, \infty, \frac{1}{15}) \leq \frac{1}{6}$, which is the worst-case gap. ■

2.8 Conclusion

In this chapter, we studied the cache-aided PIR problem from N non-communicating and replicated databases, when the cache stores uncoded bits that are unknown to the databases. We determined inner and outer bounds for the optimal normalized download cost $D^*(r)$ as a function of the total number of messages K, the

number of databases N, and the caching ratio r. Both inner and outer bounds are piece-wise linear functions in r (for fixed N, K) that consist of K line segments. The bounds match in two specific regimes: the very low caching ratio regime, i.e., $r \leq \frac{1}{1+N+N^2+\cdots+N^{K-1}}$, where $D^*(r) = (1-r)\left(1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}}\right) - r\left((K-1)+\frac{K-2}{N}+\cdots+\frac{1}{N^{K-2}}\right)$; and the very high caching ratio regime, where $D^*(r) = (1-r)(1+\frac{1}{N})-r$, for $\frac{K-2}{(N+1)K+N^2-2N-2} \leq r \leq \frac{1}{1+N}$ and $D^*(r) = 1-r$, for $r \geq \frac{1}{1+N}$. As a direct corollary for this result, we characterized the exact tradeoff between the download cost and the caching ratio for $K = 3$. For general K, N, and r, we showed that the largest gap between the achievability and the converse bounds is $\frac{1}{6}$. The outer bound shows significant reduction in the download cost with respect to the case when the cache content is fully known at all databases [28], which achieves $D^*(r) = (1-r)(1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}})$ by memory-sharing.

The achievable scheme extends the greedy scheme in [12] so that it starts with exploiting the cache bits as side information. For fixed K, N, there are $K-1$ non-degenerate corner points. These points differ in the number of cached bits that contribute in generating one side information equation. The achievability for the remaining caching ratios is done by memory-sharing between the two adjacent corner points that enclose that caching ratio r. For the converse, we extend the induction-based techniques in [12,28] to account for the availability of uncoded and unknown prefetching. The converse proof hinges on developing $K-1$ lower bounds on the length of the undesired portion of the answer string. By applying induction on each bound separately, we obtain the piece-wise linear inner bound.

CHAPTER 3

Cache-Aided Private Information Retrieval with Partially Known Uncoded Prefetching: Fundamental Limits

3.1 Introduction

We consider the problem of private information retrieval (PIR) from N non-colluding and replicated databases, when the user is equipped with a cache that holds an uncoded fraction r of the symbols from each of the K stored messages in the databases. This model operates in a two-phase scheme, namely, the prefetching phase where the user acquires side information and the retrieval phase where the user privately downloads the desired message. In the prefetching phase, the user receives $\frac{r}{N}$ uncoded fraction of each message from the nth database. This side information is known only to the nth database and unknown to the remaining databases, i.e., the user possesses *partially known* side information. We investigate the optimal normalized download cost $D^*(r)$ in the retrieval phase as a function of K, N, r. We develop lower and upper bounds for the optimal download cost. The bounds match in general for the cases of very low caching ratio ($r \leq \frac{1}{N^{K-1}}$) and very high caching ratio ($r \geq \frac{K-2}{N^2-3N+KN}$). We fully characterize the optimal download cost caching

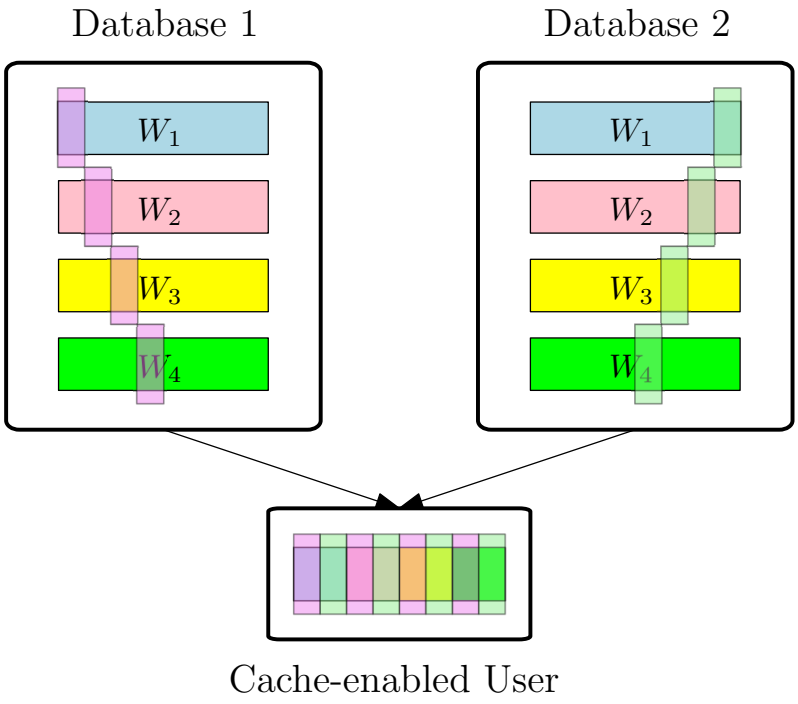

Figure 3.1: Cache-aided PIR with partially known and uncoded prefetching for $N = 2$, $K = 4$ and $r = \frac{1}{4}$.

ratio tradeoff for $K = 3$. For general K, N, and r, we show that the largest additive gap between the achievability and the converse bounds is $\frac{5}{32}$.

3.2 System Model

We consider a PIR problem with N non-communicating databases. Each database stores an identical copy of K statistically independent messages, $W_1, \ldots, W_K$. Each message is L bits long,

$$H(W_1) = \cdots = H(W_K) = L, \tag{3.1}$$

$$H(W_1, \ldots, W_K) = H(W_1) + \cdots + H(W_K). \tag{3.2}$$

The user (retriever) has a local cache memory which can store up to KLr bits, where $0 \leq r \leq 1$, and r is called the *caching ratio*. There are two phases in this system: the *prefetching phase* and the *retrieval phase*.

In the prefetching phase, for each message W_k, the user randomly and independently chooses Lr bits out of the L bits to cache. The user caches the Lr bits of each message by prefetching the same amount of bits from each database, i.e., the user prefetches $\frac{KLr}{N}$ bits from each database. $\forall n \in [N]$, where $[N] = \{1, 2, \ldots, N\}$, we denote the indices of the cached bits from the nth database by $\mathbb{H}_n$ and the cached bits from the nth database by the random variable Z_n. Therefore, the overall cached content Z is equal to $(Z_1, \ldots, Z_N)$, and $H(Z) = \sum_{n=1}^{N} H(Z_n) = KLr$. We further denote the indices of the cached bits by $\mathbb{H}$. Therefore, we have $\mathbb{H} = \bigcup_{n=1}^{N} \mathbb{H}_n$, where $\mathbb{H}_{n_1} \cap \mathbb{H}_{n_2} = \emptyset$, if $n_1 \neq n_2$. Since the user caches a subset of the bits from each message, this is called *uncoded prefetching*. Here, we consider the case where database n knows $\mathbb{H}_n$, but it does not know $\mathbb{H} \setminus \mathbb{H}_n$. We refer to Z as *partially known prefetching*; see Fig. 3.1.

In the retrieval phase, the user privately generates an index $\theta \in [K]$, and wishes to retrieve message W_θ such that it is impossible for any individual database to identify θ. Note that during the prefetching phase, the desired message is unknown a priori. Therefore, the cached bit indices $\mathbb{H}$ are independent of the desired message index θ. Note further that the cached bit indices $\mathbb{H}$ are independent of the message contents. Therefore, for random variables θ, $\mathbb{H}$, and $W_1, \ldots, W_K$, we have

$$H(\theta, \mathbb{H}, W_1, \ldots, W_K) = H(\theta) + H(\mathbb{H}) + H(W_1) + \cdots + H(W_K). \tag{3.3}$$

The user sends N queries $Q_1^{[\theta]}, \ldots, Q_N^{[\theta]}$ to the N databases, where $Q_n^{[\theta]}$ is the query sent to the nth database for message W_θ. The queries are generated according to $\mathbb{H}$, which are independent of the realizations of the K messages. Therefore,

$$I(W_1, \ldots, W_K; Q_1^{[\theta]}, \ldots, Q_N^{[\theta]}) = 0. \tag{3.4}$$

To ensure that individual databases do not know which message is retrieved, we need to satisfy the following privacy constraint, $\forall n \in [N]$, $\forall \theta \in [K]$,

$$(Q_n^{[1]}, A_n^{[1]}, W_1, \ldots, W_K, \mathbb{H}_n) \sim (Q_n^{[\theta]}, A_n^{[\theta]}, W_1, \ldots, W_K, \mathbb{H}_n), \tag{3.5}$$

where $A \sim B$ means that A and B are identically distributed.

After receiving the query $Q_n^{[\theta]}$, the nth database replies with an answering string $A_n^{[\theta]}$, which is a function of $Q_n^{[\theta]}$ and all the K messages. Therefore, $\forall \theta \in [K], \forall n \in [N]$,

$$H(A_n^{[\theta]} | Q_n^{[\theta]}, W_1, \ldots, W_K) = 0. \tag{3.6}$$

After receiving the answering strings $A_1^{[\theta]}, \ldots, A_N^{[\theta]}$ from all the N databases, the user needs to decode the desired message W_θ reliably. By using Fano's inequality, we have the following reliability constraint

$$H\left(W_\theta | Z, \mathbb{H}, Q_1^{[\theta]}, \ldots, Q_N^{[\theta]}, A_1^{[\theta]}, \ldots, A_N^{[\theta]}\right) = o(L), \tag{3.7}$$

where $o(L)$ denotes a function such that $\frac{o(L)}{L} \to 0$ as $L \to \infty$.

For a fixed N, K, and caching ratio r, a pair $(D(r), L)$ is achievable if there exists a PIR scheme for message of size L bits long with partially known uncoded prefetching satisfying the privacy constraint (3.5) and the reliability constraint (3.7), where $D(r)$ represents the expected number of downloaded bits (over all the queries) from the N databases via the answering strings $A_{1:N}^{[\theta]}$, where $A_{1:N}^{[\theta]} = (A_1^{[\theta]}, \ldots, A_N^{[\theta]})$, i.e.,

$$D(r) = \sum_{n=1}^{N} H\left(A_n^{[\theta]}\right). \tag{3.8}$$

In this work, we aim at characterizing the optimal normalized download cost $D^*(r)$ corresponding to every caching ratio $0 \leq r \leq 1$, where

$$D^*(r) = \inf\left\{\frac{D(r)}{L} : (D(r), L) \text{ is achievable}\right\}, \tag{3.9}$$

which is a function of the caching ratio r.

3.3 Main Results

We provide a PIR scheme for general K, N and r, which achieves the following normalized download cost, $\bar{D}(r)$.

Theorem 3.1 (Outer bound) *In the cache-aided PIR with partially known un-*

coded prefetching, for the caching ratio

$$r_s = \frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^{i+1}}, \tag{3.10}$$

and length of the message

$$L(s) = N\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^{i+1}N, \tag{3.11}$$

where $s \in \{1, 2, \cdots, K-1\}$*, the optimal normalized download cost* $D^*(r_s)$ *is upper bounded by,*

$$D^*(r_s) \leq \bar{D}(r_s) = \frac{\sum_{i=0}^{K-1-s} \binom{K}{s+1+i}(N-1)^{i+1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^{i+1}}. \tag{3.12}$$

Moreover, if $r_s < r < r_{s+1}$*, and* $\alpha \in (0,1)$ *such that* $r = \alpha r_s + (1-\alpha)r_{s+1}$*, then*

$$D^*(r) \leq \bar{D}(r) = \alpha\bar{D}(r_s) + (1-\alpha)\bar{D}(r_{s+1}). \tag{3.13}$$

The proof of Theorem 3.1 is provided in Section 3.4. The outer bound in Theorem 3.1 is a piece-wise linear curve, which consists of K line segments. These K line segments intersect at the points r_s.

We characterize an inner bound (converse bound), which is denoted by $\tilde{D}(r)$, for the optimal normalized download cost $D^*(r)$ for general K, N, r.

Theorem 3.2 (Inner bound) *In the cache-aided PIR with partially known un-*

coded prefetching, the normalized download cost is lower bounded as,

$$
\begin{aligned}
D^*(r) &\geq \tilde{D}(r) \\
&= \max_{i\in\{2,\cdots,K+1\}} (1-r) \sum_{j=0}^{K+1-i} \frac{1}{N^j} - r\left(1-\frac{1}{N}\right) \sum_{j=0}^{K-i} \frac{K+1-i-j}{N^j} \qquad (3.14) \\
&= \max_{i\in\{2,\cdots,K+1\}} \sum_{j=0}^{K+1-i} \frac{1}{N^j} - (K+2-i)r. \qquad (3.15)
\end{aligned}
$$

The proof of Theorem 3.2 is provided in Section 3.5. The inner bound in Theorem 3.2 is also a piece-wise linear curve, which consists of K line segments. Interestingly, these K line segments intersect at the points as follows,

$$\tilde{r}_i = \frac{1}{N^{K-i}}, \quad i = 1, \cdots, K-1. \qquad (3.16)$$

The outer bounds provided in Theorem 3.1 and the inner bounds provided in Theorem 3.2 match for some caching ratios r as summarized in the following corollary.

Corollary 3.1 (Optimal tradeoff for very low and very high caching ratios) *In the cache-aided PIR with partially known uncoded prefetching, for very low caching ratios, i.e., for $r \leq \frac{1}{N^{K-1}}$, the optimal normalized download cost is given by,*

$$D^*(r) = \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}\right) - Kr. \qquad (3.17)$$

On the other hand, for very high caching ratios, i.e., for $r \geq \frac{K-2}{N^2-3N+KN}$, the optimal

normalized download cost is given by,

$$D^*(r) = \begin{cases} 1 + \frac{1}{N} - 2r, & \frac{K-2}{N^2-3N+KN} \leq r \leq \frac{1}{N} \\ 1 - r, & \frac{1}{N} \leq r \leq 1 \end{cases} . \tag{3.18}$$

Proof: From (3.10) and (3.16), we have

$$r_1 = \tilde{r}_1 = \frac{1}{N^{K-1}}, \tag{3.19}$$

$$r_{K-2} = \frac{K-2}{N^2-3N+KN}, \tag{3.20}$$

$$r_{K-1} = \tilde{r}_{K-1} = \frac{1}{N}. \tag{3.21}$$

For the outer bound of the case of very low caching ratios, from (3.12), we have

$$\bar{D}(r_1) = \frac{\sum_{i=0}^{K-2} \binom{K}{2+i}(N-1)^{i+1}}{\binom{K-2}{0} + \sum_{i=0}^{K-2} \binom{K-1}{1+i}(N-1)^{i+1}} \tag{3.22}$$

$$= \frac{\frac{1}{(N-1)} \sum_{i=0}^{K-2} \binom{K}{2+i}(N-1)^{i+2}}{N^{K-1}} \tag{3.23}$$

$$= \frac{\frac{1}{(N-1)} \left(N^K - 1 - K(N-1)\right)}{N^{K-1}} \tag{3.24}$$

$$= \frac{N^K - KN + K - 1}{N^K - N^{K-1}}. \tag{3.25}$$

For the inner bound of the case of very low caching ratios, from (3.14), by choosing $i = 2$ and using $r = r_1$, we have

$$\tilde{D}(r_1) \geq (1 - r_1) \sum_{j=0}^{K-1} \frac{1}{N^j} - r_1 \left(1 - \frac{1}{N}\right) \sum_{j=0}^{K-2} \frac{K-1-j}{N^j} \tag{3.26}$$

$$= \left(1 - \frac{1}{N^{K-1}}\right) \frac{1 - \frac{1}{N^K}}{1 - \frac{1}{N}} - \frac{1}{N^{K-1}} \left(1 - \frac{1}{N}\right) \frac{K - \frac{K}{N} - 1 + \frac{1}{N^K}}{\left(1 - \frac{1}{N}\right)^2} \tag{3.27}$$

$$= \frac{1}{\left(1 - \frac{1}{N}\right)} \left[\left(1 - \frac{1}{N^{K-1}}\right) \left(1 - \frac{1}{N^K}\right) - \frac{1}{N^{K-1}} \left(K - \frac{K}{N} - 1 + \frac{1}{N^K}\right) \right] \tag{3.28}$$

$$= \frac{N^K - KN + K - 1}{N^K - N^{K-1}} = \bar{D}(r_1). \tag{3.29}$$

Thus, since $\tilde{D}(r_1) \leq \bar{D}(r_1)$ by definition, (3.29) implies $\tilde{D}(r_1) = \bar{D}(r_1)$.

For the outer bound of the case of very high caching ratios, from (3.12), we have

$$\bar{D}(r_{K-2}) = \frac{\sum_{i=0}^{1} \binom{K}{K-1+i} (N-1)^{i+1} N}{N\binom{K-2}{K-3} + \sum_{i=0}^{1} \binom{K-1}{K-2+i} (N-1)^{i+1} N} \tag{3.30}$$

$$= \frac{N^2 + KN - 2N - K + 1}{N^2 - 3N + KN}, \tag{3.31}$$

and for the inner bound of the case of very high caching ratios, from (3.14) by choosing $i = K$ and using $r = r_{K-2}$,

$$\tilde{D}(r_{K-2}) \geq (1 - r_{K-2}) \sum_{j=0}^{1} \frac{1}{N^j} - r_{K-2} \left(1 - \frac{1}{N}\right) \sum_{j=0}^{0} \frac{1-j}{N^j} \tag{3.32}$$

$$= 1 + \frac{1}{N} - 2r_{K-2} \tag{3.33}$$

$$= \frac{N^2 + KN - 2N - K + 1}{N^2 - 3N + KN} = \bar{D}(r_{K-2}) \tag{3.34}$$

implying $\tilde{D}(r_{K-2}) = \bar{D}(r_{K-2})$.

Finally, from (3.12), $\bar{D}(r_{K-1}) = \frac{N-1}{N}$, and from (3.14) by choosing $i = K+1$ and using $r = r_{K-1}$,

$$\tilde{D}(r_{K-1}) \geq \frac{N-1}{N} = \bar{D}(r_{K-1}) \tag{3.35}$$

implying $\tilde{D}(r_{K-1}) = \bar{D}(r_{K-1})$.

Therefore, $\tilde{D}(r) = \bar{D}(r)$ at $r = r_1$, $r = r_{K-2}$ and $r = r_{K-1}$. In addition to that $\tilde{D}(0) = \bar{D}(0)$ and $\tilde{D}(1) = \bar{D}(1)$. Since both $\bar{D}(r)$ and $\tilde{D}(r)$ are linear functions of r, and since $\tilde{D}(0) = \bar{D}(0)$ and $\tilde{D}(r_1) = \bar{D}(r_1)$, we have $\tilde{D}(r) = \bar{D}(r) = D^*(r)$ for $0 \leq r \leq r_1$. This is the very low caching ratio region. In addition, since $\tilde{D}(r_{K-2}) = \bar{D}(r_{K-2})$, $\tilde{D}(r_{K-1}) = \bar{D}(r_{K-1})$ and $\tilde{D}(1) = \bar{D}(1)$, we have $\tilde{D}(r) = \bar{D}(r) = D^*(r)$ for $r_{K-2} \leq r \leq 1$. This is the very high caching ratio region. ■

We use the example of $K = 4$, $N = 2$ to illustrate Corollary 3.1 (see Figure 3.2). In this case, $r_1 = \tilde{r}_1 = \frac{1}{8}$, $r_{K-2} = \frac{1}{3}$, and $r_{K-1} = \tilde{r}_{K-1} = \frac{1}{2}$. Therefore, we have exact results for $0 \leq r \leq \frac{1}{8}$ (very low caching ratios) and $\frac{1}{3} \leq r \leq 1$ (very high caching ratios). We have a gap between the achievability and the converse for medium caching ratios in $\frac{1}{8} \leq r \leq \frac{1}{3}$. More specifically, line segments connecting $(0, \frac{15}{8})$ and $(\frac{1}{8}, \frac{11}{8})$; connecting $(\frac{1}{3}, \frac{5}{6})$ and $(\frac{1}{2}, \frac{1}{2})$; and connecting $(\frac{1}{2}, \frac{1}{2})$ and $(1, 0)$ are tight.

For the case $K = 3$, we have exact tradeoff curve for any N, r as shown in the following corollary.

Corollary 3.2 (Optimal tradeoff for $K = 3$) *In the cache-aided PIR with partially known uncoded prefetching with $K = 3$ messages, the optimal download cost*

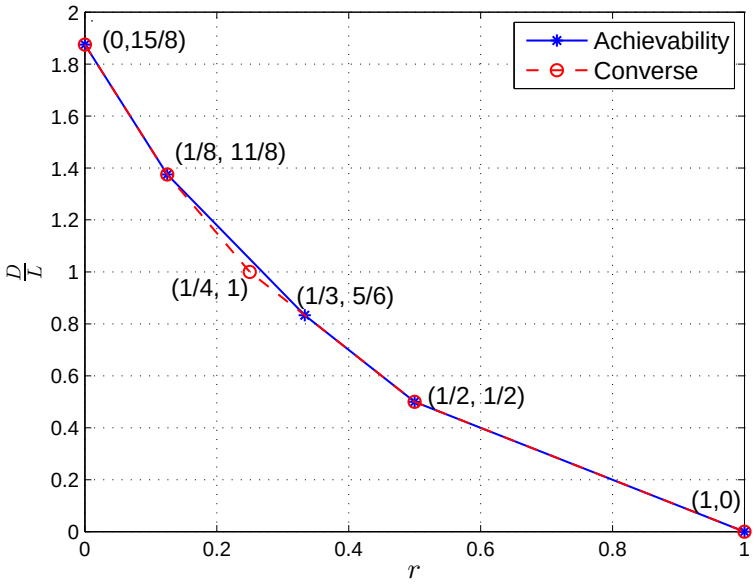

Figure 3.2: Inner and outer bounds for $K = 4$, $N = 2$.

caching ratio tradeoff is given explicitly as,

$$D^*(r) = \begin{cases} 1 + \frac{1}{N} + \frac{1}{N^2} - 3r, & 0 \leq r \leq \frac{1}{N^2} \\ 1 + \frac{1}{N} - 2r, & \frac{1}{N^2} \leq r \leq \frac{1}{N} \\ 1 - r, & \frac{1}{N} \leq r \leq 1 \end{cases} . \tag{3.36}$$

Proof: The proof follows from the proof of Corollary 3.1. Note that in this case, from (3.19) and (3.20), $r_1 = r_{K-2} = \frac{1}{N^2}$; and from (3.21), $r_2 = r_{K-1} = \frac{1}{N}$. Thus, we have a tight result for $0 \leq r \leq r_1 = \frac{1}{N^2}$ (very low caching ratios) and a tight result for $r_{K-2} = r_1 = \frac{1}{N^2} \leq r \leq 1$, i.e., a tight result for all $0 \leq r \leq 1$. We have three segments in this case: $[0, r_1]$, $[r_1, r_2]$ and $[r_2, 1]$ with three different line expressions for the exact result as given in (3.10)-(3.12) and written explicitly in (3.36). ■

3.4 Achievable Scheme

In this section, we present an achievable scheme for the outer bounds provided in Theorem 3.1. Our achievable scheme is based on [12, 28] and Chapter 2. We first provide achievable schemes for the caching ratios r_s in (3.10) by applying the principles in [12]: 1) database symmetry, 2) message symmetry within each database, and 3) exploiting undesired messages as side information. For an arbitrary caching ratio $r \neq r_s$, we apply the memory-sharing scheme in [28]. Since the cached content is partially known by the databases, the achievable scheme is different from that in Chapter 2. We first use the case of $K = 3$, $N = 2$ to illustrate the main ideas of our achievability scheme.

3.4.1 Motivating Example: $K = 3$ Messages and $N = 2$ Databases

We permute the bits of messages W_1, W_2, W_3 randomly and independently, and use a_i, b_i, and c_i to denote the bits of each permuted message, respectively. We assume that the user wants to retrieve message W_1 privately without loss of generality.

3.4.1.1 Caching Ratio $r_1 = \frac{1}{4}$

We choose the message size as 8 bits. In the prefetching phase, for caching ratio $r_1 = \frac{1}{4}$, the user caches 2 bits from each message. Therefore, the user caches 1 bit from each database for each message. Therefore, $Z_1 = (a_1, b_1, c_1)$ and $Z_2 = (a_2, b_2, c_2)$.

In the retrieval phase, for $s = 1$, we first mix 1 bit of side information with the desired bit. Therefore, the user queries $a_3 + b_2$ and $a_4 + c_2$ from database 1.

Note that database 1 knows that the user has prefetched Z_1. Therefore, the user does not use side information Z_1 to retrieve information from database 1. To keep message symmetry, the user further queries $b_3 + c_3$ from database 1. Similarly, the user queries $a_5 + b_1$, $a_6 + c_1$ and $b_4 + c_4$ from database 2. Then, the user exploits the side information $b_4 + c_4$ to query $a_7 + b_4 + c_4$ from database 1 and the side information $b_3 + c_3$ to query $a_8 + b_3 + c_3$ from database 2. After this step, no more side information can be used and the message symmetry is attained for each database. Therefore, the PIR scheme ends here. The decodability of message W_1 can be shown easily, since the desired bits are either mixed with cached side information or the side information obtained from the other database. Specifically, for the downloaded bits from database 1, the user can decode a_3 and a_4 from $a_3 + b_2$ and $a_4 + c_2$, since b_2 and c_2 are in the cache. The user can decode a_7 from $a_7 + b_4 + c_4$, since $b_4 + c_4$ is the side information obtained from database 2. A similar decoding procedure applies to the downloaded bits from database 2. Overall, the user downloads 8 bits. Therefore, the normalized download cost is 1. We summarize the queries in Table. 3.1.

Table 3.1: Query table for $K = 3$, $N = 2$, $r_1 = \frac{1}{4}$.

s	DB1	DB2
$s = 1$	$a_3 + b_2$	$a_5 + b_1$
	$a_4 + c_2$	$a_6 + c_1$
	$b_3 + c_3$	$b_4 + c_4$
	$a_7 + b_4 + c_4$	$a_8 + b_3 + c_3$

$Z_1 = (a_1, b_1, c_1)$	$Z_2 = (a_2, b_2, c_2)$

3.4.1.2 Caching Ratio $r_2 = \frac{1}{2}$

We choose the message size as 4 bits. In the prefetching phase, for caching ratio $r_2 = \frac{1}{2}$, the user caches 2 bits from each message. Therefore, the user caches 1 bit from each database for each message. Therefore, $Z_1 = (a_1, b_1, c_1)$ and $Z_2 = (a_2, b_2, c_2)$. In the retrieval phase, for $s = 2$, we first mix 2 bits of side information with the desired bit. Therefore, the user queries $a_3 + b_2 + c_2$ from database 1. Similarly, the user queries $a_4 + b_1 + c_1$ from database 2. After this, no more side information can be used and the message symmetry is attained for each database. Therefore, the PIR scheme ends here. The user can decode a_3 and a_4 from $a_3 + b_2 + c_2$ and $a_4 + b_1 + c_1$, since b_1, b_2, c_1 and c_2 are in the cache. Overall, the user downloads 2 bits. Therefore, the normalized download cost is $\frac{1}{2}$. We summarize the queries in Table. 3.2.

Table 3.2: Query table for $K = 3$, $N = 2$, $r_2 = \frac{1}{2}$.

s	DB1	DB2
$s = 2$	$a_3 + b_2 + c_2$	$a_4 + b_1 + c_1$
	$Z_1 = (a_1, b_1, c_1)$	$Z_2 = (a_2, b_2, c_2)$

3.4.1.3 Caching Ratio $r = \frac{1}{3}$

We choose the message size as 12 bits. In the prefetching phase, for caching ratio $r = \frac{1}{3}$, the user caches 4 bits from each message. Therefore, the user caches 2 bits from each database for each message. Therefore, $Z_1 = (a_1, a_2, b_1, b_2, c_1, c_2)$ and $Z_2 = (a_3, a_4, b_3, b_4, c_3, c_4)$. In the retrieval phase, we combine the achievable schemes

in Section 3.4.1.1 and 3.4.1.2 as shown in Table 3.3. The normalized download cost is $\frac{5}{6}$. By applying [28, Lemma 1] and taking $\alpha = \frac{2}{3}$, we can show that $\bar{D}(\frac{1}{3}) = \bar{D}(\frac{2}{3} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{1}{2}) = \frac{2}{3}\bar{D}(\frac{1}{4}) + \frac{1}{3}\bar{D}(\frac{1}{2}) = \frac{2}{3} \cdot 1 + \frac{1}{3} \cdot \frac{1}{2} = \frac{5}{6}$.

Table 3.3: Query table for $K = 3$, $N = 2$, $r = \frac{1}{3}$.

s	DB1	DB2
$s = 1$	$a_5 + b_3$	$a_7 + b_1$
	$a_6 + c_3$	$a_8 + c_1$
	$b_5 + c_5$	$b_6 + c_6$
	$a_9 + b_6 + c_6$	$a_{10} + b_5 + c_5$
$s = 2$	$a_{11} + b_4 + c_4$	$a_{12} + b_2 + c_2$

$Z_1 = (a_1, a_2, b_1, b_2, c_1, c_2)$	$Z_2 = (a_3, a_4, b_3, b_4, c_3, c_4)$

3.4.2 Achievable Scheme

We first present the achievable scheme for the caching ratios r_s given in (3.10). Then, we apply the memory-sharing scheme provided in [28] for the intermediate caching ratios.

3.4.2.1 Achievable Scheme for the Caching Ratio r_s

For fixed K and N, there are $K - 1$ non-degenerate corner points (in addition to degenerate caching ratios $r = 0$ and $r = 1$). The caching ratios, r_s, corresponding to these non-degenerate corner points are indexed by s, which represents the number of cached bits used in the side information mixture at the first round of the querying. For each $s \in \{1, 2, \ldots, K - 1\}$, we choose the length of the message to be $L(s)$ for

the corner point indexed by s, where

$$L(s) = N\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^{i+1}N. \tag{3.37}$$

In the prefetching phase, for each message the user randomly and independently chooses $N\binom{K-2}{s-1}$ bits to cache, and caches $\binom{K-2}{s-1}$ bits from each database for each message. Therefore, the caching ratio r_s is equal to

$$r_s = \frac{N\binom{K-2}{s-1}}{N\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^{i+1}N}. \tag{3.38}$$

In the retrieval phase, the user applies the PIR scheme in Algorithm 1.

Since the desired bits are added to the side information which is either obtained from the cached bits (if $t = s+1$) or from the remaining $(N-1)$ databases in the $(t-1)$th round when $t > s+1$, the user can decode the uncached portion of the desired message by canceling out the side information bits. In addition, for each database, each message is queried equally likely with the same set of equations, which guarantees privacy as in [12]. Therefore, the privacy constraint in (3.5) and the reliability constraint in (3.7) are satisfied.

We now calculate the total number of downloaded bits for the caching ratio r_s in (3.38). For the round $t = s+1$, we exploit s cached bits to form the side information equation. Therefore, each download is a sum of $s+1$ bits. For each database, we utilize the side information cached from other $N-1$ databases. In addition to the message symmetry step enforcing symmetry across K messages, we

Algorithm 1 PIR scheme

1. *Initialization:* Set the round index to $t = s + 1$, where the tth round involves downloading sums of every t combinations of the K messages.

2. *Exploiting side information:*

   if $t = s + 1$, **then** for the first database, the user forms queries by mixing s undesired bits cached from the other $N-1$ databases in the prefetching phase to form one side information equation. Each side information equation is added to one bit from the uncached portion of the desired message. Therefore, for the first database, the user downloads $\binom{K-1}{s}(N-1)$ equations in the form of a desired bit added to a mixture of s cached bits from other messages.

   else if $t > s + 1$, **then** for the first database, the user exploits the $\binom{K-1}{t-1}(N-1)^{t-s}$ side information equations generated from the remaining $(N-1)$ databases in the $(t-1)$th round.

3. *Symmetry across databases:* The user downloads the same number of equations with the same structure as in step 2 from every database. Consequently, the user decodes $\binom{K-1}{t-1}(N-1)^{t-s}$ desired bits from every database, which are done either using the cached bits as side information if $t = s + 1$, or the side information generated in the $(t-1)$th round if $t > s + 1$.

4. *Message symmetry:* To satisfy the privacy constraint, the user should download the same amount of bits from other messages. Therefore, the user downloads $\binom{K-1}{t}(N-1)^{t-s}$ undesired equations from each database in the form of sum of t bits from the uncached portion of the undesired messages.

5. *Repeat* steps 2, 3, 4 after setting $t = t + 1$ until $t = K$.

6. *Shuffling the order of queries:* By shuffling the order of queries uniformly, all possible queries can be made equally likely regardless of the message index.

download $\binom{K}{s+1}(N-1)$ bits from a database. Due to the database symmetry step, in total, we download $\binom{K}{s+1}(N-1)N$ bits. For the round $t = s+i > s+1$, we exploit $s+i-1$ undesired bits downloaded from the $(t-1)$th round to form the side information equation. Due to message symmetry and database symmetry, we download $\binom{K}{s+1+i}(N-1)^{i+1}N$ bits. Overall, the total number of downloaded bits is,

$$D(r_s) = \sum_{i=0}^{K-1-s} \binom{K}{s+1+i}(N-1)^{i+1}N. \tag{3.39}$$

By canceling out the undesired side information bits using the cached bits for the round $t = s+1$, we obtain $\binom{K-1}{s}(N-1)N$ desired bits. For the round $t = s+i > s+1$, we decode $\binom{K-1}{s+i}(N-1)^{i+1}N$ desired bits by using the side information obtained in $(t-1)$th round. Overall, we obtain $L(s) - N\binom{K-2}{s-1}$ desired bits. Therefore, the normalized download cost is,

$$\begin{aligned}\bar{D}(r_s) &= \frac{D(r_s)}{L(s)} \\ &= \frac{\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^{i+1}N}{N\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^{i+1}N}.\end{aligned} \tag{3.40}$$

3.4.2.2 Achievable Scheme for the Caching Ratios not Equal to r_s

For caching ratios r which are not exactly equal to (3.38) for some s, we first find an s such that $r_s < r < r_{s+1}$. We choose $0 < \alpha < 1$ such that $r = \alpha r_s + (1-\alpha)r_{s+1}$. By using the memory-sharing scheme in [28, Lemma 1], we achieve the following

normalized download cost,

$$\bar{D}(r) = \alpha \bar{D}(r_s) + (1-\alpha)\bar{D}(r_{s+1}). \tag{3.41}$$

3.5 Converse Proof

In this section, we derive an inner bound for the cache-aided PIR with partially known uncoded prefetching. We extend the techniques in [12] and Chapter 2 to our problem. The main difference between this proof and that in Chapter 2 is the usage of privacy constraint given in (3.5).

Lemma 3.1 (Interference lower bound) *For the cache-aided PIR with partially known uncoded prefetching, the interference from undesired messages within the answering strings $D(r) - L(1-r)$ is lower bounded by,*

$$D(r) - L(1-r) + o(L) \geq I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right) \tag{3.42}$$

for all $k \in \{2, \ldots, K\}$.

The proof of Lemma 3.1 is similar to Lemma 2.1 in Chapter 2. In the following lemma, we prove an inductive relation for the mutual information term on the right hand side of (3.42).

Lemma 3.2 (Induction lemma) *For all $k \in \{2, \ldots, K\}$, the mutual information*

term in Lemma 3.1 can be inductively lower bounded as,

$$\begin{aligned} & I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right) \\ & \geq \frac{1}{N} I\left(W_{k+1:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k}, Z, \mathbb{H}\right) \\ & \quad + \frac{L(1-r)}{N} + \frac{1-N}{N}(K-k+1)Lr - o(L). \end{aligned} \tag{3.43}$$

Lemma 3.2 is a generalization of [12, Lemma 6] and Lemma 2.2 in Chapter 2, and it reduces to [12, Lemma 6] when $r = 0$. Compared to Lemma 2.2 in Chapter 2, the lower bound in (3.43) is increased by $\frac{(K-k+1)Lr}{N}$, since the cached content is partially known by the databases.

Proof: We start with the left hand side of (3.43),

$$\begin{aligned} & I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right) \\ & \quad = I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}, Z, \mathbb{H} | W_{1:k-1}\right) - I(W_{k:K}; Z, \mathbb{H} | W_{1:k-1}). \end{aligned} \tag{3.44}$$

For the first term on the right hand side of (3.44), we have

$$I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}, Z, \mathbb{H} | W_{1:k-1}\right)$$

$$= \frac{1}{N} N I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}, Z, \mathbb{H} | W_{1:k-1}\right) \tag{3.45}$$

$$\geq \frac{1}{N} \sum_{n=1}^{N} I\left(W_{k:K}; Q_{n}^{[k-1]}, A_{n}^{[k-1]}, Z_n, \mathbb{H}_n | W_{1:k-1}\right) \tag{3.46}$$

$$= \frac{1}{N}\left[\sum_{n=1}^{N} I\left(W_{k:K}; Q_{n}^{[k-1]}, A_{n}^{[k-1]} | W_{1:k-1}, Z_n, \mathbb{H}_n\right) + \sum_{n=1}^{N} I\left(W_{k:K}; Z_n, \mathbb{H}_n | W_{1:k-1}\right)\right] \tag{3.47}$$

$$= \frac{1}{N}\left[\sum_{n=1}^{N} I\left(W_{k:K}; Q_n^{[k-1]}, A_n^{[k-1]} | W_{1:k-1}, Z_n, \mathbb{H}_n\right) + N \times \frac{(K-k+1)Lr}{N}\right] \tag{3.48}$$

$$\overset{(3.5)}{=} \frac{1}{N}\sum_{n=1}^{N} I\left(W_{k:K}; Q_n^{[k]}, A_n^{[k]} | W_{1:k-1}, Z_n, \mathbb{H}_n\right) + \frac{(K-k+1)Lr}{N} \tag{3.49}$$

$$\overset{(3.3),(3.4)}{=} \frac{1}{N}\sum_{n=1}^{N} I\left(W_{k:K}; A_n^{[k]} | W_{1:k-1}, Z_n, \mathbb{H}_n, Q_n^{[k]}\right) + \frac{(K-k+1)Lr}{N} \tag{3.50}$$

$$\overset{(3.6)}{=} \frac{1}{N}\sum_{n=1}^{N} H\left(A_n^{[k]} | W_{1:k-1}, Z_n, \mathbb{H}_n, Q_n^{[k]}\right) + \frac{(K-k+1)Lr}{N} \tag{3.51}$$

$$\geq \frac{1}{N}\sum_{n=1}^{N} H\left(A_n^{[k]} | W_{1:k-1}, Z, \mathbb{H}, Q_{1:N}^{[k]}, A_{1:n-1}^{[k]}\right) + \frac{(K-k+1)Lr}{N} \tag{3.52}$$

$$\overset{(3.6)}{=} \frac{1}{N}\sum_{n=1}^{N} I\left(W_{k:K}; A_n^{[k]} | W_{1:k-1}, Z, \mathbb{H}, Q_{1:N}^{[k]}, A_{1:n-1}^{[k]}\right) + \frac{(K-k+1)Lr}{N} \tag{3.53}$$

$$= \frac{1}{N} I\left(W_{k:K}; A_{1:N}^{[k]} | W_{1:k-1}, Z, \mathbb{H}, Q_{1:N}^{[k]}\right) + \frac{(K-k+1)Lr}{N} \tag{3.54}$$

$$\overset{(3.3),(3.4)}{=} \frac{1}{N} I\left(W_{k:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k-1}, Z, \mathbb{H}\right) + \frac{(K-k+1)Lr}{N} \tag{3.55}$$

$$\overset{(3.7)}{=} \frac{1}{N} I\left(W_{k:K}; W_k, Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k-1}, Z, \mathbb{H}\right) + \frac{(K-k+1)Lr}{N} - o(L) \tag{3.56}$$

$$= \frac{(K-k+1)Lr}{N} + \frac{1}{N}\Bigg[I\left(W_{k:K}; W_k | W_{1:k-1}, Z, \mathbb{H}\right) + I\left(W_{k:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k}, Z, \mathbb{H}\right)\Bigg] - o(L) \tag{3.57}$$

$$= \frac{(K-k+1)Lr}{N} + \frac{L(1-r)}{N} + \frac{1}{N} I\left(W_{k+1:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k}, Z, \mathbb{H}\right) - o(L), \tag{3.58}$$

where (3.46) follows from the non-negativity of mutual information, (3.48) is due to the fact that from the nth database, the user prefetches $\frac{KLr}{N}$ bits, (3.49) follows from the privacy constraint, (3.50) and (3.55) follow from the independence of $W_{k:K}$ and $Q_n^{[k]}$, (3.51) and (3.53) follow from the fact that the answering string $A_n^{[k]}$ is

a deterministic function of $(W_{1:K}, Q_n^{[k]})$, (3.52) follows from conditioning reduces entropy, and (3.56) follows from the reliability constraint.

For the second term on the right hand side of (3.44), we have

$$\begin{aligned} & I(W_{k:K}; Z, \mathbb{H}|W_{1:k-1}) \\ & \quad = H\left(W_{k:K}|W_{1:k-1}\right) - H(W_{k:K}|W_{1:k-1}, Z, \mathbb{H}) && (3.59) \\ & \quad = (K-k+1)\,Lr && (3.60) \end{aligned}$$

where (3.60) follows from the uncoded nature of the cached bits.

Combining (3.44), (3.58) and (3.60) yields (3.43). ■

Now, we are ready to derive the general inner bound for arbitrary K, N, r. To obtain this bound, we use Lemma 3.1 to find K lower bounds by varying the index k in the lemma from $k = 2$ to $k = K$, and by using the non-negativity of mutual information for the Kth bound. Next, we inductively lower bound each term of Lemma 3.1 by using Lemma 3.2 $(K - k + 1)$ times to get K explicit lower bounds.

Lemma 3.3 *For fixed N, K and r, we have*

$$D(r) \geq L(1-r) \sum_{j=0}^{K+1-k} \frac{1}{N^j} - Lr\left(1 - \frac{1}{N}\right) \sum_{j=0}^{K-k} \frac{K+1-k-j}{N^j} + o(L), \qquad (3.61)$$

where $k = 2, \ldots, K+1$.

Proof: We have

$$D(r) \overset{(3.42)}{\geq} I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right) + L(1-r) - o(L) \tag{3.62}$$

$$\overset{(3.43)}{\geq} \frac{1}{N} I\left(W_{k+1:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k}, Z, \mathbb{H}\right) + L(1-r)\left(1+\frac{1}{N}\right) - Lr\left(1-\frac{1}{N}\right)(K-k+1) - o(L) \tag{3.63}$$

$$\overset{(3.43)}{\geq} \frac{1}{N^2} I\left(W_{k+2:K}; Q_{1:N}^{[k+1]}, A_{1:N}^{[k+1]} | W_{1:k+1}, Z, \mathbb{H}\right) + L(1-r)\left[1+\frac{1}{N}+\frac{1}{N^2}\right] - Lr\left(1-\frac{1}{N}\right)\left[(K-k+1)+\frac{(K-k)}{N}\right] - o(L) \tag{3.64}$$

$$\overset{(3.43)}{\geq} \cdots \tag{3.65}$$

$$\overset{(3.43)}{\geq} L(1-r)\sum_{j=0}^{K+1-k}\frac{1}{N^j} - Lr\left(1-\frac{1}{N}\right)\sum_{j=0}^{K-k}\frac{K+1-k-j}{N^j} + o(L), \tag{3.66}$$

where (3.62) follows from Lemma 3.1, and the remaining steps follow from the successive application of Lemma 3.2. ■

We conclude the converse proof by dividing by L and taking the limit as $L \to \infty$. Then, for $k = 2, \cdots, K+1$, we have

$$D^*(r) \geq (1-r)\sum_{j=0}^{K+1-k}\frac{1}{N^j} - r\left(1-\frac{1}{N}\right)\sum_{j=0}^{K-k}\frac{K+1-k-j}{N^j}. \tag{3.67}$$

Since (3.67) gives K intersecting line segments, the normalized download cost is lower bounded by their maximum value as follows

$$D^*(r) \geq \max_{i\in\{2,\cdots,K+1\}} (1-r)\sum_{j=0}^{K+1-i}\frac{1}{N^j} - r\left(1-\frac{1}{N}\right)\sum_{j=0}^{K-i}\frac{K+1-i-j}{N^j}. \tag{3.68}$$

3.6 Further Examples

3.6.1 $K = 4$ Messages, $N = 2$ Databases

For $K = 4$ and $N = 2$, we present achievable PIR schemes for caching ratios $r_1 = \frac{1}{8}$ in Table 3.4, $r_2 = \frac{1}{3}$ in Table 3.5, and $r_3 = \frac{1}{2}$ in Table 3.6. The PIR schemes aim to retrieve message W_1, where we use a_i to denote its bits. The achievable normalized download costs for these caching ratios are $\frac{11}{8}$, $\frac{5}{6}$ and $\frac{1}{2}$, respectively. The plot of the inner and outer bounds can be found in Figure 3.2.

Table 3.4: Query table for $K = 4$, $N = 2$ and $r_1 = \frac{1}{8}$.

s	DB1	DB2
$s = 1$	$a_3 + b_2$	$a_6 + b_1$
	$a_4 + c_2$	$a_7 + c_1$
	$a_5 + d_2$	$a_8 + d_1$
	$b_3 + c_3$	$b_5 + c_5$
	$b_4 + d_3$	$b_6 + d_5$
	$c_4 + d_4$	$c_6 + d_6$
	$a_9 + b_5 + c_5$	$a_{12} + b_3 + c_3$
	$a_{10} + b_6 + d_5$	$a_{13} + b_4 + d_3$
	$a_{11} + c_6 + d_6$	$a_{14} + c_4 + d_4$
	$b_7 + c_7 + d_7$	$b_8 + c_8 + d_8$
	$a_{15} + b_8 + c_8 + d_8$	$a_{16} + b_7 + c_7 + d_7$
	$Z_1 = (a_1, b_1, c_1, d_1)$	$Z_2 = (a_2, b_2, c_2, d_2)$

3.6.2 $K = 5$, $K = 10$ and $K = 100$ Messages, $N = 2$ Databases

For $N = 2$, we show the numerical results for the inner and outer bounds for $K = 5$, $K = 10$ and $K = 100$ in Figures 3.3, 3.4 and 3.5. For fixed N as K grows, the gap between the achievable bound and converse bound increases. This observation will

Table 3.5: Query table for $K = 4$, $N = 2$, $r_2 = \frac{1}{3}$.

s	DB1	DB2
$s = 2$	$a_5 + b_3 + c_3$	$a_8 + b_1 + c_1$
	$a_6 + d_3 + b_4$	$a_9 + d_1 + b_2$
	$a_7 + c_4 + d_4$	$a_{10} + c_2 + d_2$
	$b_5 + c_5 + d_5$	$b_6 + c_6 + d_6$
	$a_{11} + b_6 + c_6 + d_6$	$a_{12} + b_5 + c_5 + d_5$

$Z_1 = \begin{pmatrix} a_1, a_2, b_1, b_2, \\ c_1, c_2, d_1, d_2 \end{pmatrix}$	$Z_2 = \begin{pmatrix} a_3, a_4, b_3, b_4, \\ c_3, c_4, d_3, d_4 \end{pmatrix}$

Table 3.6: Query table for $K = 4$, $N = 2$, $r_3 = \frac{1}{2}$.

s	DB1	DB2
$s = 3$	$a_3 + b_2 + c_2 + d_2$	$a_4 + b_1 + c_1 + d_1$

$Z_1 = (a_1, b_1, c_1, d_1)$	$Z_2 = (a_2, b_2, c_2, d_2)$

be made specific in Section 3.7.

3.7 Gap Analysis

In this section, we analyze the gap between the achievable bounds given in (3.12) and the converse bounds given in (3.14). We first observe that for fixed number of databases N, as the number of messages K increases, the achievable normalized download cost increases, and for large enough caching ratios $r \geq \frac{1}{N}$, the PIR schemes for different number of messages share the same normalized download cost $1 - r$. In addition to the monotonicity, the achievable normalized download cost for $K + 1$ messages has a special relationship with the achievable normalized download cost for K messages. We first use an example to illustrate this property. For $N = 2$, $K = 3$, $K = 4$, and $K = 5$, the achievable bounds are shown in Figure 3.6. The achievable bound for $K = 5$ is above the achievable bound for $K = 4$, and the achievable

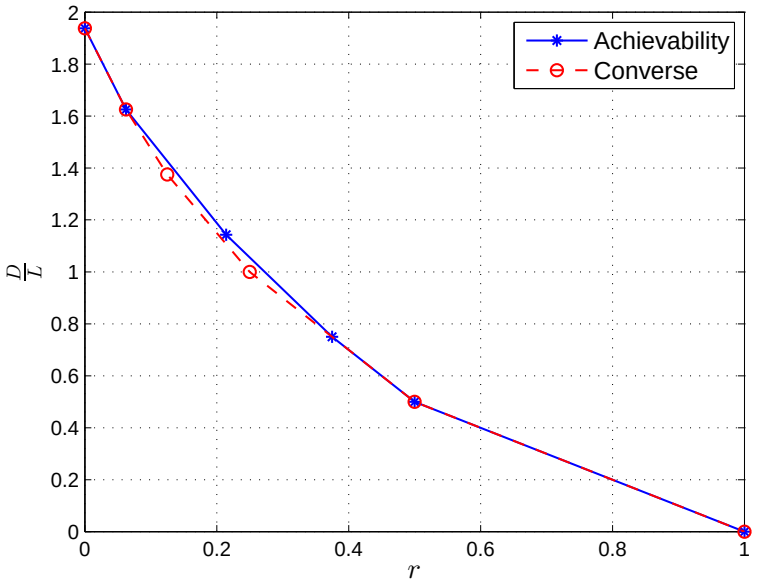

Figure 3.3: Inner and outer bounds for $K = 5$, $N = 2$.

bound for $K = 4$ is above the achievable bound for $K = 3$. By denoting $r_s^{(K)}$ as the caching ratio with total K messages and parameter s (see (3.10)), we observe that $(r_1^{(5)}, \bar{D}(r_1^{(5)}))$ falls on the line connecting $(r_0^{(4)}, \bar{D}(r_0^{(4)}))$ and $(r_1^{(4)}, \bar{D}(r_1^{(4)}))$. This observation is general, $(r_s^{(K+1)}, \bar{D}(r_s^{(K+1)}))$ falls on the line connecting $(r_{s-1}^{(K)}, \bar{D}(r_{s-1}^{(K)}))$ and $(r_s^{(K)}, \bar{D}(r_s^{(K)}))$. We summarize this result in the following lemma.

Lemma 3.4 (Monotonicity of the achievable bounds) *In cache-aided PIR with partially known uncoded prefetching, for fixed number of databases N, if the number of messages K increases, then the achievable normalized download cost increases. Furthermore, we have*

$$r_s^{(K+1)} = \alpha r_{s-1}^{(K)} + (1-\alpha) r_s^{(K)}, \tag{3.69}$$

$$\bar{D}(r_s^{(K+1)}) = \alpha \bar{D}(r_{s-1}^{(K)}) + (1-\alpha) \bar{D}(r_s^{(K)}), \tag{3.70}$$

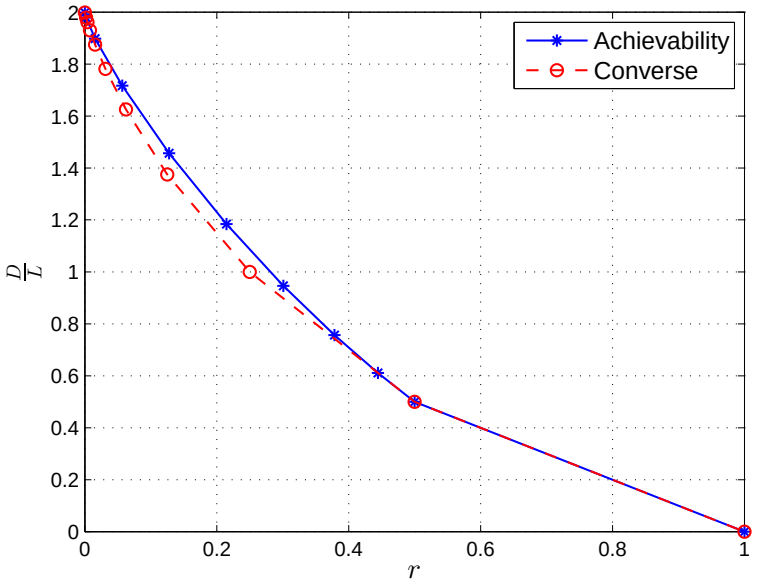

Figure 3.4: Inner and outer bounds for $K = 10$, $N = 2$.

where $0 \leq \alpha \leq 1$.

The proof of Lemma 3.4 is similar to Chapter 2[Lemma 4].

After showing the monotonicity of the achievable bounds, we show that as $K \to \infty$, the asymptotic upper bound for the achievable bounds is given as in the following lemma. With this asymptotic upper bound, we conclude that the worst-case additive gap is $\frac{5}{32}$.

Lemma 3.5 (Asymptotics and the worst-case additive gap) *In cache-aided PIR with partially known uncoded prefetching, as* $K \to \infty$*, the outer bound is upper bounded by,*

$$\bar{D}(r) \leq \frac{N}{N-1}(1-r)^2 \tag{3.71}$$

Hence, the worst-case additive gap is $\frac{5}{32}$.

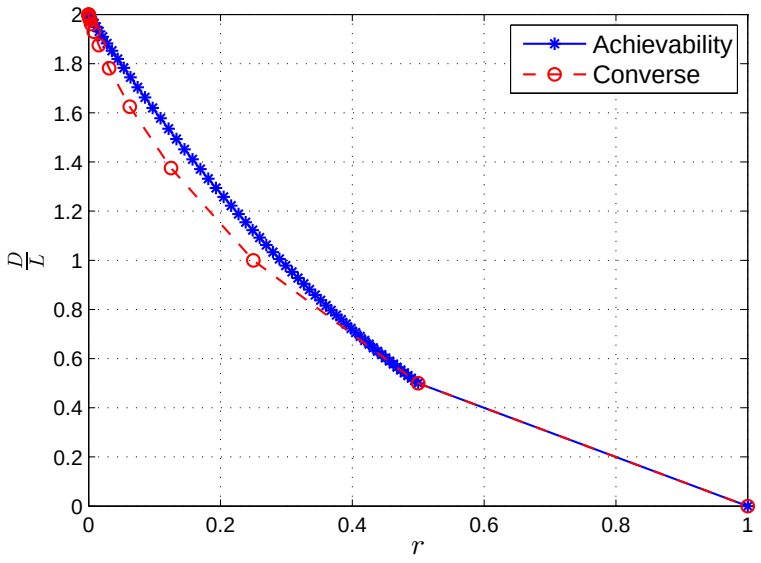

Figure 3.5: Inner and outer bounds for $K = 100$, $N = 2$.

The detailed proof of Lemma 3.5 is provided in Section 3.10. We note that the outer bound is monotonically increasing in K. Therefore, we first derive an asymptotic upper bound as $K \to \infty$ for the outer bound as in (3.71). Then, we show that most of the K inner bounds concentrate around $r = 0$. Therefore, we only need to consider a small number of the inner bounds for the worst-case gap analysis.

3.8 Comparisons with Other Cache-Aided PIR Models

In this section, we compare the normalized download costs between different cache-aided PIR models subjected to same memory size constraint. We first use an example of $N = 2$ and $K = 12$ (see Figure 3.7) to show the relative normalized download costs for different models. In [32, 33], the user caches M full messages out of total K messages. In order to compare with other cache-aided PIR schemes, we use $\frac{M}{K}$ as the caching ratio. Since the PIR schemes are only reported for the corner points

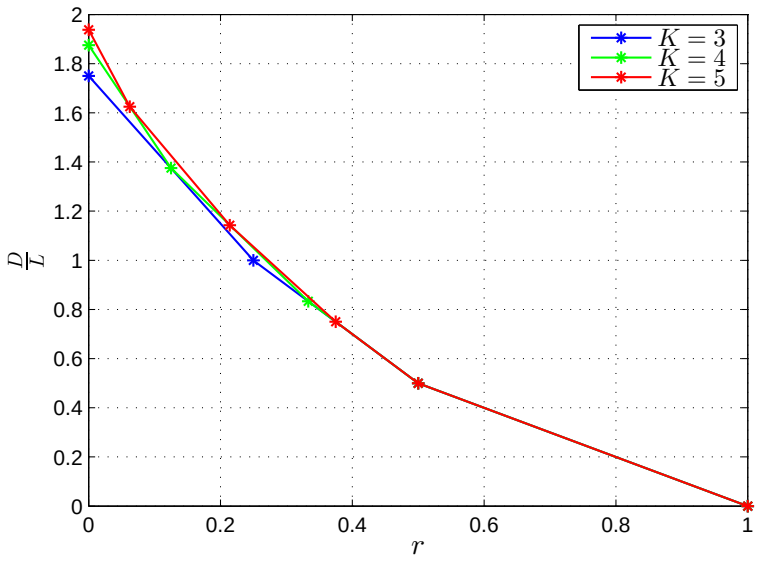

Figure 3.6: Outer bounds for $N = 2$, $K = 3$, $K = 4$ and $K = 5$.

in [32, 33], we use dotted lines to connect the corner points. For [28], Chapter 2 and this work, since we can apply memory-sharing to achieve the download costs between the corner points, we use solid lines to connect the corner points.

We first compare references [32, 33], in which the user caches M full messages out of K messages and the databases are unaware. In [33], the user not only wishes to protect the privacy of the desired messages but also wishes to protect the privacy of the cached messages. Note that the other works (Chapter 2, this chapter and [28, 32]) only consider to protect the privacy of the desired messages. Since the message privacy constraint is less restricted, reference [32] achieves lower normalized download cost than reference [33]. The main difference between Chapter 4 and [33] is that the databases are totally unaware of the cached M messages as in [33] or the nth database is aware of some of M messages cached from the nth database as in Chapter 4. Interestingly, these two models result in the same normalized

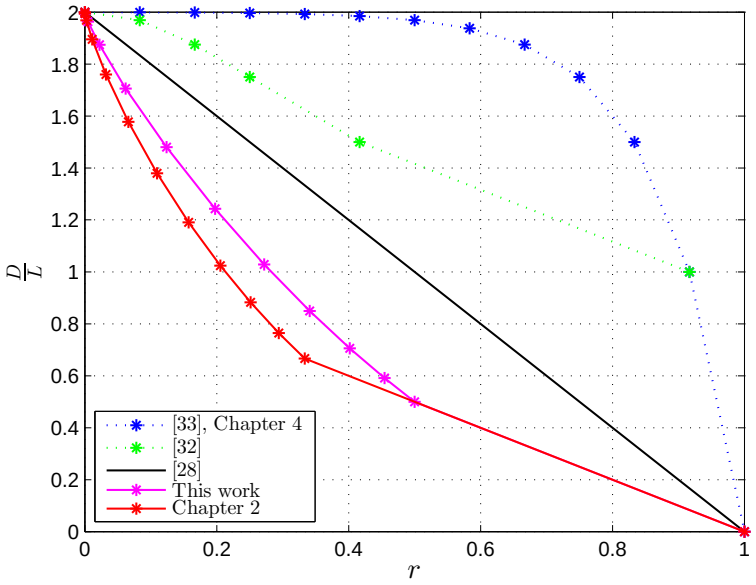

Figure 3.7: Outer bounds for $N = 2$, $K = 12$ for different cache-aided PIR models.

download costs. Although the nth database's awareness of some cached messages might increase the download cost, at the same time the user does not need to protect the privacy of these known messages from the nth database, which might reduce the download cost.

We then compare reference [28], Chapter 2 and this work. The main difference between these three works is the different level of awareness of the side information the user cached. Reference [28] considers that all the databases are aware of the side information the user cached. In contrast, Chapter 2 considers that all the databases are unaware of the side information. This work considers that the nth database is aware of the side information cached from the nth database. Corollary 2.1 in Chapter 2 shows the unawareness gain. Therefore, Chapter 2 achieves lower normalized download cost than [28]. The same proof technique in Corollary 2.1 in Chapter 2 can also show the partially unawareness gain. Therefore, this work also achieves

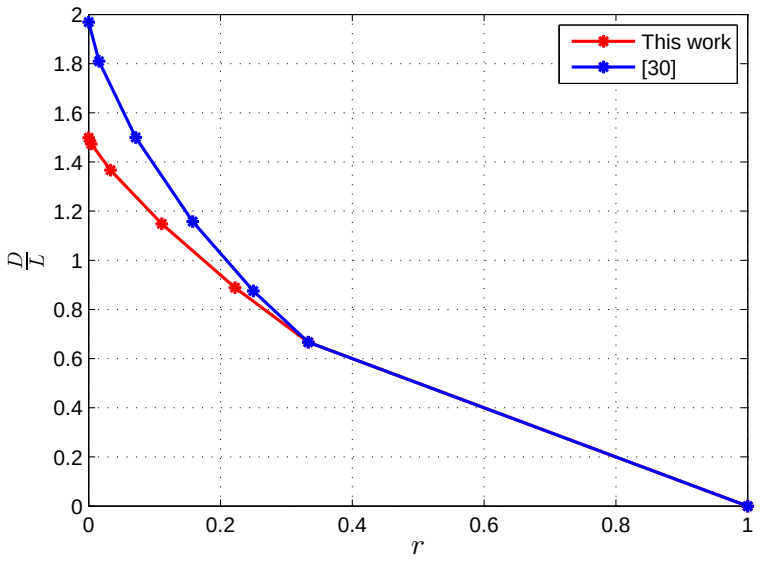

Figure 3.8: Comparison between this work and Chapter 2 for $N = 3$ and $K = 6$.

lower normalized download cost than [28]. Since these three works consider only the privacy of the desired message, different from [33] and Chapter 4, Chapter 2 achieves lower normalized download cost than this work. For high caching ratios $\frac{1}{N} \leq r \leq 1$, the proposed scheme in this work and that in Chapter 2 share the same normalized download cost $1 - r$.

We further compare Chapter 2 and this work in the following scenario. To apply the scheme in Chapter 2, for N databases, we choose one database for prefeching and use the remaining $N - 1$ databases for retrieval. Therefore, the cached side information is completely unknown to the $N-1$ databases. We also apply the scheme in this work for comparison. For a fixed caching ratio, we compare the normalized download costs. For caching ratios $\frac{1}{N} \leq r \leq 1$, the normalized download cost is $1 - r$ for both schemes. For caching ratios $\frac{K-2}{N^2+KN-4N+1} < r < \frac{1}{N}$, we can show analytically that the normalized download cost in this work is lower than that in

Chapter 2. For caching ratios $0 < r < \frac{1}{N}$, from numerical results, we observe that the scheme in this chapter achieves lower normalized download cost. For $N = 3$ and $K = 6$, numerical results are shown in Figure 3.8.

3.9 Conclusion

In this chapter, we studied the cache-aided PIR problem from N non-communicating and replicated databases, when the cache stores uncoded bits that are partially known to the databases. We determined inner and outer bounds for the optimal normalized download cost $D^*(r)$ as a function of the total number of messages K, the number of databases N, and the caching ratio r. Both inner and outer bounds are piece-wise linear functions in r (for fixed N, K) that consist of K line segments. The bounds match in two specific regimes: the very low caching ratio regime, i.e., $r \leq \frac{1}{N^{K-1}}$, and the very high caching ratio regime, where $r \geq \frac{K-2}{N^2-3N+KN}$. As a direct corollary for this result, we characterized the exact tradeoff between the download cost and the caching ratio for $K = 3$. For general K, N, and r, we showed that the largest additive gap between the achievability and the converse bounds is $\frac{5}{32}$. The achievable scheme extends the greedy scheme in [12] so that it starts with exploiting the cache bits as side information. For fixed K, N, there are $K - 1$ non-degenerate corner points. These points differ in the number of cached bits that contribute in generating one side information equation. The achievability for the remaining caching ratios is done by memory-sharing between the two adjacent corner points that enclose that caching ratio r. For the converse, we extended the induction-

based techniques in [12] and Chapter 2 to account for the availability of uncoded and partially prefetched side information at the retriever. The converse proof hinges on developing K lower bounds on the length of the undesired portion of the answer string. By applying induction on each bound separately, we obtained the piece-wise linear inner bound.

3.10 Appendix

Proof: From (3.12), we rewrite $\bar{D}(r_s)$ as

$$\bar{D}(r_s) = \frac{\sum_{i=0}^{K-1-s} \binom{K}{s+1+i}(N-1)^{i+1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^{i+1}} \tag{3.72}$$

$$= \frac{\frac{\sum_{i=0}^{K-1-s} \binom{K}{s+1+i}(N-1)^{i+1}}{\sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^{i+1}}}{\frac{\binom{K-2}{s-1}}{\sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^{i+1}} + 1} = \frac{\psi_1(N,K,s)}{\psi_2(N,K,s)+1}. \tag{3.73}$$

Let $\lambda = \frac{s}{K}$. We first upper bound $\psi_1(N,K,s)$,

$$\psi_1(N,K,s) = \frac{\sum_{i=0}^{K-1-s} \binom{K}{s+1+i}(N-1)^{i+1}}{\sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^{i+1}} \tag{3.74}$$

$$= \frac{\sum_{i=0}^{K-1-s} \frac{K}{s+1+i}\binom{K-1}{s+i}(N-1)^{i+1}}{\sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^{i+1}} \tag{3.75}$$

$$\leq \frac{\sum_{i=0}^{K-1-s} \frac{K}{s}\binom{K-1}{s+i}(N-1)^{i+1}}{\sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^{i+1}} = \frac{1}{\lambda}. \tag{3.76}$$

We then upper bound the reciprocal of $\psi_2(N,K,s)$ as,

$$\frac{1}{\psi_2(N,K,s)}$$

$$= \sum_{i=0}^{K-1-s} \frac{\binom{K-1}{s+i}(N-1)^{i+1}}{\binom{K-2}{s-1}} \tag{3.77}$$

$$= \sum_{i=0}^{K-1-s} \frac{(K-1)(K-1-s)\cdots(K-i-s)}{s(s+1)\cdots(s+i)}(N-1)^{i+1} \tag{3.78}$$

$$\leq (N-1)\sum_{i=0}^{K-1-s} \frac{K(K-s)^i}{s^{i+1}}(N-1)^i \tag{3.79}$$

$$= \frac{(N-1)}{\lambda}\sum_{i=0}^{(1-\lambda)K-1}\left(\frac{(1-\lambda)(N-1)}{\lambda}\right)^i. \tag{3.80}$$

When $\lambda > 1 - \frac{1}{N}$, $\frac{(1-\lambda)(N-1)}{\lambda} < 1$. As $K \to \infty$, $\frac{1}{\psi_2(N,K,s)}$ is upper bounded by

$$\lim_{K\to\infty}\frac{1}{\psi_2(N,K,s)} \leq \frac{N-1}{\lambda}\sum_{i=0}^{\infty}\left(\frac{(1-\lambda)(N-1)}{\lambda}\right)^i \tag{3.81}$$

$$= \frac{N-1}{N\lambda-(N-1)}. \tag{3.82}$$

Now, we lower bound (3.78) by keeping the first ϵK terms in the sum for any ϵ such that $0 < \epsilon < 1 - \lambda$,

$$\frac{1}{\psi_2(N,K,s)}$$

$$\geq \sum_{i=0}^{\epsilon K} \frac{(K-1)(K-1-s)\cdots(K-i-s)}{s(s+1)\cdots(s+i)}(N-1)^{i+1} \tag{3.83}$$

$$\geq (N-1)\sum_{i=0}^{\epsilon K} \frac{(K-1)(K-\epsilon K-s)^i}{(s+\epsilon K)^{i+1}}(N-1)^i \tag{3.84}$$

$$= (N-1)\sum_{i=0}^{\epsilon K} \frac{(1-\frac{1}{K})((1-(\lambda+\epsilon))^i}{(\lambda+\epsilon)^{i+1}}(N-1)^i. \tag{3.85}$$

As $K \to \infty$, for any $0 < \epsilon < 1 - \lambda$, we have

$$\lim_{K\to\infty} \frac{1}{\psi_2(N,K,s)} \geq \frac{N-1}{\lambda+\epsilon} \sum_{i=0}^{\infty} \left(\frac{(1-(\lambda+\epsilon))(N-1)}{\lambda+\epsilon} \right)^i \tag{3.86}$$

$$= \frac{N-1}{N(\lambda+\epsilon)-(N-1)}. \tag{3.87}$$

From (3.87) and (3.82), as $K \to \infty$, by picking $\epsilon \to 0$, we have

$$\psi_2(N,K,s) \to \frac{N}{N-1}\lambda - 1. \tag{3.88}$$

Furthermore, as $K \to \infty$, r_s converges to

$$r_s \to r = \lim_{K\to\infty} \frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^{i+1}} \tag{3.89}$$

$$= \lim_{K\to\infty} \frac{\psi_2(N,K,s)}{\psi_2(N,K,s)+1} \tag{3.90}$$

$$= \frac{N\lambda-(N-1)}{N\lambda} = 1 - \left(1 - \frac{1}{N}\right)\frac{1}{\lambda}. \tag{3.91}$$

Note that if $\lambda = 1 - \frac{1}{N}$, then $r = 0$, while if $\lambda = 1$, then $r = \frac{1}{N}$. Since we now consider the gap in the region of $0 \leq r \leq \frac{1}{N}$, without loss of generality, we consider $\lambda > 1 - \frac{1}{N}$. We express λ as

$$\lambda = \frac{1-\frac{1}{N}}{1-r}. \tag{3.92}$$

Continuing (3.73), by using (3.76), (3.88) and (3.92), we have the following

upper bound on $\bar{D}(r)$

$$\bar{D}(r) \leq \frac{\frac{1}{\lambda}}{\frac{N}{N-1}\lambda} = \frac{1}{\lambda^2}\left(1 - \frac{1}{N}\right) = \frac{N}{N-1}(1-r)^2. \tag{3.93}$$

Now, we compare the inner bound in (3.14) with the outer bound derived in (3.93). Note that the inner bound in (3.14) consists of K line segments, and these K line segments intersect at the following $K-1$ points given by,

$$\tilde{r}_i = \frac{1}{N^i}, \quad i = 1, \cdots, K-1. \tag{3.94}$$

As i increases, $\tilde{r}_i$ concentrates to $r = 0$. Therefore, for these K line segments, we only need to consider small number of them for the worst-gap analysis. Denote the gap between the inner and the outer bounds by $\Delta(N, K, r)$. We note that the gap $\Delta(N, \infty, r)$ is a piece-wise convex function for $0 \leq r \leq 1$ since it is the difference between a convex function $\bar{D}(r)$ and a piece-wise linear function. Hence, the maximizing caching ratio for the gap exists exactly at the corner points $\tilde{r}_i$ and it suffices to examine the gap at these corner points.

For the outer bound, by plugging (3.94) into (3.93), we have

$$\bar{D}(\tilde{r}_i) \leq \frac{N}{N-1}\left(1 - \frac{1}{N^i}\right)^2 = \frac{1 - (\frac{1}{N})^i}{1 - \frac{1}{N}}\left(1 - \frac{1}{N^i}\right). \tag{3.95}$$

Furthermore, for the inner bound, we have

$$\tilde{D}(\tilde{r}_i) = (1 - r_i)\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^i}\right)$$

$$- r_i \left(1 - \frac{1}{N}\right) \left(i + \frac{(i-1)}{N} + \cdots + \frac{1}{N^{i-1}}\right) \tag{3.96}$$

$$= - r_i \left[\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^i}\right) + \left(1 - \frac{1}{N}\right) \left(i + \frac{(i-1)}{N} + \cdots + \frac{1}{N^{i-1}}\right) \right] + \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^i}\right) \tag{3.97}$$

$$= - r_i(i+1) + \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^i}\right) \tag{3.98}$$

$$= \frac{1 - (\frac{1}{N})^{i+1}}{1 - \frac{1}{N}} - r_i(i+1) = \frac{1 - (\frac{1}{N})^{i+1}}{1 - \frac{1}{N}} - \frac{i+1}{N^i} \tag{3.99}$$

Consequently, we can upper bound the asymptotic gap at the corner point $\tilde{r}_i$ as

$$\Delta(N, \infty, \tilde{r}_i) = \bar{D}(\tilde{r}_i) - \tilde{D}(\tilde{r}_i) \leq \frac{1}{N^i} \left[i - \frac{1 - (\frac{1}{N})^i}{1 - \frac{1}{N}} \right] \tag{3.100}$$

Hence, $\Delta(N, \infty, \tilde{r}_i)$ is monotonically decreasing in N. Therefore,

$$\Delta(N, K, r) \leq \Delta(2, \infty, r) \leq \max_i \ \frac{1}{2^i} \left[i - \frac{1 - (\frac{1}{2})^i}{1 - \frac{1}{2}} \right] \tag{3.101}$$

For the case $N = 2$, we note that all the inner bounds after the 7th corner point are concentrated around $r = 0$ since $\tilde{r}_i \leq \frac{1}{128}$ for $i \geq 7$. Therefore, it suffices to characterize the gap only for the first 7 corner points. Considering the 7th corner point which corresponds to $\tilde{r}_6 = \frac{1}{128}$, and $\bar{D}(r) \leq 2$ trivially for all r, and $\tilde{D}(\frac{1}{128}) = 1.9297$. Hence, $\Delta(2, \infty, r) \leq 0.07$, for $r \leq \frac{1}{127}$. Now, we focus on calculating the gap at $\tilde{r}_i$, $i = 1, \cdots, 7$. Examining all the corner points, we see that $r = \frac{1}{8}$ is the maximizing caching ratio for the gap (corresponding to $i = 3$), and $\Delta(2, \infty, \frac{1}{8}) \leq \frac{5}{32}$,

which is the worst-case additive gap. ■

CHAPTER 4

The Capacity of Private Information Retrieval with Partially Known Private Side Information

4.1 Introduction

We consider the problem of private information retrieval (PIR) of a single message out of K messages from N replicated and non-colluding databases where a cache-enabled user (retriever) of cache-size M possesses side information in the form of full messages that are partially known to the databases. In this model, the user and the databases engage in a two-phase scheme, namely, the prefetching phase where the user acquires side information and the retrieval phase where the user downloads desired information. In the prefetching phase, the user receives m_n full messages from the nth database, under the cache memory size constraint $\sum_{n=1}^{N} m_n \leq M$. In the retrieval phase, the user wishes to retrieve a message such that no individual database learns anything about the identity of the desired message. In addition, the identities of the side information messages that the user did not prefetch from a database must remain private against that database. Since the side information provided by each database in the prefetching phase is known by the provid-

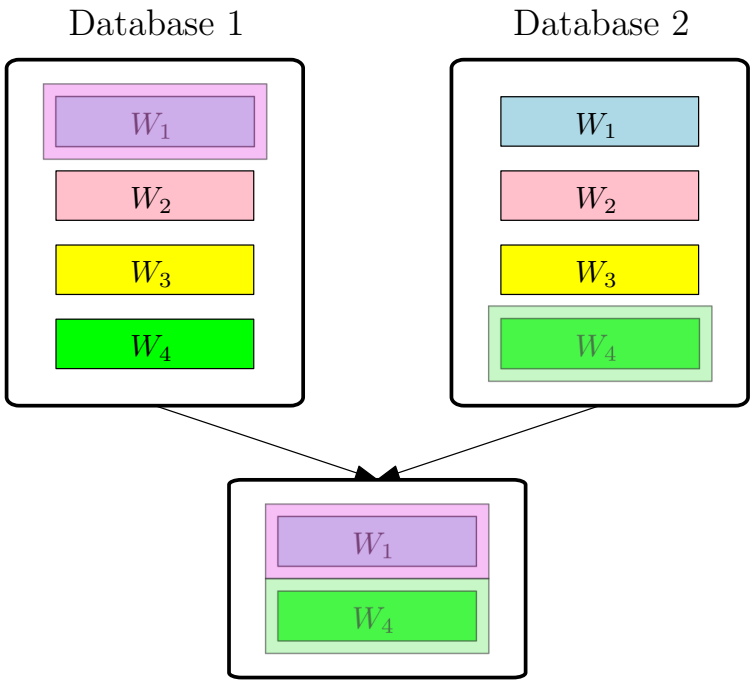

Figure 4.1: PIR with partially known PSI for $N = 2$, $K = 4$ and $M = 2$.

ing database and the side information must be kept private against the remaining databases, we coin this model as *partially known private side information*. We characterize the capacity of the PIR with partially known private side information to be $C = \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-M-1}}\right)^{-1} = \frac{1-\frac{1}{N}}{1-(\frac{1}{N})^{K-M}}$. Interestingly, this result is the same if none of the databases knows any of the prefetched side information, i.e., when the side information is obtained externally, a problem posed by Kadhe et al. [32] and settled by Chen-Wang-Jafar [33] recently. Thus, our result implies that there is no loss in using the same databases for both prefetching and retrieval phases.

4.2 System Model

We consider a classic PIR problem with K independent messages $W_1, \ldots, W_K$, where each message consists of L symbols,

$$H(W_1) = \cdots = H(W_K) = L, \qquad H(W_1, \ldots, W_K) = H(W_1) + \cdots + H(W_K). \tag{4.1}$$

There are N non-communicating databases, and each database stores all the K messages. The user (retriever) has a local cache memory which can store up to M messages.

There are two phases: a *prefetching phase* and a *retrieval phase*. In the prefetching phase, $\forall n \in [N]$, where $[N] = \{1, 2, \ldots, N\}$, the user caches m_n out of total K messages from the nth database. We denote the indices of the cached messages from the nth database as $\mathbb{H}_n$. Therefore, $|\mathbb{H}_n| = m_n$. We denote the indices of all cached messages as $\mathbb{H}$,

$$\mathbb{H} = \bigcup_{n=1}^{N} \mathbb{H}_n, \tag{4.2}$$

where $\mathbb{H}_{n_1} \cap \mathbb{H}_{n_2} = \emptyset$, if $n_1 \neq n_2$. Due to the cache memory size constraint, we require

$$|\mathbb{H}| = \sum_{n=1}^{N} m_n \leq M. \tag{4.3}$$

Since the user caches m_n messages from the nth database, $\mathbb{H}_n$ is known to the nth database. Since the databases do not communicate with each other, $\mathbb{H}_n$ is unknown to the other databases. We use $\mathbf{m} = (m_1, \ldots, m_N)$ to represent the prefetching phase. After the prefetching phase, the user learns $|\mathbb{H}|$ messages, denoted as $\mathcal{W}_{\mathbb{H}} = \{W_{i_1}, \ldots, W_{i_{|\mathbb{H}|}}\}$. We refer to $\mathcal{W}_{\mathbb{H}}$ as *partially known private side information*; see Fig. 4.1.

In the retrieval phase, the user privately generates a desired message index $\theta \in [K] \setminus \mathbb{H}$, and wishes to retrieve message W_θ such that no database knows which message is retrieved. Since the desired message index θ and cached message indices $\mathbb{H}$ are independent of the message contents, for random variables θ, $\mathbb{H}$, and $W_1, \ldots, W_K$, we have

$$H(\theta, \mathbb{H}, W_1, \ldots, W_K) = H(\theta, \mathbb{H}) + H(W_1) + \cdots + H(W_K). \tag{4.4}$$

In order to retrieve W_θ, the user sends N queries $Q_1^{[\theta,\mathbb{H}]}, \ldots, Q_N^{[\theta,\mathbb{H}]}$ to the N databases, where $Q_n^{[\theta,\mathbb{H}]}$ is the query sent to the nth database for message W_θ given the user has partially known private side information $\mathcal{W}_{\mathbb{H}}$. The queries are generated according to $\mathbb{H}$, which is independent of the realizations of the K messages. Therefore, we have

$$I(W_1, \ldots, W_K; Q_1^{[\theta,\mathbb{H}]}, \ldots, Q_N^{[\theta,\mathbb{H}]}) = 0. \tag{4.5}$$

To ensure that individual databases do not know which message is retrieved

and also do not know the cached messages from other databases, i.e., to guarantee the privacy of $(\theta, \mathbb{H} \setminus \mathbb{H}_n)$, we need to satisfy the following privacy constraint, $\forall n \in [N]$, $\forall \mathbb{H}, \mathbb{H}'$ such that $|\mathbb{H}| = |\mathbb{H}'| \leq M$, $\mathbb{H}_n \subset \mathbb{H}$, $\mathbb{H}_n \subset \mathbb{H}'$, and $\forall \theta \in [K] \setminus \mathbb{H}$, $\forall \theta' \in [K] \setminus \mathbb{H}'$,

$$(Q_n^{[\theta,\mathbb{H}]}, A_n^{[\theta,\mathbb{H}]}, W_1, \ldots, W_K, \mathbb{H}_n) \sim (Q_n^{[\theta',\mathbb{H}']}, A_n^{[\theta',\mathbb{H}']}, W_1, \ldots, W_K, \mathbb{H}_n), \tag{4.6}$$

where $A \sim B$ means that A and B are identically distributed.

Upon receiving the query $Q_n^{[\theta,\mathbb{H}]}$, the nth database replies with an answering string $A_n^{[\theta,\mathbb{H}]}$, which is a function of $Q_n^{[\theta,\mathbb{H}]}$ and all the K messages. Therefore, $\forall \theta \in [K] \setminus \mathbb{H}, \forall n \in [N]$,

$$H(A_n^{[\theta,\mathbb{H}]} | Q_n^{[\theta,\mathbb{H}]}, W_1, \ldots, W_K) = 0. \tag{4.7}$$

After receiving the answering strings $A_1^{[\theta,\mathbb{H}]}, \ldots, A_N^{[\theta,\mathbb{H}]}$ from all the N databases, the user needs to decode the desired message W_θ reliably. By using Fano's inequality, we have the following reliability constraint

$$H\left(W_\theta | \mathcal{W}_\mathbb{H}, \mathbb{H}, Q_1^{[\theta,\mathbb{H}]}, \ldots, Q_N^{[\theta,\mathbb{H}]}, A_1^{[\theta,\mathbb{H}]}, \ldots, A_N^{[\theta,\mathbb{H}]}\right) = o(L), \tag{4.8}$$

where $o(L)$ denotes a function such that $\frac{o(L)}{L} \to 0$ as $L \to \infty$.

For fixed N, K, and pretching scheme $\mathbf{m} = (m_1, \ldots, m_N)$, a pair $(D(\mathbf{m}), L(\mathbf{m}))$ is achievable if there exists a PIR scheme for messages of size $L(\mathbf{m})$ symbols long with partially known private side information satisfying the privacy constraint (4.6)

and the reliability constraint (4.8), where $D(\mathbf{m})$ represents the expected number of downloaded symbols (over all the queries) from the N databases via the answering strings $A_{1:N}^{[\theta,\mathbb{H}]}$, where $A_{1:N}^{[\theta,\mathbb{H}]} = (A_1^{[\theta,\mathbb{H}]}, \ldots, A_N^{[\theta,\mathbb{H}]})$, i.e.,

$$D(\mathbf{m}) = \sum_{n=1}^{N} H\left(A_n^{[\theta,\mathbb{H}]}\right). \tag{4.9}$$

In this work, for fixed N, K, and M, we aim to characterize the optimal normalized download cost D^*, where

$$D^* = \inf_{\mathbf{m}:(4.3)} \left\{ \frac{D(\mathbf{m})}{L(\mathbf{m})} : (D(\mathbf{m}), L(\mathbf{m})) \text{ is achievable} \right\}. \tag{4.10}$$

4.3 Main Results

We characterize the exact normalized download cost for the PIR problem with partially known private side information as shown in the following theorem.

Theorem 4.1 *In the PIR problem with partially known private side information under the cache memory size constraint $|\mathbb{H}| \leq M$, the optimal normalized download cost is*

$$D^* = 1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-M-1}} \tag{4.11}$$

$$= \frac{1 - (\frac{1}{N})^{K-M}}{1 - \frac{1}{N}}. \tag{4.12}$$

The converse proof for Theorem 4.1 is given in Section 4.4, and the achievability proof for Theorem 4.1 is given in Section 4.5. Theorem 4.1 does not assume any particular property for the prefetching strategy, i.e., $\mathbf{m}$ is arbitrary except for satisfying the memory size constraint. We have a few remarks.

Remark 4.1 *Theorem 4.1 implies that* $C = \frac{1}{D^*} = \frac{1-\frac{1}{N}}{1-(\frac{1}{N})^{K-M}}$. *Surprisingly, this capacity expression is exactly the same as the capacity for the PIR problem with completely unknown private side information in [33]. This implies that there is no loss in capacity due to employing the same databases for both prefetching and retrieval phases. The reason for this phenomenon is that although each database has a partial knowledge about some of the cached messages at the user, the privacy constraint on this known side information is relaxed.*

Remark 4.2 *The normalized download cost in Theorem 4.1 is the same as the normalized download cost for the classical PIR problem [12] if the number of messages is* $K - M$. *That is, a cache of size* M *messages effectively reduces the total number of messages by* M. *Noting that the download cost in [12] monotonically increases in the number of messages, the effective reduction in the number of messages by the cache size results in a significant reduction in the download cost due to the presence of side information at the user even though it is partially known by the databases and it needs to be kept private against other databases.*

Remark 4.3 *The optimal prefetching strategy exploits the entire cache memory of the user as the capacity expression is monotonically increasing in* M.

Remark 4.4 *In Section 4.5, we present the capacity achieving schemes for the partially known private side information. We note that, in general the PIR scheme in [33] is a valid achievable scheme for our problem as well. Nevertheless, in the special case of* uniform prefetching*, i.e.,* $m_n = \frac{M}{N} = m \in \mathbb{N}$*, we provide a different achievable scheme that exploits the prefetching uniformity to work with message size* $L = N^{K-m} = N^{K-\frac{M}{N}}$ *in contrast to* $L = N^K$ *needed for the scheme in [33], i.e., the message size is decreased by an exponential factor* $N^{\frac{M}{N}}$*. Furthermore, we note that although both schemes need an MDS code to reduce the number of downloaded equations, we note that the field size needed to realize this MDS code is significantly smaller with our scheme (if* $\frac{M}{N} \in \mathbb{N}$*) compared with the field size needed in the scheme in [33]. This implies that although* uniform prefetching *does not affect the PIR capacity, it significantly simplifies the achievable scheme.*

4.4 Converse Proof

In this section, we derive a general lower bound for the normalized download cost D^* given in (4.10). We extend the techniques presented in [12, 33] to the PIR problem with partially known private side information.

For the prefetching vector $\mathbf{m} = (m_1, \ldots, m_N)$ satisfying (4.3), we note that satisfying the memory size constraint with equality leads to a valid lower bound on (4.10). Consequently, we first consider the case $\sum_{n=1}^{N} m_n = \tilde{M} \leq M$, i.e., we study the case when the user learns $\tilde{M}$ messages after the prefetching phase. Since we do not specify the prefetching strategy $\mathbf{m}$ in advance, the following lower bound is

valid for all $\mathbf{m}$ such that $\sum_{n=1}^{N} m_n = \tilde{M}$. Without loss of generality, we relabel the $\tilde{M}$ cached messages as $W_1, W_2, \ldots, W_{\tilde{M}}$, i.e., $\mathbb{H} = \{1, 2, \ldots, \tilde{M}\}$ and $\mathcal{W}_{\mathbb{H}} = W_{1:\tilde{M}}$. We first need the following lemma, which characterizes a lower bound on the length of the undesired portion of the answering strings as a consequence of the privacy constraint on the retrieved message.

Lemma 4.1 (Interference lower bound) *For the PIR with partially known private side information, the interference from undesired messages within the answering strings, $D - L$, is lower bounded by,*

$$D - L + o(L) \geq I\left(W_{\tilde{M}+2:K}; \mathbb{H}, Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]}, A_{1:N}^{[\tilde{M}+1,\mathbb{H}]} | \mathcal{W}_{\mathbb{H}}, W_{\tilde{M}+1}\right). \tag{4.13}$$

If the privacy constraint is absent, the user downloads only L symbols for the desired message, however, when the privacy constraint is present, it should download D symbols. The difference between D and L, i.e., $D - L$, corresponds to the undesired portion of the answering strings. Note that Lemma 4.1 is an extension of [12, Lemma 5] if $\tilde{M} = 0$, i.e., the user has no partially known private side information. Lemma 4.1 differs from its counterpart in Chapter 2 [Lemma 1] in two aspects, namely, the left hand side is $D(r) - L(1 - r)$ in Chapter 2 as the user requests to download the uncached bits only, and the bound in Chapter 2 [Lemma 1] constructs $K - 1$ distinct lower bounds by changing k in contrast to one bound here as it always starts from $W_{\tilde{M}+2}$. Finally, we note that a similar argument to Lemma 4.1 can be implied from [33].

Proof: We start with the right hand side of (4.13),

$$I\left(W_{\tilde{M}+2:K};\mathbb{H},Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]},A_{1:N}^{[\tilde{M}+1,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+1}\right)$$
$$=I\left(W_{\tilde{M}+2:K};\mathbb{H},Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]},A_{1:N}^{[\tilde{M}+1,\mathbb{H}]},W_{\tilde{M}+1}|\mathcal{W}_{\mathbb{H}}\right)-I\left(W_{\tilde{M}+2:K};W_{\tilde{M}+1}|\mathcal{W}_{\mathbb{H}}\right). \tag{4.14}$$

For the first term on the right hand side of (4.14), we have

$$I\left(W_{\tilde{M}+2:K};\mathbb{H},Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]},A_{1:N}^{[\tilde{M}+1,\mathbb{H}]},W_{\tilde{M}+1}|\mathcal{W}_{\mathbb{H}}\right)$$
$$=I\left(W_{\tilde{M}+2:K};\mathbb{H},Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]},A_{1:N}^{[\tilde{M}+1,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}}\right)$$
$$+I\left(W_{\tilde{M}+2:K};W_{\tilde{M}+1}|\mathbb{H},Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]},A_{1:N}^{[\tilde{M}+1,\mathbb{H}]},\mathcal{W}_{\mathbb{H}}\right) \tag{4.15}$$
$$\overset{(4.8)}{=}I\left(W_{\tilde{M}+2:K};\mathbb{H},Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]},A_{1:N}^{[\tilde{M}+1,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}}\right)+o(L) \tag{4.16}$$
$$\overset{(4.4),(4.5)}{=}I\left(W_{\tilde{M}+2:K};A_{1:N}^{[\tilde{M}+1,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},\mathbb{H},Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]}\right)+o(L) \tag{4.17}$$
$$=H\left(A_{1:N}^{[\tilde{M}+1,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},\mathbb{H},Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]}\right)$$
$$-H\left(A_{1:N}^{[\tilde{M}+1,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},\mathbb{H},Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]},W_{\tilde{M}+2:K}\right)+o(L) \tag{4.18}$$
$$\overset{(4.8)}{=}H\left(A_{1:N}^{[\tilde{M}+1,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},\mathbb{H},Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]}\right)$$
$$-H\left(A_{1:N}^{[\tilde{M}+1,\mathbb{H}]},W_{\tilde{M}+1}|\mathcal{W}_{\mathbb{H}},\mathbb{H},Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]},W_{\tilde{M}+2:K}\right)+o(L) \tag{4.19}$$
$$\leq H\left(A_{1:N}^{[\tilde{M}+1,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},\mathbb{H},Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]}\right)$$
$$-H\left(W_{\tilde{M}+1}|\mathcal{W}_{\mathbb{H}},\mathbb{H},Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]},W_{\tilde{M}+2:K}\right)+o(L) \tag{4.20}$$
$$\overset{(4.4),(4.5)}{=}H\left(A_{1:N}^{[\tilde{M}+1,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},\mathbb{H},Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]}\right)-H\left(W_{\tilde{M}+1}|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+2:K}\right)+o(L) \tag{4.21}$$
$$=H\left(A_{1:N}^{[\tilde{M}+1,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},\mathbb{H},Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]}\right)-L+o(L) \tag{4.22}$$
$$\leq H\left(A_{1:N}^{[\tilde{M}+1,\mathbb{H}]}\right)-L+o(L) \tag{4.23}$$

$$\leq D - L + o(L), \tag{4.24}$$

where (4.16), (4.19) follow from the decodability of $W_{\tilde{M}+1}$ given $\left(\mathbb{H}, Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]}, A_{1:N}^{[\tilde{M}+1,\mathbb{H}]}, \mathcal{W}_{\mathbb{H}}\right)$, (4.17) follows from the independence of $W_{\tilde{M}+2:K}$ and $\left(\mathbb{H}, Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]}\right)$, (4.21) follows from the independence of $W_{\tilde{M}+1}$ and $\left(\mathbb{H}, Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]}\right)$, and (4.24) follows from the independence bound.

For the second term on the right hand side of (4.14), we have

$$I\left(W_{\tilde{M}+2:K}; W_{\tilde{M}+1}|\mathcal{W}_{\mathbb{H}}\right) = H\left(W_{\tilde{M}+1}|\mathcal{W}_{\mathbb{H}}\right) - H\left(W_{\tilde{M}+1}|\mathcal{W}_{\mathbb{H}}, W_{\tilde{M}+2:K}\right) \tag{4.25}$$

$$= L - L = 0. \tag{4.26}$$

Combining (4.14), (4.24), and (4.26) yields (4.13). ■

In the following lemma, we prove an inductive relation for the mutual information term on the right hand side of (4.13).

Lemma 4.2 (Induction lemma) *For all $k \in \{\tilde{M}+2, \ldots, K\}$, the mutual information term in Lemma 4.1 can be inductively lower bounded as,*

$$I\left(W_{k:K}; \mathbb{H}, Q_{1:N}^{[k-1,\mathbb{H}]}, A_{1:N}^{[k-1,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}}, W_{\tilde{M}+1:k-1}\right)$$
$$\geq \frac{1}{N} I\left(W_{k+1:K}; \mathbb{H}, Q_{1:N}^{[k,\mathbb{H}]}, A_{1:N}^{[k,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}}, W_{\tilde{M}+1:k}\right) + \frac{L - o(L)}{N}. \tag{4.27}$$

Lemma 4.2 is a generalization of [12, Lemma 6] to our setting. The main difference between Lemma 4.2 and [33] is that in order to apply the *partial* privacy constraint, the random variable $\mathbb{H}$ should be used in its local form $\mathbb{H}_n$ as it

corresponds to the partial knowledge of the nth database.

Proof: We start with the left hand side of (4.27),

$$
\begin{aligned}
&I\left(W_{k:K};\mathbb{H},Q_{1:N}^{[k-1,\mathbb{H}]},A_{1:N}^{[k-1,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+1:k-1}\right)\\
&\quad=\frac{1}{N}\times N\times I\left(W_{k:K};\mathbb{H},Q_{1:N}^{[k-1,\mathbb{H}]},A_{1:N}^{[k-1,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+1:k-1}\right) && (4.28)\\
&\quad\geq\frac{1}{N}\sum_{n=1}^{N}I\left(W_{k:K};\mathbb{H}_n,Q_n^{[k-1,\mathbb{H}]},A_n^{[k-1,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+1:k-1}\right) && (4.29)\\
&\quad\geq\frac{1}{N}\sum_{n=1}^{N}I\left(W_{k:K};Q_n^{[k-1,\mathbb{H}]},A_n^{[k-1,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+1:k-1},\mathbb{H}_n\right) && (4.30)\\
&\quad\overset{(4.6)}{=}\frac{1}{N}\sum_{n=1}^{N}I\left(W_{k:K};Q_n^{[k,\mathbb{H}]},A_n^{[k,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+1:k-1},\mathbb{H}_n\right) && (4.31)\\
&\quad\overset{(4.4),(4.5)}{=}\frac{1}{N}\sum_{n=1}^{N}I\left(W_{k:K};A_n^{[k,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+1:k-1},\mathbb{H}_n,Q_n^{[k,\mathbb{H}]}\right) && (4.32)\\
&\quad\overset{(4.7)}{=}\frac{1}{N}\sum_{n=1}^{N}H\left(A_n^{[k,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+1:k-1},\mathbb{H}_n,Q_n^{[k,\mathbb{H}]}\right) && (4.33)\\
&\quad\geq\frac{1}{N}\sum_{n=1}^{N}H\left(A_n^{[k,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+1:k-1},\mathbb{H},Q_{1:N}^{[k,\mathbb{H}]},A_{1:n-1}^{[k,\mathbb{H}]}\right) && (4.34)\\
&\quad\overset{(4.7)}{=}\frac{1}{N}\sum_{n=1}^{N}I\left(W_{k:K};A_n^{[k,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+1:k-1},\mathbb{H},Q_{1:N}^{[k,\mathbb{H}]},A_{1:n-1}^{[k,\mathbb{H}]}\right) && (4.35)\\
&\quad=\frac{1}{N}I\left(W_{k:K};A_{1:N}^{[k,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+1:k-1},\mathbb{H},Q_{1:N}^{[k,\mathbb{H}]}\right) && (4.36)\\
&\quad\overset{(4.4),(4.5)}{=}\frac{1}{N}I\left(W_{k:K};\mathbb{H},Q_{1:N}^{[k,\mathbb{H}]},A_{1:N}^{[k,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+1:k-1}\right) && (4.37)\\
&\quad\overset{(4.8)}{=}\frac{1}{N}I\left(W_{k:K};W_k,\mathbb{H},Q_{1:N}^{[k,\mathbb{H}]},A_{1:N}^{[k,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+1:k-1}\right)-\frac{o(L)}{N} && (4.38)\\
&\quad=\frac{1}{N}I\left(W_{k:K};W_k|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+1:k-1}\right)+\frac{1}{N}I\left(W_{k:K};\mathbb{H},Q_{1:N}^{[k,\mathbb{H}]},A_{1:N}^{[k,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+1:k}\right)\\
&\qquad-\frac{o(L)}{N} && (4.39)\\
&\quad=\frac{1}{N}I\left(W_{k+1:K};\mathbb{H},Q_{1:N}^{[k,\mathbb{H}]},A_{1:N}^{[k,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}},W_{\tilde{M}+1:k}\right)+\frac{L-o(L)}{N}, && (4.40)
\end{aligned}
$$

where (4.29) follows from the non-negativity of mutual information, (4.31) follows from the privacy constraint, (4.32) follows from the independence of the messages and the queries, (4.33), (4.35) follow from the fact that answer strings are deterministic functions of the messages and the queries, (4.34) follows from the fact that conditioning reduces entropy, (4.37) follows from the independence of $W_{k:K}$ and $\left(\mathbb{H}, Q_{1:N}^{[k,\mathbb{H}]}\right)$, (4.38) follows from the reliability constraint on W_k, and (4.40) follows from the independence of W_k and $(\mathcal{W}_{\mathbb{H}}, W_{\tilde{M}+1:k-1})$. ■

Now, we are ready to derive the lower bound for arbitrary K, N, and $\tilde{M}$. This can be obtained by applying Lemma 4.1 and Lemma 4.2 successively.

Lemma 4.3 *For fixed N, K, and $\tilde{M} \leq M$, we have*

$$D \geq L\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-\tilde{M}-1}}\right) - o(L). \tag{4.41}$$

Proof: We have

$$\begin{aligned}
D &\overset{(4.13)}{\geq} L + I\left(W_{\tilde{M}+2:K}; \mathbb{H}, Q_{1:N}^{[\tilde{M}+1,\mathbb{H}]}, A_{1:N}^{[\tilde{M}+1,\mathbb{H}]} | \mathcal{W}_{\mathbb{H}}, W_{\tilde{M}+1}\right) - o(L) && (4.42)\\
&\overset{(4.27)}{\geq} L + \frac{L}{N} + \frac{1}{N} I\left(W_{\tilde{M}+3:K}; \mathbb{H}, Q_{1:N}^{[\tilde{M}+2,\mathbb{H}]}, A_{1:N}^{[\tilde{M}+2,\mathbb{H}]} | \mathcal{W}_{\mathbb{H}}, W_{\tilde{M}+1:\tilde{M}+2}\right) - o(L) && (4.43)\\
&\overset{(4.27)}{\geq} L + \frac{L}{N} + \frac{L}{N^2} + \frac{1}{N} I\left(W_{\tilde{M}+4:K}; \mathbb{H}, Q_{1:N}^{[\tilde{M}+3,\mathbb{H}]}, A_{1:N}^{[\tilde{M}+3,\mathbb{H}]} | \mathcal{W}_{\mathbb{H}}, W_{\tilde{M}+1:\tilde{M}+3}\right) \\
&\qquad - o(L) && (4.44)\\
&\overset{(4.27)}{\geq} \cdots && (4.45)\\
&\overset{(4.27)}{\geq} L\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-\tilde{M}-1}}\right) - o(L), && (4.46)
\end{aligned}$$

where (4.42) follows from Lemma 4.1, (4.43)-(4.46) follow from applying Lemma 4.2 starting from $k = \tilde{M} + 2$ to $k = K$, which differs from [12] in terms of the starting point of the induction. ■

We conclude the converse proof by dividing by L and taking $L \to \infty$ in (4.41), to have

$$D^* \geq 1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-\tilde{M}-1}}. \tag{4.47}$$

Finally, we note that the right hand side of (4.47) is monotonically decreasing in $\tilde{M}$. Since $\tilde{M} \leq M$, the lowest lower bound is obtained by taking $\tilde{M} = M$, which yields the final converse bound,

$$D^* \geq 1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-M-1}}. \tag{4.48}$$

Remark 4.5 *We note that if* (4.48) *is tight, any prefetching strategy* $\mathbf{m}$ *such that* $\sum_{n=1}^{N} m_n < M$ *is strictly suboptimal. Furthermore, the lower bound in* (4.48) *is the same for all prefetching strategies* $\mathbf{m}$ *satisfying* $\sum_{n=1}^{N} m_n = M$*. In Section 4.5, we show that this lower bound is tight.*

4.5 Achievability Proof

We first note that the achievability scheme proposed in [33] for the PIR problem with completely unknown private side information also works for the PIR problem with partially known private side information here. The PIR scheme in [33] is based

on MDS codes and consists of two stages. The first stage determines the systematic part of the MDS code according to the queries generated in [12], which protects the privacy of the desired message, i.e., in the first stage, the user designs the queries such that no information is leaked about which message out of the K messages is the desired one. In the second stage, the user reduces the number of the downloaded equations by downloading the parity part of the MDS code only. For the case of partially known private side information here, two privacy constraints should be satisfied: the desired message privacy constraint and the side information privacy constraint. For the desired message, we note that the user should guarantee that the queries designed to retrieve any of the $K - m_n$ messages should be indistinguishable at the nth database (i.e., with the exception of the m_n messages that the nth database has provided). Due to the first stage, the privacy of the desired message holds as it was designed to protect the privacy of all K messages, which is more restricted. Furthermore, the PIR scheme in [33] also protects the privacy of the side information. The scheme in [33] ensures that the queries do not reveal the identity of the M messages that are possessed by the user as side information. In our model, we note that we need to protect the privacy of $M - m_n$ messages from the nth database, as the remaining m_n messages are known to the nth database. Since the privacy constraint imposed on the side information in our model is less restricted than [33], using the scheme in [33] satisfies the privacy constraint of the side information in our case as well. That is, the nth database cannot infer which other $M - m_n$ messages the user holds. The PIR scheme in [33] achieves the normalized download cost in Theorem 4.1. The PIR scheme in [33] requires a message size of N^K symbols. In the

following, we propose another achievability scheme which requires a message size of $N^{K-\frac{M}{N}}$, if $m_n = \frac{M}{N} \in \mathbb{N}$. Thus, this scheme requires smaller sub-packetization and smaller field size for the MDS code.

Our PIR scheme for partially known private side information is based on the PIR schemes in [12, 33]. To protect the privacy of the partially known private side information and the privacy of the desired message, similar to [12], we apply the following three principles recursively: 1) database symmetry, 2) message symmetry within each database, and 3) exploiting undesired messages as side information. We reduce the download cost by utilizing the reconstruction property of MDS codes by exploiting partially known private side information as in [33]. The side information enables the user to request reduced number of equations as a consequence of the user's knowledge of M messages from the prefetching phase. Nevertheless, to protect the privacy of the side information, the user actually queries MDS coded symbols which is mixture of $K-m_n$ messages. The main difference between our achievability scheme and that in [12, 33] is that since the nth database knows that the user has prefetched m_n messages, the user does not need to protect the privacy for these m_n messages from the nth database. This effectively reduces the number of messages that the scheme in [33] needs to operate on to $K - m_n$ messages in contrast to K in [33]. When $\frac{M}{N} \in \mathbb{N}$, we show that if the user caches the same number of messages from each database, i.e., $m_n = \frac{M}{N}$, for all n, then the lower bound in (4.11) is achievable by this scheme. This scheme reduces the message size requirement from $L = N^K$ in [33] to $L = N^{K-\frac{M}{N}}$ here, simplifying the achievable scheme.

4.5.1 Motivating Examples

4.5.1.1 $N = 2$ Databases, $K = 4$ Messages, and $M = 2$ Cached Messages

Assume that each message is of size 8 symbols. We use a_i, b_i, c_i and d_i, for $i = 1, \ldots, 8$, to denote the symbols of messages W_1, W_2, W_3 and W_4, respectively. In this example, in the prefetching phase, the user caches message W_3 from database 1, and message W_4 from database 2; and in the retrieval phase, the user wishes to retrieve message W_1 privately. The user first generates the query table in Table 4.1. In Table 4.1, the user queries 7 symbols. Since the user knows d_1 from the cached message W_4, in order to use the partially known private side information, the user can in fact reduce the number of queries to 6 equations per database by ignoring d_1. However, if the user simply does not download d_1, it compromises the privacy of W_4 at database 1. Alternatively, the user queries the MDS coded version of the 7 symbols. By using these 7 symbols as the systematic part, we can use a $(13, 7)$ MDS code. By downloading the 6 parity symbols, the user can reconstruct the whole 7 symbols utilizing the knowledge of d_1. Therefore, the normalized download cost for our achievability scheme is $\frac{6+6}{8} = \frac{3}{2}$, which matches the lower bound in (4.11) for this case.

For database 1, the query table in Table 4.1 induces the same distribution on the messages W_1, W_2 and W_4. Therefore, we guarantee the privacy of the desired message. The reliability constraint can also be verified. Note that b_2 is downloaded

from database 2, and d_2 is downloaded in the prefetching phase. Therefore, a_3 and a_4 are decodable. By getting $b_4 + c_3$ from database 2, the user can get b_4 due to the private side information W_3. Therefore, the user can decode a_7 from $a_7 + b_4 + d_4$. Similar arguments follow for database 2.

Table 4.1: Query table for $K = 4$, $N = 2$, $M = 2$.

DB1	DB2
a_1	a_2
b_1	b_2
d_1	c_1
$a_3 + b_2$	$a_5 + b_1$
$a_4 + d_2$	$a_6 + c_2$
$b_3 + d_3$	$b_4 + c_3$
$a_7 + b_4 + d_4$	$a_8 + b_3 + c_4$

$\mathcal{W}_{\mathbb{H}_1}=\{W_3\}$	$\mathcal{W}_{\mathbb{H}_2}=\{W_4\}$

4.5.1.2 $N = 2$ Databases, $K = 5$ Messages, and $M = 2$ Cached Messages

Assume that each message is of size 16 symbols. We use a_i, b_i, c_i, d_i and e_i, for $i = 1, \ldots, 16$, to denote the symbols of messages W_1, W_2, W_3, W_4, and W_5, respectively. In this example, in the prefetching phase, the user caches message W_4 from database 1, and message W_5 from database 2; and in the retrieval phase, the user wishes to retrieve message W_1 privately. The user first generates the query table in Table 4.2. In Table 4.2, the user queries 15 symbols. Since the user knows e_1 from the cached message W_5, in order to use the partially known private side information, the user in fact queries the MDS coded version of the 15 symbols. By using these 15 symbols as the systematic part, we can use a $(29, 15)$ MDS code. By

downloading the 14 parity symbols, the user can reconstruct the whole 15 symbols. Therefore, the normalized download cost for our achievability scheme is $\frac{14+14}{16} = \frac{7}{4}$, which matches the lower bound in (4.11) for this case.

For database 1, the query table in Table 4.2 induces the same distribution on the messages W_1, W_2, W_3 and W_5. Therefore, we guarantee the privacy of the desired message. The reliability constraint can also be verified. Note that b_2, c_2 are downloaded from database 2, and e_2 is downloaded in the prefetching phase. Therefore, a_3, a_4 and a_5 are decodable. By getting $b_6 + d_3$ from database 2, the user can get b_6 due to the private side information W_4. Similarly, c_6 is also decodable. Therefore, the user can decode a_{10} from $a_{10} + b_6 + e_5$ and a_{11} from $a_{11} + c_6 + e_6$. By getting $b_8 + c_8 + d_7$ from database 2, the user can get $b_8 + c_8$ due to the private side information W_4. Therefore, the user can decode a_{15} from $a_{15} + b_8 + c_8 + e_8$. Similar arguments follow for database 2.

4.5.2 General Achievable Scheme for $\frac{M}{N} \in \mathbb{N}$

Let $\frac{M}{N} = m$. In the prefetching phase, the user caches m messages from each database. To achieve the lower bound shown in (4.11), in the retrieval phase, we choose the message size as $L = N^{K-m}$ symbols. The details of the achievable scheme are as follows:

1. *Initialization:* The user permutes each message randomly and independently. After the random permutation, we use $U_i(j)$ to denote the jth symbol of the permuted message W_i. Suppose the user wishes to retrieve W_θ privately. We

Table 4.2: Query table for $K = 5$, $N = 2$, $M = 2$.

DB1	DB2
a_1	a_2
b_1	b_2
c_1	c_2
e_1	d_1
$a_3 + b_2$	$a_6 + b_1$
$a_4 + c_2$	$a_7 + c_1$
$a_5 + e_2$	$a_8 + d_2$
$b_3 + c_3$	$b_5 + c_5$
$b_4 + e_3$	$b_6 + d_3$
$c_4 + e_4$	$c_6 + d_4$
$a_9 + b_5 + c_5$	$a_{12} + b_3 + c_3$
$a_{10} + b_6 + e_5$	$a_{13} + b_4 + d_5$
$a_{11} + c_6 + e_6$	$a_{14} + c_4 + d_6$
$b_7 + c_7 + e_7$	$b_8 + c_8 + d_7$
$a_{15} + b_8 + c_8 + e_8$	$a_{16} + b_7 + c_7 + d_8$
$\mathcal{W}_{\mathbb{H}_1}=\{W_4\}$	$\mathcal{W}_{\mathbb{H}_2}=\{W_5\}$

then prepare the query table by first querying $U_\theta(1)$ from database 1. Set the round index to $r = 1$.

2. *Symmetry across databases:* The user queries the same number of equations with the same structure as database 1 from the remaining databases.

3. *Message symmetry:* For each database, to satisfy the privacy constraint, the user should query equal amount of symbols from all other $K - m$ messages. Since the user has cached m messages from each database in the prefetching phase, the user does not need to protect the privacy for these m messages. For the rth round, the user queries sums of every r combinations of the $K - m$ messages.

4. *Exploiting side information:* For database 1, the user exploits the side infor-

mation equations obtained from the other $(N-1)$ databases to query sum of $r+1$ combinations of the $K-m$ messages, where sum of r combinations is the side information. If the r combinations contain the cached message from database 1, we replace the overlapping symbols through the symbols cached from other databases.

5. *Repeat* steps 2, 3, 4 after setting $r = r+1$ until $r = K-m+1$.

6. *Shuffling the order of queries:* By shuffling the order of queries uniformly, all possible queries can be made equally likely regardless of the message index. This guarantees the privacy of the desired message.

7. *Downloading MDS parity parts:* Now, the query table is finished. For each database, let p be the number of queried symbols in the query table, and let q be the number of queried symbols which are determined by the side information the user cached in the prefetching phase. Apply a $(2p-q, p)$ MDS code to the queried symbols by letting the p symbols to be the systematic part. Finally, the user downloads the parity parts of the MDS-coded answering strings which are $p-q$ symbols for each database.

4.5.3 Normalized Download Cost

We now calculate the total number of downloaded symbols. We first calculate p, which is the number of queried symbols in the query table for each database,

$$p = \binom{K-m}{1} + \binom{K-m}{2}(N-1) + \cdots + \binom{K-m}{K-m}(N-1)^{K-m-1} \qquad (4.49)$$

$$= \frac{1}{N-1}\left[\binom{K-m}{1}(N-1) + \binom{K-m}{2}(N-1)^2 + \dots + \binom{K-m}{K-m}(N-1)^{K-m}\right] \tag{4.50}$$

$$= \frac{1}{N-1}\left(N^{K-m}-1\right), \tag{4.51}$$

where $\binom{K-m}{r}$ in (4.49) corresponds to the queries of sums of every r combinations of the $K-m$ messages, and $(N-1)^{r-1}$ corresponds to the number of sets of the available side information from other $(N-1)$ databases.

We then calculate q, which is the number of queried symbols which are determined by the side information the user cached in the prefetching phase,

$$q = \binom{(N-1)m}{1} + \binom{(N-1)m}{2}(N-1) + \cdots + \binom{(N-1)m}{(N-1)m}(N-1)^{(N-1)m-1} \tag{4.52}$$

$$= \frac{1}{N-1}\left[\binom{(N-1)m}{1}(N-1) + \cdots + \binom{(N-1)m}{(N-1)m}(N-1)^{(N-1)m}\right] \tag{4.53}$$

$$= \frac{1}{N-1}\left(N^{(N-1)m}-1\right), \tag{4.54}$$

where $\binom{(N-1)m}{r}$ in (4.52) corresponds to the queries which can be determined by the partially known private side information, and $(N-1)^{r-1}$ corresponds to the number of sets of queries consisting of r combinations.

Next, we calculate the number of symbols for the desired message,

$$L = N\left[\binom{K-m-1}{0} + \binom{K-m-1}{1}(N-1) + \dots\right.$$

$$+\binom{K-m-1}{K-m-1}(N-1)^{K-m-1}\Bigg] \tag{4.55}$$

$$= N \times N^{K-m-1} = N^{K-m}, \tag{4.56}$$

where $\binom{K-m-1}{r-1}$ in (4.55) corresponds to the queries containing the desired message and $(N-1)^{r-1}$ corresponds to the number of sets of queries consisting of r combinations.

Therefore, the normalized download cost becomes,

$$\frac{D}{L} = \frac{N(p-q)}{L} \tag{4.57}$$

$$= \frac{\frac{N}{N-1}\left(N^{K-m}-1\right) - \frac{N}{N-1}\left(N^{(N-1)m}-1\right)}{N^{K-m}} \tag{4.58}$$

$$= \frac{N}{N-1} \times \frac{N^{K-m} - N^{(N-1)m}}{N^{K-m}} \tag{4.59}$$

$$= \frac{1}{1-\frac{1}{N}} \times \left[1 - \left(\frac{1}{N}\right)^{K-M}\right], \tag{4.60}$$

which matches the lower bound in (4.11).

Remark 4.6 *Note that although our achievable scheme and the scheme in [33] are both using MDS coding to exploit the available side information, the field size requirements for realizing the MDS codes are different. For the scheme of [33], a* $(2\tilde{p}-\tilde{q}, \tilde{p})$ *MDS code is used, where* $\tilde{p} = \frac{1}{N-1}(N^K - 1)$ *and* $\tilde{q} = \frac{1}{N-1}(N^M - 1)$. *This requires larger field size than the* $(2p-q, p)$ *MDS code used in our scheme (if* $\frac{M}{N} \in \mathbb{N}$*), since* $2\tilde{p} - \tilde{q} > (2p-q)$.

4.6 Conclusion

In this chapter, we have introduced a new PIR model, namely, PIR with partially known private side information as a natural model for studying practical PIR problems with cached side information. In this model, the user and the databases engage in a caching/PIR scenario which consists of two phases, namely, prefetching phase and retrieval phase. The nth database provides the user with m_n side information messages in the prefetching phase such that $\sum_{n=1}^{N} m_n \leq M$, hence, each database has *partial knowledge* about the side information in contrast to full knowledge in [28] and no knowledge in [32,33] and Chapter 2. Based on this side information, the user designs a retrieval scheme that does not reveal the identity of the desired message or the identities of the remaining $M - m_n$ messages to the nth database. For this model, we determined the exact capacity to be $C = \frac{1-\frac{1}{N}}{1-(\frac{1}{N})^{K-M}}$. The capacity is attained for any prefetching strategy that satisfies the cache memory size constraint with equality. The achievable scheme in [33] can also be used for this model. We further proposed another PIR scheme which requires smaller sub-packetization and field size for the case of uniform prefetching. Uniform prefetching, when feasible, is optimal. Interestingly, the capacity expression we derive for this problem is exactly the same as the capacity expression for the PIR problem with completely unknown side information [33]. Therefore, our result implies that there is no loss in employing the same databases for prefetching and retrieval purposes.

CHAPTER 5

The Capacity of Private Information Retrieval with Private Side Information Under Storage Constraints

5.1 Introduction

We consider the problem of private information retrieval (PIR) of a single message out of K messages from N replicated and non-colluding databases where a cache-enabled user (retriever) of cache-size S possesses side information in the form of uncoded portions of the messages that are unknown to the databases. The identities of these side information messages need to be kept private from the databases, i.e., we consider PIR with private side information (PSI). We characterize the optimal normalized download cost for this PIR-PSI problem under the storage constraint S as $D^* = 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1-M}} + \frac{1-r_M}{N^{K-M}} + \frac{1-r_{M-1}}{N^{K-M+1}} + \cdots + \frac{1-r_1}{N^{K-1}}$, where r_i is the portion of the ith side information message that is cached with $\sum_{i=1}^{M} r_i = S$. Based on this capacity result, we prove two facts: First, for a fixed memory size S and a fixed number of accessible messages M, uniform caching achieves the lowest normalized download cost, i.e., $r_i = \frac{S}{M}$, for $i = 1, \ldots, M$, is optimum. Second, for a fixed memory size S, among all possible $K - \lceil S \rceil + 1$ uniform caching schemes,

the uniform caching scheme which caches $M = K$ messages achieves the lowest normalized download cost.

5.2 System Model

We consider a system consisting of N non-communicating databases and a user (retriever). Each database stores the same set of K independent messages $W_1, \ldots, W_K$, and each message is of size L symbols, i.e.,

$$H(W_1) = \cdots = H(W_K) = L, \qquad H(W_1, \ldots, W_K) = H(W_1) + \cdots + H(W_K). \tag{5.1}$$

The user has a local cache memory which is of size SL symbols, where $S \in [0, K]$. There are two phases in the system: the prefetching phase and the retrieval phase.

In the prefetching phase, the user can randomly access M messages out of total K messages, where $M \geq S$. For each of the M accessed messages, the user caches the first Lr_i symbols out of the total L symbols for $i = 1, \ldots, M$. The caching scheme is subject to a memory size constraint of S, i.e.,

$$\sum_{i=1}^{M} r_i = S. \tag{5.2}$$

We denote the indices (identities) of the cached M messages as $\mathbb{H}$, and denote $\mathcal{W}_{\mathbb{H}}$ as the cached messages. Therefore, $|\mathbb{H}| = M$, and $H(\mathcal{W}_{\mathbb{H}}) = SL$.

Note that M and $(r_1, \ldots, r_M)$ specify a caching scheme. If $r_1 = \cdots = r_M$, we

call this a uniform caching scheme. For fixed S, there are $K-\lceil S\rceil+1$ uniform caching schemes depending on the number of accessible messages since $M \geq S$. For instance, if there are $K = 3$ messages in the databases and $S = 1.5$, then since $M \geq S$, M can take one of two possible values: either 2 or 3. Thus, there are two uniform caching schemes depending on the value of M. Note, $K - \lceil S\rceil + 1 = 3 - \lceil 1.5\rceil + 1 = 2$.

We assume that all the databases are aware of the caching scheme but are unaware of which messages are cached. For example, if $S = 2$, $M = 3$, and we say that the user has applied a uniform caching scheme, the databases know that the user has chosen 3 messages out of the total K messages to cache, and for each chosen message, the user has cached the first $\frac{2}{3}L$ symbols out of the total L symbols. However, the databases do not know which messages are cached by the user.

In the retrieval phase, the user privately generates an index $\theta \in [K] = \{1, \ldots, K\}$, and wishes to retrieve message W_θ such that it is impossible for any individual database to identify θ. At the same time, the user also wishes to keep the indices of the M cached messages private, i.e., in the retrieval phase the databases cannot learn which messages are cached. For random variables θ, $\mathbb{H}$, and $W_1, \ldots, W_K$, we have

$$H\left(\theta, \mathbb{H}, W_1, \ldots, W_K\right) = H\left(\theta\right) + H\left(\mathbb{H}\right) + H(W_1) + \cdots + H(W_K). \tag{5.3}$$

In order to retrieve message W_θ, the user sends N queries $Q_1^{[\theta,\mathbb{H}]}, \ldots, Q_N^{[\theta,\mathbb{H}]}$ to the N databases, where $Q_n^{[\theta,\mathbb{H}]}$ is the query sent to the nth database for message W_θ. Note that the queries are generated according to $\mathbb{H}$, which are independent of the

realization of the K messages. Therefore,

$$I(W_1, \ldots, W_K; Q_1^{[\theta,\mathbb{H}]}, \ldots, Q_N^{[\theta,\mathbb{H}]}) = 0. \tag{5.4}$$

Upon receiving the query $Q_n^{[\theta,\mathbb{H}]}$, the nth database replies with an answering string $A_n^{[\theta,\mathbb{H}]}$, which is a function of $Q_n^{[\theta,\mathbb{H}]}$ and all the K messages. Therefore, $\forall \theta \in [K], \forall n \in [N]$,

$$H(A_n^{[\theta,\mathbb{H}]} | Q_n^{[\theta,\mathbb{H}]}, W_1, \ldots, W_K) = 0. \tag{5.5}$$

After receiving the answering strings $A_1^{[\theta,\mathbb{H}]}, \ldots, A_N^{[\theta,\mathbb{H}]}$ from all the N databases, the user needs to decode the desired message W_θ reliably. By using Fano's inequality, we have the following reliability constraint

$$H\left(W_\theta | \mathcal{W}_{\mathbb{H}}, \mathbb{H}, Q_1^{[\theta,\mathbb{H}]}, \ldots, Q_N^{[\theta,\mathbb{H}]}, A_1^{[\theta,\mathbb{H}]}, \ldots, A_N^{[\theta,\mathbb{H}]}\right) = o(L), \tag{5.6}$$

where $o(L)$ denotes a function such that $\frac{o(L)}{L} \to 0$ as $L \to \infty$.

To ensure that individual databases do not know which message is retrieved and to keep the M cached messages private, we have the following privacy constraint, $\forall n \in [N]$, $\forall \theta, \theta' \in [K]$, $\forall \mathbb{H}, \mathbb{H}' \subset [K]$ such that $|\mathbb{H}| = |\mathbb{H}'| = M$,

$$(Q_n^{[\theta,\mathbb{H}]}, A_n^{[\theta,\mathbb{H}]}, W_1, \ldots, W_K) \sim (Q_n^{[\theta',\mathbb{H}']}, A_n^{[\theta',\mathbb{H}']}, W_1, \ldots, W_K), \tag{5.7}$$

where $A \sim B$ means that A and B are identically distributed.

For a fixed N, K, S and caching scheme $(r_1, \ldots, r_M)$, a pair (D, L) is achievable if there exists a PIR scheme for the message which is of size L symbols satisfying the reliability constraint (5.6) and the privacy constraint (5.7), where D represents the expected number of downloaded bits (over all the queries) from the N databases via the answering strings $A_{1:N}^{[\theta,\mathbb{H}]}$, where $A_{1:N}^{[\theta,\mathbb{H}]} = (A_1^{[\theta,\mathbb{H}]}, \ldots, A_N^{[\theta,\mathbb{H}]})$, i.e.,

$$D = \sum_{n=1}^{N} H\left(A_n^{[\theta,\mathbb{H}]}\right). \tag{5.8}$$

In this work, we aim at characterizing the optimal normalized download cost D^*, where

$$D^* = \inf\left\{\frac{D}{L} : (D, L) \text{ is achievable}\right\}. \tag{5.9}$$

We use an example shown in Fig. 5.1 to illustrate the system model. Consider a user wanting to download a message from $N = 3$ non-communicating databases, each storing the same set of $K = 5$ messages. Assume that the user is already in possession of $M = 3$ messages through some unspecified means; the user may have obtained these from another user, or it may have prefetched them from another database. The databases do not know the identities of these messages, but they know that the user has access to $M = 3$ messages. (For this example, say these messages are W_2, W_4 and W_5.) However, the user has limited local storage with size $S = 1$ message. What should the user keep in order to minimize the download cost of the desired message during the PIR phase while keeping the identities of

both desired and cached messages private? Should the user keep 1 full message in its cache and discard the other 2 messages, shown as caching option 1 in Fig. 5.1? Should the user choose 2 messages, store half of each chosen message and discard the remaining 1 message, shown as caching option 2 in Fig. 5.1? Or, should the user keep all 3 messages and store a portion of each? In that case, what portions of messages should the user store? E.g., should it store 25% of W_2, 25% of W_4 and 50% of W_5, shown as caching option 3, or should it store $\frac{1}{3}$ of all 3 messages, shown as caching option 4 in Fig. 5.1?

Different caching schemes result in different download costs for the PIR-PSI problem. Intuition may say that if portions of many messages are kept in the cache, then the user will need to protect many identities from the databases due to the PSI requirement, which may seem disadvantageous. On the other hand, intuition may also say that keeping portions of many messages may improve the diversity of side information for the PIR phase, which may seem advantageous. What is the optimum way to utilize the user's limited cache memory? In this chapter, we characterize the optimal normalized download cost for any given caching strategy, and determine the optimal caching strategy under a given storage constraint.

5.3 Main Results and Discussions

We characterize the exact normalized download cost for PIR-PSI under a storage constraint in the following theorem.

Theorem 5.1 *In PIR-PSI under a storage constraint, the optimal normalized down-*

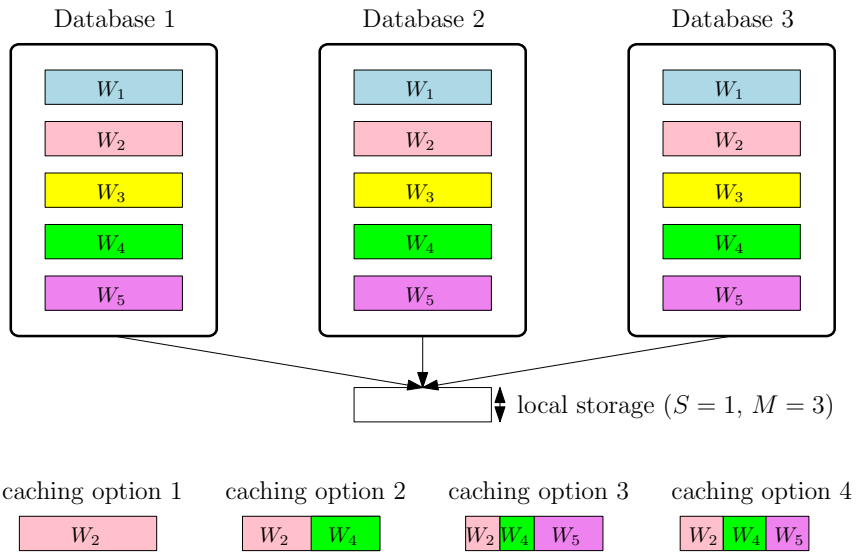

Figure 5.1: PIR-PSI under a storage constraint. Here $N = 3$, $K = 5$, $S = 1$, and $M = 3$.

load cost is

$$D^* = 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1-M}} + \frac{1-r_M}{N^{K-M}} + \frac{1-r_{M-1}}{N^{K-M+1}} + \cdots + \frac{1-r_1}{N^{K-1}} \quad (5.10)$$

where $r_1 \geq r_2 \geq \cdots \geq r_M$ *without loss of generality.*

The converse proof for Theorem 5.1 is given in Section 5.4, and the achievability proof for Theorem 5.1 is given in Section 5.5.

Remark 5.1 *For* $S = 0$*, by letting* $r_i = 0$*, for* $i = 1, \ldots, M$*,* (5.10) *reduces to*

$$D^* = 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1}}, \quad (5.11)$$

which is the optimal normalized download cost of the original PIR problem as shown in [12].

Remark 5.2 *For $S \in [K]$ and $M = S$, by letting $r_i = 1$ for $i = 1, \ldots, M$,* (5.10) *reduces to*

$$D^* = 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1-M}}, \tag{5.12}$$

which is the optimal normalized download cost of the PIR with PSI problem as shown in [33]. We can further generalize the result to the PIR with partially known PSI as shown in Chapter 4. Note further that for $M > S$, $(\frac{1-r_M}{N^{K-M}} + \frac{1-r_{M-1}}{N^{K-M+1}} + \cdots + \frac{1-r_1}{N^{K-1}})$ is the penalty to the download cost under the storage constraint.

Corollary 5.1 *For fixed $M \geq S$, uniform caching scheme achieves the lowest normalized download cost.*

Proof: The user has access to M messages. To achieve a low normalized download cost in (5.10), we need to solve the following optimization problem,

$$\begin{aligned} \min_{\alpha_i, i=1,\ldots,M} \quad & \alpha_M \frac{1}{N^{K-M}} + \alpha_{M-1} \frac{1}{N^{K-M+1}} + \cdots + \alpha_1 \frac{1}{N^{K-1}} \\ \text{s.t.} \quad & \alpha_M + \alpha_{M-1} + \cdots + \alpha_1 = M - S, \\ & 1 \geq \alpha_M \geq \alpha_{M-1} \geq \cdots \geq \alpha_1 \geq 0, \end{aligned} \tag{5.13}$$

which is obtained by replacing $1 - r_i$ in (5.10) with α_i for $i = 1, \ldots, M$. We prove by contradiction that the minimum is achieved when $\alpha_M = \alpha_{M-1}$. Suppose not, then we have optimum $\alpha_M^* > \alpha_{M-1}^*$. Choose $\delta = \frac{\alpha_M^* - \alpha_{M-1}^*}{3}$, and let $\alpha_M' = \alpha_M^* - \delta$, $\alpha_{M-1}' = \alpha_{M-1}^* + \delta$. Then, with α_M' and α_{M-1}', we achieve a lower normalized

download cost than with α_M^* and α_{M-1}^*, which gives a contradiction. Therefore, we have $\alpha_M = \alpha_{M-1}$. Intuitively, note that the coefficient of α_M is larger than the coefficient of α_{M-1} in the objective function in (5.13). Therefore, in order to minimize the objective function, we need to choose α_M as small as possible. But, since α_M needs to be larger than α_{M-1} according to the constraint set of (5.13), the smallest α_M we can choose is $\alpha_M = \alpha_{M-1}$. Using similar arguments, we also have $\alpha_{M-1} = \alpha_{M-2} = \cdots = \alpha_1$. Therefore, uniform caching achieves the lowest normalized download cost for fixed M. ■

Corollary 5.2 *For fixed S, among all the $K - \lceil S \rceil + 1$ uniform caching schemes, the uniform caching scheme with $M = K$ achieves the lowest normalized download cost.*

Proof: For the uniform caching scheme M, the user caches the first $\frac{S}{M}L$ symbols of each chosen message. From (5.10), the normalized download cost is

$$D^*(M) = 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1-M}} + \left(1 - \frac{S}{M}\right)\left(\frac{1}{N^{K-M}} + \cdots + \frac{1}{N^{K-1}}\right). \tag{5.14}$$

Considering the difference of the normalized download costs between $D^*(M+1)$ and $D^*(M)$,

$$\begin{aligned} &D^*(M+1) - D^*(M) \\ &\quad = 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-2-M}} + \left(1 - \frac{S}{M+1}\right)\left(\frac{1}{N^{K-M-1}} + \cdots + \frac{1}{N^{K-1}}\right) \end{aligned}$$

$$- \left[1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1-M}} + \left(1 - \frac{S}{M}\right)\left(\frac{1}{N^{K-M}} + \cdots + \frac{1}{N^{K-1}}\right)\right] \tag{5.15}$$

$$= -\frac{S}{M+1}\left(\frac{1}{N^{K-M-1}} + \cdots + \frac{1}{N^{K-1}}\right) + \frac{S}{M}\left(\frac{1}{N^{K-M}} + \cdots + \frac{1}{N^{K-1}}\right) \tag{5.16}$$

$$= -\frac{S}{M+1} \times \frac{1}{N^{K-M-1}} + \left(\frac{S}{M} - \frac{S}{M+1}\right)\left(\frac{1}{N^{K-M}} + \cdots + \frac{1}{N^{K-1}}\right) \tag{5.17}$$

$$= \frac{S}{M(M+1)}\left(\frac{1}{N^{K-M}} + \cdots + \frac{1}{N^{K-1}}\right) - \frac{S}{M(M+1)} \times \frac{M}{N^{K-M-1}} \tag{5.18}$$

$$\leq 0. \tag{5.19}$$

Thus, the uniform caching scheme with $M = K$ achieves the lowest normalized download cost among all possible uniform caching schemes. ■

Corollary 5.3 *For fixed S, among all possible caching schemes, the uniform caching scheme with $M = K$ achieves the lowest normalized download cost.*

Proof: From Corollary 5.1, we know that for fixed M, uniform caching scheme achieves the lowest normalized download cost. From Corollary 5.2, we know that among all uniform caching schemes, the uniform caching scheme with $M = K$ achieves the lowest normalized download cost. Combining these two corollaries, we conclude that among all possible caching schemes, the uniform caching scheme with $M = K$ achieves the lowest normalized download cost. ■

5.4 Converse Proof

In this section, we provide a lower bound for PIR-PSI under a storage constraint. In the following, without loss of generality, we relabel the messages according to $\mathbb{H}$, such that $W_{1:M}$ are the messages accessed by the user in the prefetching phase, where $W_{1:M} = (W_1, W_2, \ldots, W_M)$. Here, W_i denotes the message whose first Lr_i symbols are cached by the user, for $i = 1, 2, \ldots, M$, and without loss of generality, $r_1 \geq r_2 \geq \cdots \geq r_M$.

We first need the following lemma, which develops a lower bound on the length of the undesired portion of the answering strings as a consequence of the privacy constraint.

Lemma 5.1 (Interference lower bound) *For PIR-PSI under a storage constraint, the interference from undesired messages within the answering strings, $D - L$, is lower bounded by,*

$$D - L + o(L) \geq I\left(W_{1:K-1}; Q_{1:N}^{[K,\mathbb{H}]}, A_{1:N}^{[K,\mathbb{H}]} | \mathcal{W}_{\mathbb{H}}, \mathbb{H}, W_K\right). \tag{5.20}$$

If the privacy constraint is absent, the user downloads only L symbols of the desired message, however, when the privacy constraint is present, it should download D symbols. The difference between D and L, i.e., $D - L$, corresponds to the undesired portion of the answering strings. Note that Lemma 5.1 is an extension of [12, Lemma 5], where $M = 0$, i.e., the user has no PSI. Lemma 5.1 differs from its counterpart in Chapter 2 [Lemma 1] in two aspects; first, the left hand side is

$D(r) - L(1-r)$ in Chapter 2 as the user requests to download the uncached bits only, and second, Chapter 2 [Lemma 1] constructs $K-1$ distinct lower bounds by changing k, in contrast to only one bound here. In addition, we note that a similar argument to Lemma 5.1 can be implied from [33] and Chapter 4. The main difference between Lemma 5.1 and [33] and Chapter 4 is that $\mathcal{W}_{\mathbb{H}}$ refers to parts of messages here, while in [33] and Chapter 4, $\mathcal{W}_{\mathbb{H}}$ refers to full messages.

Proof: We start with the right hand side of (5.20),

$$I\left(W_{1:K-1}; Q_{1:N}^{[K,\mathbb{H}]}, A_{1:N}^{[K,\mathbb{H}]} | \mathcal{W}_{\mathbb{H}}, \mathbb{H}, W_K\right) \leq I\left(W_{1:K-1}; Q_{1:N}^{[K,\mathbb{H}]}, A_{1:N}^{[K,\mathbb{H}]}, W_K | \mathcal{W}_{\mathbb{H}}, \mathbb{H}\right). \tag{5.21}$$

For the right hand side of (5.21), we have

$$\begin{aligned}
&I\left(W_{1:K-1}; Q_{1:N}^{[K,\mathbb{H}]}, A_{1:N}^{[K,\mathbb{H}]}, W_K | \mathcal{W}_{\mathbb{H}}, \mathbb{H}\right) \\
&\quad = I\left(W_{1:K-1}; Q_{1:N}^{[K,\mathbb{H}]}, A_{1:N}^{[K,\mathbb{H}]} | \mathcal{W}_{\mathbb{H}}, \mathbb{H}\right) + I\left(W_{1:K-1}; W_K | Q_{1:N}^{[K,\mathbb{H}]}, A_{1:N}^{[K,\mathbb{H}]}, \mathcal{W}_{\mathbb{H}}, \mathbb{H}\right) && (5.22)\\
&\quad \overset{(5.6)}{=} I\left(W_{1:K-1}; Q_{1:N}^{[K,\mathbb{H}]}, A_{1:N}^{[K,\mathbb{H}]} | \mathcal{W}_{\mathbb{H}}, \mathbb{H}\right) + o(L) && (5.23)\\
&\quad \overset{(5.3),(5.4)}{=} I\left(W_{1:K-1}; A_{1:N}^{[K,\mathbb{H}]} | \mathcal{W}_{\mathbb{H}}, \mathbb{H}, Q_{1:N}^{[K,\mathbb{H}]}\right) + o(L) && (5.24)\\
&\quad = H\left(A_{1:N}^{[K,\mathbb{H}]} | \mathcal{W}_{\mathbb{H}}, \mathbb{H}, Q_{1:N}^{[K,\mathbb{H}]}\right) - H\left(A_{1:N}^{[K,\mathbb{H}]} | \mathcal{W}_{\mathbb{H}}, \mathbb{H}, Q_{1:N}^{[K,\mathbb{H}]}, W_{1:K-1}\right) + o(L) && (5.25)\\
&\quad \leq D - H\left(A_{1:N}^{[K,\mathbb{H}]} | \mathcal{W}_{\mathbb{H}}, \mathbb{H}, Q_{1:N}^{[K,\mathbb{H}]}, W_{1:K-1}\right) + o(L) && (5.26)\\
&\quad \overset{(5.6)}{=} D - H\left(A_{1:N}^{[K,\mathbb{H}]}, W_K | \mathcal{W}_{\mathbb{H}}, \mathbb{H}, Q_{1:N}^{[K,\mathbb{H}]}, W_{1:K-1}\right) + o(L) && (5.27)
\end{aligned}$$

$$\leq D - H\left(W_K|\mathcal{W}_{\mathbb{H}}, \mathbb{H}, Q_{1:N}^{[K,\mathbb{H}]}, W_{1:K-1}\right) + o(L) \tag{5.28}$$

$$\overset{(5.3),(5.4)}{=} D - L + o(L) \tag{5.29}$$

where (5.23), (5.27) follow from the decodability of W_K given $\left(Q_{1:N}^{[K,\mathbb{H}]}, A_{1:N}^{[K,\mathbb{H}]}, \mathcal{W}_{\mathbb{H}}, \mathbb{H}\right)$, (5.24), (5.29) follow from the independence of $W_{1:K}$ and $Q_{1:N}^{[K,\mathbb{H}]}$ given $\mathbb{H}$, and (5.26) follows from the independence bound. Combining (5.21) and (5.29) yields (5.20).

■

For the conditional mutual information term on the right hand side of (5.20), we have

$$I\left(W_{1:K-1}; Q_{1:N}^{[K,\mathbb{H}]}, A_{1:N}^{[K,\mathbb{H}]}|\mathcal{W}_{\mathbb{H}}, \mathbb{H}, W_K\right)$$

$$= \sum_h p(h) I\left(W_{1:K-1}; Q_{1:N}^{[K,h]}, A_{1:N}^{[K,h]}|\mathcal{W}_h, h, W_K\right) \tag{5.30}$$

$$= \sum_h p(h) I\left(W_{1:K-1}; Q_{1:N}^{[K,h]}, A_{1:N}^{[K,h]}|\mathcal{W}_h, W_K\right). \tag{5.31}$$

where we have written the mutual information in (5.20) as an expectation over all possible caching scheme realizations, as the databases do not know which messages are cached.

In the following lemma, we develop an inductive relation for the mutual information term on the right hand side of (5.31).

Lemma 5.2 (Fractional induction lemma) *For all $k \in \{1, \ldots, K-1\}$, the mu-*

tual information term in (5.31) *can be inductively lower bounded as,*

$$I\left(W_{1:k}; Q_{1:N}^{[k+1,h]}, A_{1:N}^{[k+1,h]} | \mathcal{W}_h, W_{k+1:K}\right) \geq \frac{1}{N} I\left(W_{1:k-1}; Q_{1:N}^{[k,h]}, A_{1:N}^{[k,h]} | \mathcal{W}_h, W_{k:K}\right) + \frac{L}{N}(1 - r_k) - o(L), \tag{5.32}$$

where $r_k = 0$ *when* $k > M$.

Lemma 5.2 is a generalization of [12, Lemma 6] to our setting. The main difference between Lemma 5.2 and [12, Lemma 6] is that the cached PSI results in a different induction relation.

Proof: We start with the left hand side of (5.32),

$$\begin{aligned}
&I\left(W_{1:k}; Q_{1:N}^{[k+1,h]}, A_{1:N}^{[k+1,h]} | \mathcal{W}_h, W_{k+1:K}\right) \\
&= \frac{1}{N} \times N \times I\left(W_{1:k}; Q_{1:N}^{[k+1,h]}, A_{1:N}^{[k+1,h]} | \mathcal{W}_h, W_{k+1:K}\right) && (5.33)\\
&\geq \frac{1}{N} \sum_{n=1}^{N} I\left(W_{1:k}; Q_n^{[k+1,h]}, A_n^{[k+1,h]} | \mathcal{W}_h, W_{k+1:K}\right) && (5.34)\\
&\overset{(5.7)}{=} \frac{1}{N} \sum_{n=1}^{N} I\left(W_{1:k}; Q_n^{[k,h]}, A_n^{[k,h]} | \mathcal{W}_h, W_{k+1:K}\right) && (5.35)\\
&\geq \frac{1}{N} \sum_{n=1}^{N} I\left(W_{1:k}; A_n^{[k,h]} | \mathcal{W}_h, W_{k+1:K}, Q_n^{[k,h]}\right) && (5.36)\\
&\overset{(5.5)}{=} \frac{1}{N} \sum_{n=1}^{N} H\left(A_n^{[k,h]} | \mathcal{W}_h, W_{k+1:K}, Q_n^{[k,h]}\right) && (5.37)\\
&\geq \frac{1}{N} \sum_{n=1}^{N} H\left(A_n^{[k,h]} | \mathcal{W}_h, W_{k+1:K}, Q_{1:N}^{[k,h]}, A_{1:n-1}^{[k,h]}\right) && (5.38)\\
&\overset{(5.5)}{=} \frac{1}{N} \sum_{n=1}^{N} I\left(W_{1:k}; A_n^{[k,h]} | \mathcal{W}_h, W_{k+1:K}, Q_{1:N}^{[k,h]}, A_{1:n-1}^{[k,h]}\right) && (5.39)\\
&= \frac{1}{N} I\left(W_{1:k}; A_{1:N}^{[k,h]} | \mathcal{W}_h, W_{k+1:K}, Q_{1:N}^{[k,h]}\right) && (5.40)
\end{aligned}$$

$$\overset{(5.3),(5.4)}{=} \frac{1}{N} I\left(W_{1:k}; Q_{1:N}^{[k,h]}, A_{1:N}^{[k,h]} | \mathcal{W}_h, W_{k+1:K}\right) \tag{5.41}$$

$$\overset{(5.6)}{=} \frac{1}{N} I\left(W_{1:k}; W_k, Q_{1:N}^{[k,h]}, A_{1:N}^{[k,h]} | \mathcal{W}_h, W_{k+1:K}\right) - o(L) \tag{5.42}$$

$$= \frac{1}{N} I\left(W_{1:k}; W_k | \mathcal{W}_h, W_{k+1:K}\right) + \frac{1}{N} I\left(W_{1:k}; Q_{1:N}^{[k,h]}, A_{1:N}^{[k,h]} | \mathcal{W}_h, W_{k:K}\right) - o(L) \tag{5.43}$$

$$= \frac{1}{N} I\left(W_{1:k}; Q_{1:N}^{[k,h]}, A_{1:N}^{[k,h]} | \mathcal{W}_h, W_{k:K}\right) + \frac{L}{N}(1 - r_k) - o(L), \tag{5.44}$$

where (5.34) and (5.36) follow from the chain rule and the non-negativity of mutual information, (5.35) follows from the privacy constraint, (5.37), (5.39) follow from the fact that answer strings are deterministic functions of the messages and the queries, (5.38) follows from the fact that conditioning reduces entropy, (5.41) follows from the independence of $W_{1:K}$ and $Q_{1:N}^{[k,h]}$, (5.42) follows from the reliability constraint on W_k, and (5.44) is due to the fact that $H\left(W_k | \mathcal{W}_h, W_{k+1:K}\right) = L(1 - r_k)$, where if $k \notin h$ then $r_k = 0$. ■

By applying Lemma 5.2 recursively to the right hand side of (5.31)

$$I\left(W_{1:K-1}; Q_{1:N}^{[K,h]}, A_{1:N}^{[K,h]} | \mathcal{W}_h, W_K\right)$$

$$\overset{(5.32)}{\geq} \frac{1}{N} I\left(W_{1:K-2}; Q_{1:N}^{[K-1,h]}, A_{1:N}^{[K-1,h]} | \mathcal{W}_h, W_{K-1:K}\right) + \frac{L}{N} - o(L) \tag{5.45}$$

$$\overset{(5.32)}{\geq} \frac{1}{N^2} I\left(W_{1:K-3}; Q_{1:N}^{[K-2,h]}, A_{1:N}^{[K-2,h]} | \mathcal{W}_h, W_{K-2:K}\right) + \frac{L}{N^2} + \frac{L}{N} - o(L) \tag{5.46}$$

$$\overset{(5.32)}{\geq} \dots \tag{5.47}$$

$$\overset{(5.32)}{\geq} \frac{1}{N^{K-1-M}} I\left(W_{1:M}; Q_{1:N}^{[M+1,h]}, A_{1:N}^{[M+1,h]} | \mathcal{W}_h, W_{M+1:K}\right) + \frac{L}{N^{K-1-M}} + \dots + \frac{L}{N^2} + \frac{L}{N} - o(L) \tag{5.48}$$

$$\overset{(5.32)}{\geq} \frac{1}{N^{K-M}} I\left(W_{1:M-1}; Q_{1:N}^{[M,h]}, A_{1:N}^{[M,h]} | \mathcal{W}_h, W_{M:K}\right) + \frac{L}{N^{K-M}}(1 - r_M)$$

$$+\frac{L}{N^{K-1-M}}+\cdots+\frac{L}{N^2}+\frac{L}{N}-o(L) \tag{5.49}$$

$$\overset{(5.32)}{\geq} \ldots \tag{5.50}$$

$$\overset{(5.32)}{\geq} \frac{L(1-r_1)}{N^{K-1}}+\cdots+\frac{L(1-r_M)}{N^{K-M}}+\cdots+\frac{L}{N^2}+\frac{L}{N}-o(L). \tag{5.51}$$

Note that in (5.45) to (5.48), we apply the fractional induction lemma with $r=0$, since $W_{M+1:K}$ are not cached in $\mathcal{W}_h$. In (5.49) to (5.51), $r_k>0$ for the fractional induction lemma, since $W_{1:M}$ are cached in $\mathcal{W}_h$ partially.

By combining (5.20), (5.31), and (5.51), and dividing by L on both sides, we obtain a lower bound for the normalized download cost as

$$D^* \geq 1+\frac{1}{N}+\frac{1}{N^2}+\cdots+\frac{1}{N^{K-1-M}}+\frac{1-r_M}{N^{K-M}}+\frac{1-r_{M-1}}{N^{K-M+1}}+\cdots+\frac{1-r_1}{N^{K-1}}, \tag{5.52}$$

which proves (5.10).

5.5 Achievability Proof

Our achievability scheme is based on the PIR schemes in [12] and [33]. For the portion of the messages not cached by the user, we use the PIR scheme in [12], which applies the following three principles recursively: 1) database symmetry, 2) message symmetry within each database, and 3) exploiting undesired messages as side information. For the portion of the messages cached by the user, we use the PIR scheme in [33], which is based on MDS codes and consists of two stages: The first stage determines the systematic part of the MDS code according to the queries

generated in [12]. In the second stage, the user reduces the download cost by downloading the parity part of the MDS code only. By applying the two PIR schemes, the user retrieves the desired message privately while keeping the cached messages private.

5.5.1 Motivating Examples

5.5.1.1 $N = 2$ Databases, $K = 5$ Messages, $M = 2$ Accessed Messages, and $S = 1$ with Uniform Caching

In this example, in the prefetching phase, the user randomly chooses two messages to cache, say W_1 and W_4. Since $S = 1$ and the user uses uniform caching scheme, the user caches the first half of W_1 and the first half of W_4. We note that the databases are aware of the caching scheme, i.e., the databases know that two out of five messages are chosen by the user, and the first halves of the chosen messages are cached. However, the databases do not know which are the two chosen messages.

In the retrieval phase, assume that the user wishes to retrieve message W_3 privately. For the first half portion of the message, i.e., for the symbols in the interval $[0, \frac{L}{2}]$, since the user has cached messages W_1 and W_4, the user applies the PIR scheme in [33] with $M = 2$. The total download cost for the first half portion of the message, as shown in (5.12), is

$$\frac{L}{2} \times \left(1 + \frac{1}{2} + \frac{1}{2^{5-1-2}}\right). \tag{5.53}$$

For the remaining half portion of the message, i.e., for the symbols in the interval $[\frac{L}{2}, L]$, since the user has not cached any messages, the user applies the PIR scheme in [12]. The total download cost for the remaining half portion of the message, as shown in (5.11), is

$$\frac{L}{2} \times \left(1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^{5-1}}\right). \tag{5.54}$$

The overall download cost is the sum of (5.53) and (5.54). Therefore, the optimal normalized download cost is $\frac{59}{32}$, which can also be obtained through (5.10) by letting $r_1 = \frac{1}{2}$ and $r_2 = \frac{1}{2}$. Note that since we have applied the PIR scheme in [33] to retrieve the first half portion of the message, the databases cannot learn which messages are cached by the user. In addition, both PIR schemes in [12] and [33] keep the identity of the desired message private. Therefore, the combination of these two PIR schemes is a feasible PIR scheme for PIR-PSI a under storage constraint [18].

5.5.1.2 $N = 2$ Databases, $K = 5$ Messages, $S = 1$, $M = 3$ with $r_1 = \frac{1}{2}$, and $r_2 = r_3 = \frac{1}{4}$

In this example, see Fig. 5.2, in the prefetching phase, since $r_1 = \frac{1}{2}$, the user first randomly chooses one message to cache, say W_3, and the user caches the first half of W_3. Since $r_2 = r_3 = \frac{1}{4}$, the user then randomly chooses two other messages to cache, say W_2 and W_5, and the user caches the first $\frac{1}{4}$ portions of W_2 and W_5. Note that $S = 1$ and $\frac{1}{2} \times 1 + \frac{1}{4} \times 2 = 1$, and the local cache memory size constraint is satisfied. We note that the databases are aware of the caching strategy, i.e., the

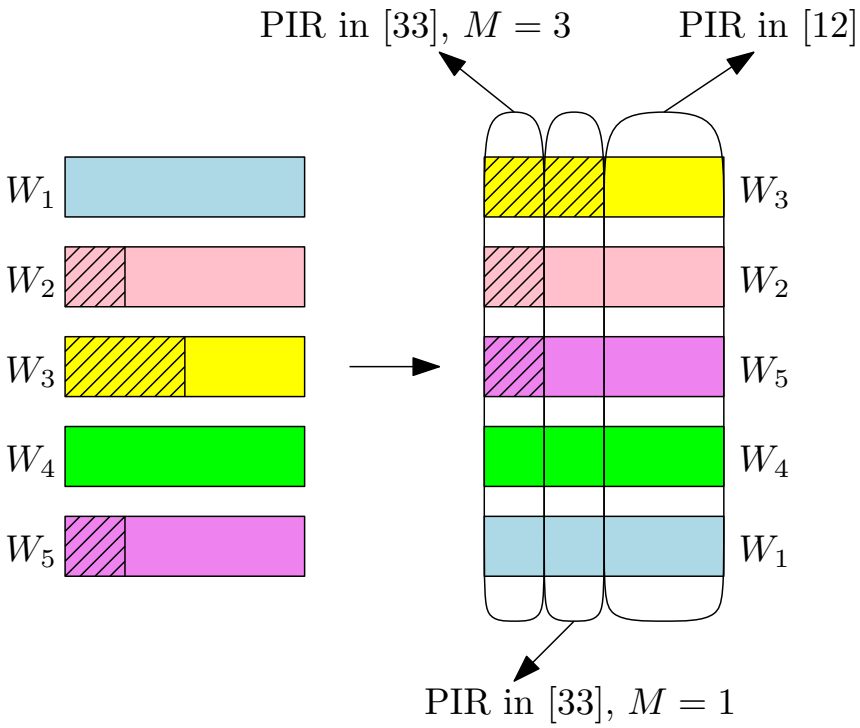

Figure 5.2: Achievable scheme: $K = 5$, $S = 1$, and $M = 3$ with $r_1 = \frac{1}{2}$, and $r_2 = r_3 = \frac{1}{4}$.

databases know that three out of five messages are chosen by the user, and for one of the chosen message, the first half of the message is cached, and for the remaining two chosen messages, the first $\frac{1}{4}$ portions are cached. However, the databases do not know which three messages are chosen.

In the retrieval phase, assume that the user wishes to retrieve message W_1 privately. For the first $\frac{1}{4}$ portion of messages, i.e., for the symbols in the interval $[0, \frac{L}{4}]$, since the user caches messages W_2, W_3 and W_5, the user applies the PIR scheme in [33] with $M = 3$. The total download cost for the first $\frac{1}{4}$ portion of the message, as shown in (5.12), is

$$\frac{L}{4} \times \left(1 + \frac{1}{2^{5-1-3}}\right). \tag{5.55}$$

For the following $\frac{1}{4}$ portion of messages, i.e., for the symbols in the interval $[\frac{L}{4}, \frac{L}{2}]$, since the user caches message W_3, the user applies the PIR scheme in [33] with $M = 1$. The total download cost for the second $\frac{1}{4}$ portion of the message, as shown in (5.12), is

$$\frac{L}{4} \times \left(1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^{5-1-1}}\right). \tag{5.56}$$

For the last half portion of messages, i.e., for the symbols in the interval $[\frac{L}{2}, L]$, since the user has not cached any messages, the user applies the PIR scheme in [12]. The total download cost for the last half portion of the message, as shown in (5.11), is

$$\frac{L}{2} \times \left(1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^{5-1}}\right). \tag{5.57}$$

The overall download cost is the sum of (5.55), (5.56) and (5.57). Therefore, the optimal normalized download cost is $\frac{29}{16}$, which can also be obtained through (5.10) by letting $r_1 = \frac{1}{2}$, and $r_2 = r_3 = \frac{1}{4}$. Note that by applying the PIR scheme in [33] to retrieve the first $\frac{1}{4}$ portion and the middle $\frac{1}{4}$ portion of the message, the databases cannot learn which messages have been cached by the user. In addition, both PIR schemes in [12] and [33] hide the identity of the desired message. Therefore, the combination of these two PIR schemes is a feasible PIR scheme for PIR-PSI under a storage constraint [18].

5.5.2 General Achievable Scheme

We now describe the general achievable scheme for $r_1 \geq r_2 \geq \cdots \geq r_M$. We first consider the first r_M fraction of messages, i.e., for the symbols in the interval $[0, Lr_M]$. Since $r_1 \geq r_2 \geq \cdots \geq r_M$, the user caches M messages for this portion. The user applies the PIR scheme in [33] which results in the download cost

$$Lr_M \times \left(1 + \frac{1}{N} + + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1-M}}\right). \tag{5.58}$$

Following the same logic, for the symbols in the interval $[Lr_i, Lr_{i-1}]$, $i \geq 2$, the user caches i messages for this portion. The user applies the PIR scheme in [33] which results in the download cost

$$L(r_{i-1} - r_i) \times \left(1 + \frac{1}{N} + + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-i}}\right). \tag{5.59}$$

Lastly, for the symbols in the interval $[Lr_1, L]$, the user caches no messages for this portion. The user applies the PIR scheme in [12] which results in the download cost

$$L(1 - r_1) \times \left(1 + \frac{1}{N} + + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1}}\right). \tag{5.60}$$

The overall download cost is the sum of (5.58), (5.59) for $i = 2, 3, \ldots, M$, and (5.60), which is (5.10). By applying the PIR scheme in [33] to retrieve symbols in the interval of $[0, Lr_1]$, the databases cannot learn which messages have been cached by the user. In addition, both PIR schemes in [12] and [33] protect the identity

of the desired message. Therefore, the combination of these two PIR schemes is a feasible PIR scheme for PIR-PSI under a storage constraint [18].

5.6 Conclusion

In this chapter, we have introduced a new PIR model, namely PIR-PSI under a storage constraint. In this model, the user randomly chooses M messages and caches the first r_i portion of the chosen messages for $i = 1, \ldots, M$ subject to the memory size constraint $\sum_{i=1}^{M} r_i = S$. In the retrieval phase, the user wishes to retrieve a message such that no individual database can learn the identity of the desired message and the identities of the cached messages. For each caching scheme, i.e., $(r_1, \ldots, r_M)$, we characterized the optimal normalized download cost to be $D^* = 1+\frac{1}{N}+\frac{1}{N^2}+\cdots+\frac{1}{N^{K-1-M}}+\frac{1-r_M}{N^{K-M}}+\frac{1-r_{M-1}}{N^{K-M+1}}+\cdots+\frac{1-r_1}{N^{K-1}}$. In addition, we showed that, for a fixed memory size S, and a fixed number of accessible messages M, uniform caching achieves the lowest normalized download cost, where uniform caching means $r_i = \frac{S}{M}$, $i = 1, \ldots, M$. Then, we showed that, for a fixed memory size S, among all $K - \lceil S \rceil + 1$ uniform caching schemes, the uniform caching scheme caching $M = K$ messages achieves the lowest normalized download cost. Finally, we conclude that for a fixed memory size S, the uniform caching scheme caching K messages achieves the lowest normalized download cost.

CHAPTER 6

The Capacity of Private Information Retrieval from Decentralized Uncoded Caching Databases

6.1 Introduction

We consider the private information retrieval (PIR) problem from decentralized uncoded caching databases. There are two phases in our problem setting, a caching phase, and a retrieval phase. In the caching phase, a data center containing all the K files, where each file is of size L bits, and several databases with storage size constraint μKL bits exist in the system. Each database independently chooses μKL bits out of the total KL bits from the data center to cache through the same probability distribution in a decentralized manner. In the retrieval phase, a user (retriever) accesses N databases in addition to the data center, and wishes to retrieve a desired file privately. We characterize the optimal normalized download cost to be $\frac{D}{L} = \sum_{n=1}^{N+1} \binom{N}{n-1} \mu^{n-1} (1-\mu)^{N+1-n} \left(1 + \frac{1}{n} + \cdots + \frac{1}{n^{K-1}}\right)$. We show that uniform and random caching scheme which is originally proposed for decentralized coded caching by Maddah-Ali and Niesen [54], along with Sun and Jafar retrieval scheme [12] which is originally proposed for PIR from replicated databases surprisingly result in

the lowest normalized download cost. This is the decentralized counterpart of the recent result of Attia, Kumar and Tandon [43] for the centralized case. The converse proof contains several ingredients such as interference lower bound, induction lemma, replacing queries and answering string random variables with the content of distributed databases, the nature of decentralized uncoded caching databases, and bit marginalization of joint caching distributions.

6.2 System Model

We consider a system consisting of one data center and several databases. The data center stores K independent files, labeled as W_1, W_2, $\ldots$, W_K, where each file is of size L bits. Therefore,

$$H(W_1) = \cdots = H(W_K) = L, \qquad H(W_1, \ldots, W_K) = H(W_1) + \cdots + H(W_K). \tag{6.1}$$

Each database has a storage capacity of μKL bits, where $0 \leq \mu \leq 1$.

The system operates in two phases: In the caching phase, we consider the case of *uncoded* caching, i.e., each database stores a subset of bits from the data center. Due to the storage size constraint, each database at most stores μKL bits out of the total KL bits from the data center. Here, we denote ith database as DB_i and use random variable Z_i to denote the stored content in DB_i. Therefore, the storage

size constraint for DB_i is

$$H(Z_i) \leq \mu KL. \tag{6.2}$$

We consider the *decentralized* setting for the caching phase, i.e., each database chooses a subset of bits to store independently according to the same probability distribution, denoted by P_H. Rigorously, let random variable H_i denote the indices of the stored bits in DB_i. For N databases, the decentralized caching scheme $\mathcal{H}$ can be specified as

$$\mathbb{P}(\mathcal{H} = (H_1, \ldots, H_N)) = \prod_{i=1}^{N} P_H(H_i). \tag{6.3}$$

In the retrieval phase, the user accesses N databases and the data center. We note that we do not know in advance which N databases are available or which N databases the user will have access to. Here, we also assume that in the retrieval phase, the data center and N databases do not communicate with each other (no collusion). To simplify the notation, we use DB_0 to denote the data center, and therefore $Z_0 = (W_1, \ldots, W_K)$ since the data center stores all the K files. The user privately generates an index $\theta \in [K] = \{1, \ldots, K\}$, and wishes to retrieve file W_θ such that it is impossible for either the data center or any individual database to identify θ. For random variables θ, and $W_1, \ldots, W_K$, we have

$$H(\theta, W_1, \ldots, W_K) = H(\theta) + H(W_1) + \cdots + H(W_K). \tag{6.4}$$

In order to retrieve file W_θ, the user sends $N+1$ queries $Q_0^{[\theta]}, \ldots, Q_N^{[\theta]}$ to DB_0, $\ldots$, DB_N, where $Q_n^{[\theta]}$ is the query sent to DB_n for file W_θ. Note that the queries are independent of the realization of the K files. Therefore,

$$I(W_1, \ldots, W_K; Q_0^{[\theta]}, \ldots, Q_N^{[\theta]}) = 0. \tag{6.5}$$

Upon receiving the query $Q_n^{[\theta]}$, DB_n replies with an answering string $A_n^{[\theta]}$, which is a function of $Q_n^{[\theta]}$ and Z_n. Therefore, $\forall \theta \in [K], \forall n \in \{0\} \cup [N]$,

$$H(A_n^{[\theta]} | Q_n^{[\theta]}, Z_n) = 0. \tag{6.6}$$

After receiving the answering strings $A_0^{[\theta]}, \ldots, A_N^{[\theta]}$ from DB_0, $\ldots$, DB_N, the user needs to decode the desired file W_θ reliably. By using Fano's inequality, we have the following reliability constraint

$$H\left(W_\theta | Q_0^{[\theta]}, \ldots, Q_N^{[\theta]}, A_0^{[\theta]}, \ldots, A_N^{[\theta]}\right) = o(L), \tag{6.7}$$

where $o(L)$ denotes a function such that $\frac{o(L)}{L} \to 0$ as $L \to \infty$.

To ensure that individual databases do not know which file is retrieved, we have the following privacy constraint, $\forall n \in \{0\} \cup [N]$, $\forall \theta \in [K]$,

$$(Q_n^{[1]}, A_n^{[1]}, W_1, \ldots, W_K) \sim (Q_n^{[\theta]}, A_n^{[\theta]}, W_1, \ldots, W_K), \tag{6.8}$$

where $A \sim B$ means that A and B are identically distributed.

Given that each file is of size L bits, for a fixed K, μ and decentralized caching probability distribution P_H, let $\mathcal{H}$ denote the indices of the cached bits in the N databases available in the retrieval phase. The probability distribution of $\mathcal{H}$ is specified in (6.3). Let $D_{\mathcal{H}}^{[\theta]}$ represent the number of downloaded bits via the answering strings $A_{0:N}^{[\theta]}$, where $A_{0:N}^{[\theta]} = (A_0^{[\theta]}, \ldots, A_N^{[\theta]})$. Then,

$$D_{\mathcal{H}}^{[\theta]} = \sum_{n=0}^{N} H\left(A_n^{[\theta]}\right). \tag{6.9}$$

We further denote $D_{\mathcal{H}}$ as the expected number of downloaded bits with respect to different file requests, i.e., $D_{\mathcal{H}} = E_\theta\left[D_{\mathcal{H}}^{[\theta]}\right]$. Finally, we denote D as the expected number of downloaded bits with respect to different realization of the cached bit indices, i.e., $D = E_{\mathcal{H}}\left[D_{\mathcal{H}}\right]$. A pair (D, L) is achievable if there exists a PIR scheme satisfying the reliability constraint (6.7) and the privacy constraint (6.8). The optimal normalized download cost D^* is defined as

$$D^* = \inf\left\{\frac{D}{L} : (D, L) \text{ is achievable}\right\}. \tag{6.10}$$

In this work, we aim at characterizing the optimal normalized download cost and finding the optimal decentralized caching probability distribution.

Next, we illustrate the system model and the problem considered with a simple example of $K = 3$ files and $N = 2$ databases in the retrieval phase; see Fig. 6.1. Consider a data center storing $K = 3$ files where each file is of size 4 bits. In the caching phase, there are 4 databases in the system, and each database can at most

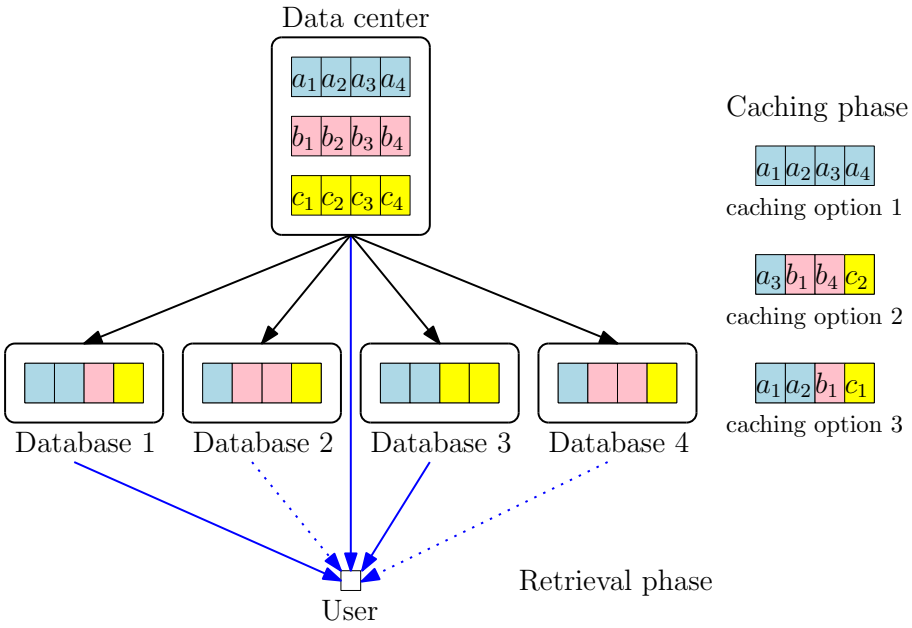

Figure 6.1: PIR from decentralized caching databases with $K = 3$, $N = 2$, and $\mu = \frac{1}{3}$.

store 4 bits. Each database can always store the first file, which is of size 4 bits, as caching option 1 in Fig. 6.1. Or each database can uniformly and randomly choose 4 bits out of total 12 bits from the data center to store. One of the realization is shown as caching option 2 in Fig. 6.1. Each database can also choose 2 bits from the first file and 1 bit each from the remaining two files to store, where one of the realization is shown as caching option 3 in Fig. 6.1. We require each database to use the same probability distribution to choose the bits to store in order to satisfy the decentralized requirement. In this example, we assume that the user can access the data center and $N = 2$ databases in the retrieval phase, say the first and the third database, and the user wishes to download a file privately. Our questions are as follows: What is the optimal probability distribution to use in the caching phase? What is the optimal PIR scheme to use in the retrieval phase? How can we jointly

design the schemes in the two phases such that the expected normalized download cost is the lowest in the second phase?

6.3 Main Results and Discussions

We characterize the optimal normalized download cost for PIR from decentralized uncoded caching databases in the following theorem.

Theorem 6.1 *For PIR from decentralized uncoded caching databases with K files, where each file is of size L bits, N databases in addition to a data center available in the retrieval phase, and a storage size constraint μKL, $0 < \mu < 1$, bits for each database, the optimal normalized download cost is*

$$\frac{D}{L} = \sum_{n=1}^{N+1} \binom{N}{n-1} \mu^{n-1} (1-\mu)^{N+1-n} \left(1 + \frac{1}{n} + \cdots + \frac{1}{n^{K-1}}\right). \tag{6.11}$$

The achievability scheme is provided in Section 6.4, and the converse proof is shown in Section 6.5. We first use the following example to show the main ingredients of Theorem 6.1.

6.3.1 Motivating Example: $K = 3$ and $N = 2$

In this example, we consider the case where the data center stores $K = 3$ independent files labeled as A, B, and C, where each file is of size L bits. In the caching phase, several databases with storage capacity of $3\mu L$ bits are present in the system. We will show that the optimal normalized download cost is $\frac{D}{L} = \frac{17}{18}\mu^2 - \frac{5}{2}\mu + 3$ when

$N = 2$ databases in addition to the data center are available in the retrieval phase.

6.3.1.1 Achievability Scheme

In the caching phase, to satisfy the storage size constraint, each database randomly and uniformly stores $3\mu L$ bits out of total $3L$ bits from the data center. Each database operates independently through the same probability distribution resulting in decentralized caching.

In the retrieval phase, suppose $N = 2$ databases, labeled as DB_1 and DB_2, in addition to the data center, labeled as DB_0, are available to the user, and the user wishes to retrieve file A privately. Let us first focus on one file, say A. We can partition file A into four subfiles

$$A = (A_0, A_{0,1}, A_{0,2}, A_{0,1,2}), \tag{6.12}$$

where, for $S \subseteq \{0, 1, 2\}$, A_S denotes the bits of file A which are stored in databases in S. For example, A_0 denotes the bits of file A only stored in DB_0 and $A_{0,2}$ denotes the bits of file A stored in DB_0 and DB_2 and so on. Since each bit is stored in the data center, 0 exists in the label of every partition. By the law of large numbers,

$$|A_S| = L\mu^{|S|-1}(1-\mu)^{3-|S|} + o(L), \tag{6.13}$$

when the file size is large enough. We can do the same partitions for files B and C.

To retrieve file A privately, we first retrieve the subfile $A_{0,1,2}$ privately. We

apply the PIR scheme proposed in [12] to retrieve the subfile $A_{0,1,2}$. Subfile $A_{0,1,2}$ is replicated in 3 databases and the total number of files is 3 since we also have $B_{0,1,2}$ and $C_{0,1,2}$. Therefore, we download

$$L\mu^2\left(1+\frac{1}{3}+\frac{1}{9}\right)+o(L) \tag{6.14}$$

bits. We also need to retrieve the subfile $A_{0,1}$ privately. Subfile $A_{0,1}$ is replicated in 2 databases and the total number of files is 3 since we also have $B_{0,1}$ and $C_{0,1}$. By applying the PIR scheme in [12], we download

$$L\mu(1-\mu)\left(1+\frac{1}{2}+\frac{1}{4}\right)+o(L) \tag{6.15}$$

bits. Next, we need to retrieve the subfile $A_{0,2}$ privately. Using [12], we download

$$L\mu(1-\mu)\left(1+\frac{1}{2}+\frac{1}{4}\right)+o(L) \tag{6.16}$$

bits. Finally, we need to retrieve A_0 privately. Using [12], we download

$$L(1-\mu)^2(1+1+1)+o(L) \tag{6.17}$$

bits. By adding (6.14), (6.15), (6.16) and (6.17), we show that the normalized download cost

$$\frac{17}{18}\mu^2-\frac{5}{2}\mu+3 \tag{6.18}$$

is achievable.

6.3.1.2 Converse Proof

Here, we show that among all the decentralized caching probability distributions P_H, the lowest normalized download cost for $N = 2$ databases is as shown in (6.18). Given a decentralized caching probability distribution P_H, we have a resulting $\mathcal{H}$ in the retrieval phase.

We lower bound $D_{\mathcal{H}}$ first. In the retrieval phase, the stored content of DB_0, DB_1, and DB_2 are fixed and uncoded, i.e., Z_0, Z_1 and Z_2 are fixed and uncoded. We can apply the lower bound in [43, Eqn. (31)] as the lower bound for $D_{\mathcal{H}}$. Therefore,

$$D_{\mathcal{H}} \geq L + \frac{4}{27}\sum_{k=1}^{3} H(W_k) + \frac{11}{108}\sum_{i=0}^{2}\sum_{k=1}^{3} H(W_k|Z_i) + \frac{17}{54}\sum_{i=0}^{2}\sum_{k=1}^{3} H(W_k|Z_{[0:2]\setminus i}) + o(L) \tag{6.19}$$

$$= \frac{13}{9}L + \frac{11}{108}\sum_{i=1}^{2}\sum_{k=1}^{3} H(W_k|Z_i) + \frac{17}{54}\sum_{k=1}^{3} H(W_k|Z_1, Z_2) + o(L) \tag{6.20}$$

$$\geq \frac{13}{9}L + \frac{11}{108}\left(3L - 3\mu L + 3L - 3\mu L\right) + \frac{17}{54}\sum_{k=1}^{3} H(W_k|Z_1, Z_2) + o(L) \tag{6.21}$$

$$= \frac{37}{18}L - \frac{11}{18}\mu L + \frac{17}{54}H(W_{1:3}|Z_1, Z_2) + o(L), \tag{6.22}$$

where (6.20) holds due to $Z_0 = (W_1, W_2, W_3)$, and (6.21) holds due to (6.2). We note that different $\mathcal{H}$ results in different Z_1 and Z_2.

We lower bound D now. From (6.22), we have

$$D = E_{\mathcal{H}}\left[D_{\mathcal{H}}\right] \geq \frac{37}{18}L - \frac{11}{18}\mu L + \frac{17}{54}E_{\mathcal{H}}\left[H(W_{1:3}|Z_1, Z_2)\right] + o(L). \tag{6.23}$$

Let random variables $X_{i,j}^{(n)}$, $i = 1, \ldots, L$, $j = 1, \ldots, K$, be the indicator functions showing that the ith bit of file W_j is cached in DB_n or not, i.e., $X_{i,j}^{(n)} = 1$ means that the ith bit of file W_j is stored in DB_n and $X_{i,j}^{(n)} = 0$ means that it is not stored in DB_n. For DB_1 we have

$$X_{1,1}^{(1)} + \cdots + X_{L,1}^{(1)} + X_{1,2}^{(1)} + \cdots + X_{L,2}^{(1)} + X_{1,3}^{(1)} + \cdots + X_{L,3}^{(1)} \leq 3\mu L \tag{6.24}$$

due to the storage size constraint in (6.2). We note that P_H induces probability measures on random variables $X_{i,j}^{(n)}$, and let $X_{i,j}^{(n)} = 1$ with probability $p_{i,j}$, where we remove the superscript n since each database adopts the same probability distribution P_H to choose the cached bits due to the decentralized property. By taking expectation on (6.24) and applying the linearity of expectation, we have

$$E[X_{1,1}^{(1)}] + \cdots + E[X_{L,3}^{(1)}] \leq 3\mu L, \tag{6.25}$$

which yields

$$p_{1,1} + \cdots + p_{L,3} \leq 3\mu L. \tag{6.26}$$

Let random variables $V_{i,j}$, $i = 1, \ldots, L$, $j = 1, \ldots, K$, be the indicator functions showing that the ith bit of file W_j is not cached in DB_1 and DB_2, i.e., $V_{i,j} = 1$ means that the ith bit of file W_j is not stored in either DB_1 or DB_2. Therefore, we

have

$$V_{i,j} = (1 - X_{i,j}^{(1)})(1 - X_{i,j}^{(2)}). \tag{6.27}$$

Now, we can evaluate $E_{\mathcal{H}}\left[H(W_{1:3}|Z_1, Z_2)\right]$ in (6.23) as follows

$$E_{\mathcal{H}}\left[H(W_{1:3}|Z_1, Z_2)\right] = E[V_{1,1} + \cdots + V_{L,3}] \tag{6.28}$$

$$= E[V_{1,1}] + \cdots + E[V_{L,3}] \tag{6.29}$$

$$= (1 - p_{1,1})^2 + \cdots + (1 - p_{L,3})^2. \tag{6.30}$$

Therefore, continuing from (6.23), we have

$$D \geq \frac{37}{18}L - \frac{11}{18}\mu L + \frac{17}{54}\left[(1 - p_{1,1})^2 + \cdots + (1 - p_{L,3})^2\right] + o(L), \tag{6.31}$$

where $p_{1,1}, \ldots, p_{L,3}$ are subject to (6.26). To further lower bound the right hand side of (6.31), we minimize the right hand side with respect to $p_{i,j}$ subject to (6.26). Hence, we consider the following Lagrangian

$$L(p_{1,1}, \ldots, p_{L,3}, \lambda) = (1 - p_{1,1})^2 + \cdots + (1 - p_{L,3})^2 + \lambda\left(p_{1,1} + \cdots + p_{L,3} - 3\mu L\right). \tag{6.32}$$

From the KKT conditions, we have

$$\lambda = 2(1 - p_{i,j}), \quad i = 1, \ldots, L, \quad j = 1, 2, 3. \tag{6.33}$$

Thus, we can further lower bound (6.31) by letting $p_{1,1} = \cdots = p_{L,3} = \mu$, and we have

$$\frac{D}{L} \geq \frac{37}{18} - \frac{11}{18}\mu + \frac{17}{54}\left[3(1-\mu)^2\right] + \frac{o(L)}{L} \tag{6.34}$$

$$= \frac{17}{18}\mu^2 - \frac{5}{2}\mu + 3 + \frac{o(L)}{L}. \tag{6.35}$$

Therefore, we show that the optimal normalized download cost is $\frac{17}{18}\mu^2 - \frac{5}{2}\mu + 3$ when $N = 2$ databases in addition to the data center are available in the retrieval phase. To achieve the optimal normalized download cost, each database should randomly and uniformly store the bits in the caching phase.

6.3.2 Further Examples and Numerical Results

Now, we use different scenarios to illustrate the optimal normalized download cost in (6.11). We first consider the scenario where the data center contains $K = 10$ files, each database with storage size constraint $\mu = \frac{1}{2}$, and in the retrieval phase, the user can access $N = 0, \ldots, 30$ databases in addition to the data center. We plot the expected normalized download cost versus different number of available databases in Fig. 6.2. When $N = 0$, in order to download the desired file privately, the user should download all the files in the data center, and this results in a download cost of $\frac{D}{L} = K = 10$. As the number of accessible databases increases, the normalized download cost decreases. We next consider the scenario where the data center contains $K = 10$ files, and the user can access $N = 5$ databases in addition to the data center in the retrieval phase. We plot the expected normalized download

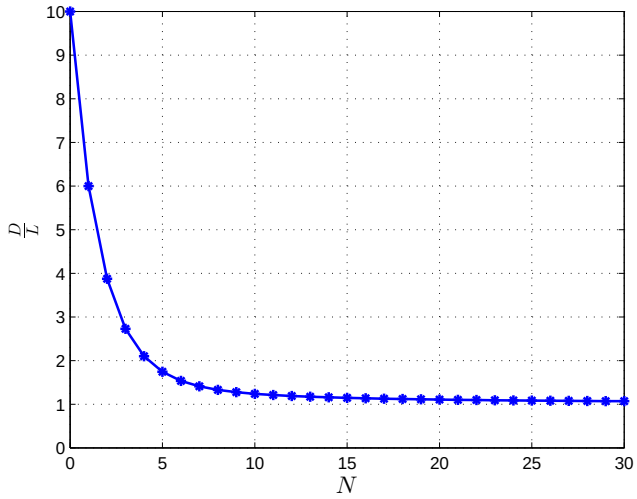

Figure 6.2: PIR from different number of available databases in the retrieval phase with $K = 10$ and $\mu = \frac{1}{2}$.

cost versus different storage size constraint μ in Fig. 6.3. When $\mu = 0$, in order to download the desired file privately, the user should download all the files in the data center resulting in $\frac{D}{L} = K = 10$. As μ increases, the normalized download cost decreases. Finally, we conclude this section with the following general remarks about our main result.

6.3.3 Remarks

Remark 6.1 *The achievability scheme consists of two parts, the design of the probability distribution in the caching phase and the PIR scheme in the retrieval phase. We find that the uniform and random caching scheme, originally proposed in [54] for decentralized coded caching, results in the optimal normalized download cost in the retrieval phase. We remark here that the symmetric batch caching scheme, originally proposed in [53] for centralized coded caching, also results in the optimal normalized*

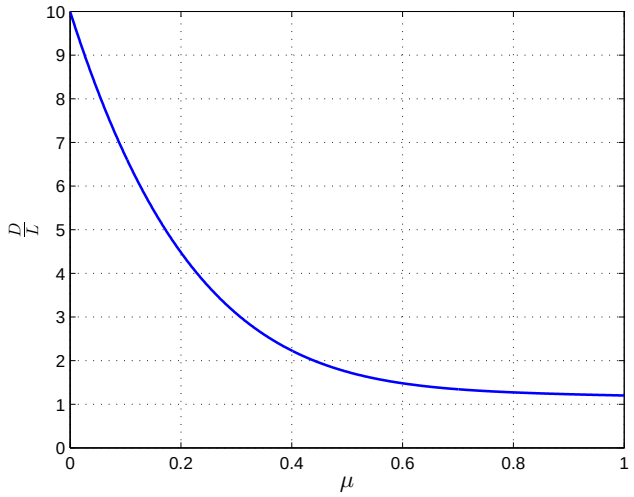

Figure 6.3: PIR from $N = 5$ databases with different storage constraint μ with $K = 10$.

download cost for PIR from centralized uncoded caching databases [43]. In the retrieval phase, according to the distribution of the subfiles, we apply the PIR scheme proposed in [12] for all subfiles to retrieve the desired file.

Remark 6.2 *For the converse, we first apply the lower bound derived in [43] which introduces new ingredients in addition to the interference lower bound lemma and induction lemma in [12, Lemma 5 and Lemma 6]. We note that in [43] the authors replace random variables for queries and answering strings by the contents of the distributed databases in a novel way which is crucial for the converse. With this replacement, we can account for different cached content in the caching phase resulting in different lower bound in the normalized download cost in the retrieval phase. Due to the nature of uncoded caching, this replacement facilitates further lower bound. For the decentralized problem here, to compare different probability distributions in the caching phase, we focus on the marginal distributions on each bit. This transfor-*

mation allows us to use linearity of expectation, and the nature of decentralization and uncoded caching to further lower bound the expected normalized download cost.

Remark 6.3 *A more directly related PIR problem from centralized uncoded caching databases for our setting is the one where, in the caching phase, the data center arranges the files in N databases in a centralized manner, and in the retrieval phase, the user has access also to the data center in addition to the N databases. This is different from the problem setting in [38, 43], since there the user can only access the N databases in the retrieval phase. As a side note, we can show that symmetric batch caching scheme is still optimal for this extended problem setting where the data center also participates in the PIR stage. Rigorously, the optimal trade-off between storage and download cost in this case is given by the lower convex envelope of the following $(\mu, D(\mu))$ pairs, for $t = 0, 1, \ldots, N$,*

$$\left(\mu = \frac{t}{N}, D(\mu) = \sum_{k=0}^{K-1} \frac{1}{(t+1)^k}\right). \tag{6.36}$$

To achieve this trade-off, the data center arranges the files into the N databases as in [38, 43]. In the retrieval phase, the user accesses also the data center; therefore, the subfiles are stored in one more database. For the converse, we no longer require all the N databases to reconstruct the entire K files as in [38, 43]. Thus, while in [38,43] the smallest allowable μ is $\mu = \frac{1}{N}$, since the N databases need to reconstruct the entire K files, here since the user can access the data center, the parameter μ starts from 0. Now, we can compare PIR from centralized caching databases and PIR from decentralized caching databases fairly, since in the retrieval phase, the

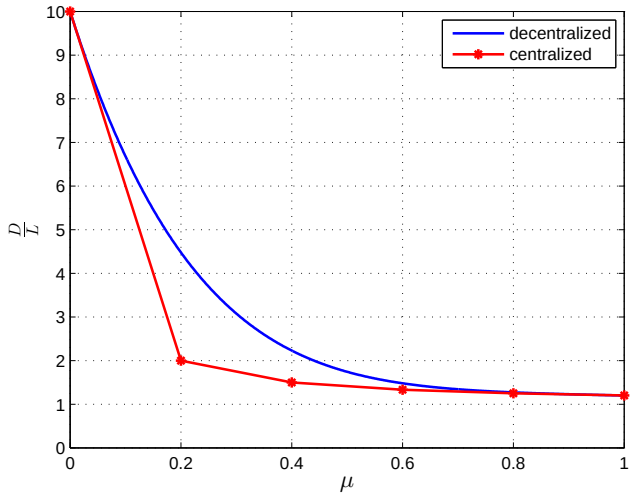

Figure 6.4: PIR from centralized caching databases and decentralized caching databases.

user can access the data center in both cases. We consider the case where $K = 10$ *and* $N = 5$*, and plot the result in Fig. 6.4.*

6.4 Achievability Scheme

The achievability scheme consists of two parts: the design of the probability distribution used in the caching phase and the PIR scheme used in the retrieval phase. In the caching phase, each database uniformly and randomly stores μKL bits from the data center. The storage size constraint in (6.2) is satisfied directly. Each database operates independently through the same probability distribution resulting in decentralized caching.

In the retrieval phase, suppose there are N databases in addition to the data

center available to the user. Each file W_j can be expressed as

$$W_j = \bigcup_{\{0\}\subseteq S\subseteq\{0,1,\ldots,N\}} W_{j,S}, \tag{6.37}$$

where $W_{j,S}$ represents the bits of file W_j which are stored in databases in S. Since each bit must be stored in the data center, i.e., DB_0, we have $\{0\} \subseteq S$. By the law of large numbers,

$$|W_{j,S}| = L\mu^{|S|-1}(1-\mu)^{N+1-|S|} + o(L), \tag{6.38}$$

when the file size is large enough.

To retrieve the desired file, say W_j, privately, we retrieve each subfile, $W_{j,S}$, privately. Subfile $W_{j,S}$ is replicated in $|S|$ databases, and for each of these $|S|$ databases, there are K subfiles, i.e., $W_{k,S}$, $k = 1, \ldots, K$. We apply the PIR scheme in [12] to retrieve $W_{j,S}$ privately by downloading

$$L\mu^{|S|-1}(1-\mu)^{N+1-|S|}\left(1 + \frac{1}{|S|} + \cdots + \frac{1}{|S|^{K-1}}\right) + o(L) \tag{6.39}$$

bits. We also note that there are $\binom{N}{|S|-1}$ types of $W_{j,S}$. Therefore, the following normalized download cost

$$\frac{D}{L} = \sum_{n=1}^{N+1} \binom{N}{n-1} \mu^{n-1}(1-\mu)^{N+1-n}\left(1 + \frac{1}{n} + \cdots + \frac{1}{n^{K-1}}\right) \tag{6.40}$$

is achievable.

6.5 Converse Proof

We first derive a lower bound for $D_{\mathcal{H}}$. Since in the retrieval phase the content of DB_0, $\ldots$, DB_N, are fixed to be Z_0, $\ldots$, Z_N, we can use the lower bound derived in [43, Eqn. (71)] to serve as the lower bound for $D_{\mathcal{H}}$. A key step to obtain [43, Eqn.(71)] is to replace the query and answering string random variables with the content of each database, i.e., replacement of $Q_{\mathcal{N}}^{[k]}$ and $A_{\mathcal{N}}^{[k]}$ with $Z_{\mathcal{N}}$. With this replacement, one can account for different cached content in the caching phase resulting in different lower bound in the normalized download cost in the retrieval phase. In addition, due to the nature of uncoded caching, this replacement facilitates a further lower bound. Moreover, to obtain [43, Eqn. (71)], the authors find interesting recursive relationships to compactly deal with the nested harmonic sums. Therefore, from [43, Eqn.(71)] we have

$$D_{\mathcal{H}} \geq L + \sum_{l=1}^{N+1} \binom{N+1}{l} \left(\frac{1}{l} + \frac{1}{l^2} + \cdots + \frac{1}{l^{K-1}} \right) x_l, \tag{6.41}$$

where

$$x_l \triangleq \frac{1}{K\binom{N+1}{l}} \sum_{\{0\} \subseteq S \subseteq [0:N],\ |S|=l} H(W_{1:K,S}), \quad l \in [1 : N+1], \tag{6.42}$$

and $W_{1:K,S}$ represents the bits of files $W_{1:K}$ which are stored in databases in S.

In the following lemma, we develop a lower bound for $E[x_l]$.

Lemma 6.1 *For $l \in [1 : N+1]$, and x_l given in* (6.42)*, we have*

$$E[x_l] \geq L\mu^{l-1}(1-\mu)^{N+1-l}\frac{\binom{N}{l-1}}{\binom{N+1}{l}}. \tag{6.43}$$

Proof: By taking expectation on (6.42) and using the linearity of expectation, we have

$$E[x_l] = \frac{1}{K\binom{N+1}{l}} \sum_{\{0\}\subseteq S\subseteq[0:N],\ |S|=l} E[H(W_{1:K,S})]. \tag{6.44}$$

Let random variables $X_{i,j}^{(n)}$, $i = 1, \ldots, L$, $j = 1, \ldots, K$, be the indicator functions showing that the ith bit of file W_j is cached in DB_n, $n = 0, \ldots, N$, or not, i.e., $X_{i,j}^{(n)} = 1$ means that the ith bit of file W_j is stored in DB_n and $X_{i,j}^{(n)} = 0$ means that it is not stored in DB_n. For DB_n we have

$$X_{1,1}^{(n)} + \cdots + X_{L,1}^{(n)} + \cdots + X_{1,K}^{(n)} + \cdots + X_{L,K}^{(n)} \leq \mu KL \tag{6.45}$$

due to the storage size constraint in (6.2). We note that P_H induces probability measures on random variables $X_{i,j}^{(n)}$, and let $X_{i,j}^{(n)} = 1$ with probability $p_{i,j}$, where we remove the superscript n since each database adopts the same probability distribution P_H to choose the cached bits due to the decentralized caching property. By taking expectation on (6.45) and applying the linearity of expectation, we have

$$E[X_{1,1}^{(n)}] + \cdots + E[X_{L,K}^{(n)}] \leq \mu KL, \tag{6.46}$$

which yields

$$p_{1,1} + \cdots + p_{L,K} \leq \mu KL. \tag{6.47}$$

Let random variables $Y_{i,j}^S$, $i = 1, \ldots, L$, $j = 1, \ldots, K$, be the indicator functions showing that the ith bit of file W_j is cached in DB_n, $n \in S$, i.e., $Y_{i,j} = 1$ means that the ith bit of the file W_j is stored in DB_n, $n \in S$. Therefore, we have

$$Y_{i,j}^S = \prod_{n \in S} X_{i,j}^{(n)} \prod_{n \in [0:N] \setminus S} (1 - X_{i,j}^{(n)}). \tag{6.48}$$

Now, we can evaluate $E\left[H(W_{1:K,S})\right]$ in (6.44) as follows

$$E\left[H(W_{1:K,S})\right] = E[Y_{1,1}^S + \cdots + Y_{L,K}^S] \tag{6.49}$$

$$= E[Y_{1,1}^S] + \cdots + E[Y_{L,K}^S] \tag{6.50}$$

$$= p_{1,1}^{|S|-1}(1 - p_{1,1})^{N+1-|S|} + \cdots + p_{L,K}^{|S|-1}(1 - p_{L,K})^{N+1-|S|}, \tag{6.51}$$

where $p_{1,1}, \ldots, p_{L,K}$ are subject to (6.47). Now, continuing from (6.44), we have

$$E[x_l] = \frac{1}{K\binom{N+1}{l}} \sum_{\{0\} \subseteq S \subseteq [0:N],\ |S|=l} p_{1,1}^{l-1}(1 - p_{1,1})^{N+1-l} + \cdots + p_{L,K}^{l-1}(1 - p_{L,K})^{N+1-l}. \tag{6.52}$$

To further lower bound (6.52), we consider the following Lagrangian

$$L(p_{1,1}, \ldots, p_{L,K}, \lambda) = p_{1,1}^{l-1}(1 - p_{1,1})^{N+1-l} + \cdots + p_{L,K}^{l-1}(1 - p_{L,K})^{N+1-l}$$

$$+ \lambda \left(p_{1,1} + \cdots + p_{L,K} - \mu KL \right). \tag{6.53}$$

From the KKT conditions, we have

$$\lambda = p_{i,j}^{l-1}(N+1-l)(1-p_{i,j})^{N-l} - (l-1)p_{i,j}^{l-2}(1-p_{i,j})^{N+1-l}, \tag{6.54}$$

where $i = 1, \ldots, L$, $j = 1, \ldots, K$. Therefore, we can further lower bound (3.34) by letting $p_{1,1} = \cdots = p_{L,K} = \mu$, then we have

$$E[x_l] \geq \frac{1}{K\binom{N+1}{l}} \sum_{\{0\} \subseteq S \subseteq [0:N],\ |S|=l} KL\mu^{l-1}(1-\mu)^{N+1-l} \tag{6.55}$$

$$= L\mu^{l-1}(1-\mu)^{N+1-l} \frac{\binom{N}{l-1}}{\binom{N+1}{l}}, \tag{6.56}$$

which completes the proof. ■

Finally, by taking expectation and applying Lemma 6.1 to (6.41), we obtain

$$\frac{D}{L} \geq 1 + \sum_{l=1}^{N+1} \binom{N}{l-1} \left(\frac{1}{l} + \frac{1}{l^2} + \cdots + \frac{1}{l^{K-1}} \right) \mu^{l-1}(1-\mu)^{N+1-l} \tag{6.57}$$

$$= (\mu + (1-\mu))^N + \sum_{l=1}^{N+1} \binom{N}{l-1} \left(\frac{1}{l} + \frac{1}{l^2} + \cdots + \frac{1}{l^{K-1}} \right) \mu^{l-1}(1-\mu)^{N+1-l} \tag{6.58}$$

$$= \sum_{l=1}^{N+1} \binom{N}{l-1} \left(1 + \frac{1}{l} + \frac{1}{l^2} + \cdots + \frac{1}{l^{K-1}} \right) \mu^{l-1}(1-\mu)^{N+1-l} \tag{6.59}$$

which matches (6.40).

6.6 Conclusion

We considered the PIR problem from decentralized uncoded caching databases. Due to the nature of decentralization and the storage size constraint, we allowed the user to access the data center in the retrieval phase to guarantee that the user can reconstruct the entire desired file. We showed that uniform and random decentralized caching scheme, originally proposed in [54] for the problem of decentralized coded caching, results in the lowest expected normalized download cost in the PIR phase. We characterized the expected normalized download cost to be $\frac{D}{L} = \sum_{n=1}^{N+1} \binom{N}{n-1} \mu^{n-1}(1-\mu)^{N+1-n} \left(1 + \frac{1}{n} + \cdots + \frac{1}{n^{K-1}}\right)$. For the achievability, we applied the PIR scheme in [12] for all subfiles. For the converse, we first applied the lower bound derived in [43], and to compare different probability distributions in the caching phase, we focused on the marginal distributions on individual bits. By using the nature of decentralization and uncoded caching, we further lower bounded the normalized download cost. Finally, we showed the matching converse for the expected normalized download cost, obtaining the exact capacity of the resulting PIR problem.

CHAPTER 7

Polar Coding for the General Wiretap Channel with Extensions to Multiuser Scenarios

7.1 Introduction

Information-theoretic work for wiretap channels is mostly based on random coding schemes. Designing practical coding schemes to achieve information-theoretic secrecy is an important problem. By applying two recently developed techniques for polar codes, namely, universal polar coding and polar coding for asymmetric channels, we propose a polar coding scheme to achieve the secrecy capacity of the general wiretap channel. We then apply this coding scheme to achieve the best-known inner bounds for the multiple access wiretap channel (MAC-WTC), and the broadcast and interference channels with confidential messages (BC-CM and IC-CM).

7.2 System Model

7.2.1 Wiretap Channel Model

A wiretap channel consists of a legitimate transmitter who wishes to send messages to a legitimate receiver secretly in the presence of an eavesdropper. Let X denote the single-letter input to the main and eavesdropper channels. Let Y and Z denote the corresponding single-letter outputs of the main and the eavesdropper channels, respectively. W represents the message to be sent to Bob and kept secret from Eve with $W \in \mathcal{W} = \{1, \cdots, 2^{nR}\}$. Let $P_e = \Pr(\hat{\mathrm{W}} \neq \mathrm{W})$ denote the probability of error for Bob's decoding.

The equivocation rate is given by

$$\frac{1}{n} H(W|Z^n), \tag{7.1}$$

which reflects the uncertainty of the message given the eavesdropper's channel observation. A rate pair (R, R_e) is achievable if for any $\epsilon > 0$, as $n \to \infty$,

$$\Pr(\hat{\mathrm{W}} \neq \mathrm{W}) \leq \epsilon, \qquad \frac{1}{n} H(W|Z^n) \geq R_e - \epsilon. \tag{7.2}$$

Perfect (weak) secrecy is achieved if $R = R_e$ [72]. Therefore, perfect secrecy is achieved if $\frac{1}{n} I(W; Z^n) \to 0$, and the *secrecy capacity* C_s is the highest achievable perfect secrecy rate R, which is also the highest possible equivocation rate [72]. Csiszár and Körner characterized the secrecy capacity for the general wiretap chan-

nel as [72]

$$C_s = \max_{V \to X \to Y,Z} I(V;Y) - I(V;Z). \tag{7.3}$$

7.2.2 Multiple Access Wiretap Channel

A MAC-WTC consists of two transmitters, one receiver and an eavesdropper. For $k \in 1, 2$, the two transmitters, with channel inputs X_k, wish to send independent messages $W_k \in \mathcal{W}_k = \{1, \cdots, 2^{nR_k}\}$ to the legitimate receiver, with channel output Y, in the presence of an eavesdropper, with channel output Z. A rate pair (R_1, R_2) is achievable if for any $\epsilon > 0$, as $n \to \infty$,

$$\Pr(\hat{\mathrm{W}}_\mathrm{k} \neq \mathrm{W}_\mathrm{k}) \leq \epsilon, \qquad \frac{1}{n} H(W_1, W_2 | Z^n) \geq R_1 + R_2 - \epsilon. \tag{7.4}$$

The secrecy capacity region of the MAC-WTC is still an open problem. The best-known achievable rate region is [95, 96] (see also [101–103]):

$$\begin{aligned} R_1 &\leq [I(V_1;Y|V_2,T) - I(V_1;Z|T)]^+, \\ R_2 &\leq [I(V_2;Y|V_1,T) - I(V_2;Z|T)]^+, \\ R_1 + R_2 &\leq [I(V_1,V_2;Y|T) - I(V_1,V_2;Z|T)]^+, \end{aligned} \tag{7.5}$$

for any distribution of the form

$$P(t)P(v_1|t)P(v_2|t)P(x_1|v_1)P(x_2|v_2)P(y,z|x_1,x_2). \tag{7.6}$$

7.2.3 Broadcast Channel With Confidential Messages

A BC-CM consists of a transmitter and two receivers. For $k \in 1, 2$, the transmitter wishes to send independent messages, $W_k \in \mathcal{W}_k = \{1, \cdots, 2^{nR_k}\}$, to their respective receiver k, while keeping the messages secret from the unintended receiver. Let X, Y_1, Y_2 denote the single-letter input and outputs of the broadcast channel. A rate pair (R_1, R_2) is achievable if for any $\epsilon > 0$, as $n \to \infty$,

$$\Pr(\hat{\mathrm{W}}_\mathrm{k} \neq \mathrm{W}_\mathrm{k}) \leq \epsilon, \qquad \frac{1}{n} H(W_1|Y_2^n) \geq R_1 - \epsilon, \qquad \frac{1}{n} H(W_2|Y_1^n) \geq R_2 - \epsilon. \quad (7.7)$$

The secrecy capacity region of the BC-CM is still an open problem. The best-known achievable rate region [97] is:

$$\begin{aligned} R_1 &\leq I(V_1; Y_1|T) - I(V_1; V_2|T) - I(V_1; Y_2|V_2, T), \\ R_2 &\leq I(V_2; Y_2|T) - I(V_1; V_2|T) - I(V_2; Y_1|V_1, T), \end{aligned} \quad (7.8)$$

over all distributions of the form

$$P(t)P(v_1, v_2|t)P(x|v_1, v_2)P(y_1, y_2|x). \quad (7.9)$$

7.2.4 Interference Channel With Confidential Messages

An IC-CM consists of two transmitters and two receivers. The two transmitters wish to send independent messages to their respective receivers, and keep the messages

confidential from the other receiver. For $k \in 1, 2$, let X_k, Y_k denote the single-letter input and output of the interference channel with messages $W_k \in \mathcal{W}_k = \{1, \cdots, 2^{nR_k}\}$. A rate pair (R_1, R_2) is achievable if for any $\epsilon > 0$, as $n \to \infty$,

$$\Pr(\hat{\mathrm{W}}_\mathrm{k} \neq \mathrm{W}_\mathrm{k}) \leq \epsilon, \qquad \frac{1}{n} H(W_1|Y_2^n) \geq R_1 - \epsilon, \qquad \frac{1}{n} H(W_2|Y_1^n) \geq R_2 - \epsilon. \quad (7.10)$$

The secrecy capacity region of the IC-CM is still an open problem. The best-known achievable rate region [97] is:

$$\begin{aligned} R_1 &\leq I(V_1; Y_1|T) - I(V_1; Y_2|V_2, T), \\ R_2 &\leq I(V_2; Y_2|T) - I(V_2; Y_1|V_1, T), \end{aligned} \quad (7.11)$$

over all distribution of the form

$$P(t)P(v_1|t)P(v_2|t)P(x_1|v_1)P(x_2|v_2)P(y_1, y_2|x_1, x_2). \quad (7.12)$$

7.3 Existing Polar Coding Techniques

7.3.1 Polar Codes for Asymmetric Channels

Let P_{XY} be the joint distribution of a pair of random variables (X, Y), where X is a binary random variable and Y is any finite-alphabet random variable. Let us define

the Bhattacharyya parameter as follows:

$$Z(X|Y) = 2\sum_{y} P_Y(y)\sqrt{P_{X|Y}(0|y)P_{X|Y}(1|y)}. \tag{7.13}$$

Let $U^n = X^n G_n$, where X^n denotes n independent copies of the random variable X with $X \sim P_X$, and $G_n = G^{\otimes k}$ where $G = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $\otimes$ denotes the Kronecker product of matrices for $n = 2^k$. Reference [74] shows that as $n \to \infty$, U_i is almost independent of U^{i-1} and uniformly distributed, or otherwise U_i is almost determined by U^{i-1}. Therefore, $[n]$, the index set $\{1, 2, \ldots, n\}$, is almost polarized into two sets $\mathcal{H}_X$ and $\mathcal{L}_X$ [80]:

$$\mathcal{H}_X = \{i \in [n] : Z(U_i|U^{i-1}) \geq 1 - \delta_n\}, \tag{7.14}$$

$$\mathcal{L}_X = \{i \in [n] : Z(U_i|U^{i-1}) \leq \delta_n\}, \tag{7.15}$$

where $\delta_n = 2^{-n^\beta}$ and $\beta \in (0, 1/2)$. Moreover,

$$\lim_{n\to\infty} \frac{1}{n}|\mathcal{H}_X| = H(X), \tag{7.16}$$

$$\lim_{n\to\infty} \frac{1}{n}|\mathcal{L}_X| = 1 - H(X). \tag{7.17}$$

Let P be a discrete memoryless channel with a binary input X and finite alphabet output Y. Here, P does not have to be a symmetric channel. Fix a distribution P_X for X. Reference [93] generalizes the above argument to achieve a

rate close to $I(X;Y)$. Consider two subsets of $[n]$, $\mathcal{H}_{X|Y}$ and $\mathcal{L}_{X|Y}$,

$$\mathcal{H}_{X|Y} = \{i \in [n] : Z(U_i|U^{i-1}, Y^n) \geq 1 - \delta_n\}, \tag{7.18}$$

$$\mathcal{L}_{X|Y} = \{i \in [n] : Z(U_i|U^{i-1}, Y^n) \leq \delta_n\}. \tag{7.19}$$

Similar to (7.16) and (7.17), we have

$$\lim_{n\to\infty} \frac{1}{n}|\mathcal{H}_{X|Y}| = H(X|Y), \tag{7.20}$$

$$\lim_{n\to\infty} \frac{1}{n}|\mathcal{L}_{X|Y}| = 1 - H(X|Y). \tag{7.21}$$

With (7.14) and (7.19), we define the following three sets

$$\mathcal{I} = \mathcal{H}_X \cap \mathcal{L}_{X|Y}, \tag{7.22}$$

$$\mathcal{F}_r = \mathcal{H}_X \cap \mathcal{L}^{\mathrm{c}}_{X|Y}, \tag{7.23}$$

$$\mathcal{F}_d = \mathcal{H}^{\mathrm{c}}_X. \tag{7.24}$$

In the following, we call the set $\mathcal{I}$ the *information set*, and sets $\mathcal{F}_r$ and $\mathcal{F}_d$ the *frozen set*. Although we call them the *frozen set*, $\mathcal{F}_r$ and $\mathcal{F}_d$ have different operational meanings which will be illustrated below. Note that for the symmetric channel capacity achieving code design, $\mathcal{F}_d$ is an empty set [73].

To achieve rate $I(X;Y)$ for channel P, let us consider the following coding scheme. First, the encoder transmits the information bits in the index set $\mathcal{I}$. For $i \in \mathcal{I}$ in (7.22), since $i \in \mathcal{H}_X$, U_i is almost independent of U^{i-1} and uniformly

distributed. Therefore, the encoder can freely assign values to $U_{\mathcal{I}}$, where $U_{\mathcal{I}}$ denotes a sub-vector $\{U_i\}_{i\in\mathcal{I}}$. Moreover, since $i \in \mathcal{L}_{X|Y}$, U_i is almost determined by U^{i-1} and Y^n, which means that given the channel output Y^n, U_i can be decoded in a successive manner.

Second, for $i \in \mathcal{F}_r$ in (7.23), U_i is almost independent of U^{i-1} and uniformly distributed, and given the channel output Y^n, U_i cannot be reliably decoded. The encoder transmits $U_{\mathcal{F}_r}$ with a uniformly random sequence and the randomness is shared between the transmitter and receiver.

Last, for $i \in \mathcal{F}_d$ in (7.24), U_i is almost determined by U^{i-1}. The values of $U_{\mathcal{F}_d}$ are computed in successive order through the following mapping:

$$u_i = \arg \max_{u\in\{0,1\}} P_{U_i|U^{i-1}}(u|u^{i-1}). \tag{7.25}$$

By (7.16) and (7.20), it is easy to verify that

$$\lim_{n\to\infty} \frac{1}{n}|\mathcal{I}| = I(X;Y). \tag{7.26}$$

Moreover, by applying successive cancellation decoder, the block error probability P_e can be upper bounded by [104]

$$P_e \leq \sum_{i\in\mathcal{I}} Z(U_i|U^{i-1}, Y^n) = O(2^{-n^\beta}) \tag{7.27}$$

for any $\beta \in (0, 1/2)$, with complexity $O(n \log n)$. Therefore, the rate $I(X;Y)$ is achieved.

7.3.2 Universal Polar Coding

Consider two B-DMCs $P : X \to Y$ and $Q : X \to Z$, and assume that these two channels have identical capacities, i.e., $C(P) = C(Q)$. Let $U^n = X^n G_n$, and denote $\mathcal{P}$ and $\mathcal{Q}$ as the information set defined in (7.22), i.e.,

$$\mathcal{P} = \{i \in [n] : Z(U_i|U^{i-1}, Y^n) \leq \delta_n\}, \tag{7.28}$$

$$\mathcal{Q} = \{i \in [n] : Z(U_i|U^{i-1}, Z^n) \leq \delta_n\}, \tag{7.29}$$

where $\delta_n = 2^{-n^\beta}$ and $\beta \in (0, 1/2)$. Since we assume $C(P) = C(Q)$, we also have $|\mathcal{P}| = |\mathcal{Q}|$.

In general, the differences $\mathcal{P} \setminus \mathcal{Q}$ and $\mathcal{Q} \setminus \mathcal{P}$ are not empty sets [88–90]; therefore, it is not straightforward to apply standard polar coding to achieve the capacity of the compound channel consisting of P and Q. Reference [91] proposes a method, called *chaining construction*, to solve this problem.

Definition 7.1 *(Chaining construction [91]) Let $m \geq 2$. The m-chain of $\mathcal{P}$ and $\mathcal{Q}$ is a code of length mn that consists of m polar blocks of length n. In each of the m blocks, the set $\mathcal{P} \cap \mathcal{Q}$ is set to be an information set. In the ith block, $1 \leq i < m$, the set $\mathcal{P} \setminus \mathcal{Q}$ is also set to be an information set. Moreover, the set $\mathcal{P} \setminus \mathcal{Q}$ in the ith block is chained to the set $\mathcal{Q} \setminus \mathcal{P}$ in the $(i+1)$th block in the sense that the information is repeated in these two sets. All other indices are frozen. Therefore, in each block, the set $(\mathcal{P} \cup \mathcal{Q})^c$ is frozen, and the set $\mathcal{Q} \setminus \mathcal{P}$ in the 1st block and the set $\mathcal{P} \setminus \mathcal{Q}$ in*

the mth block are frozen, too. The rate of the chaining construction is

$$\frac{|\mathcal{P} \cap \mathcal{Q}| + \frac{m-1}{m}|\mathcal{P} \setminus \mathcal{Q}|}{n}. \tag{7.30}$$

Next, we discuss the decoding procedure for the compound channel consisting of P and Q. If channel P is used, then we decode from the first block. On the other hand, if channel Q is used, then we decode from the last block.

First, suppose that channel P is used and a code of length mn has been received. For this case, we decode from the first block. In the 1st block, all the information bits are put in the set $\mathcal{P}$; thus, the decoder can decode correctly. For the 2nd block, through chaining construction, the set $\mathcal{P} \setminus \mathcal{Q}$ in the 1st block is chained to the set $\mathcal{Q} \setminus \mathcal{P}$ in the 2nd block, and the set $(\mathcal{P} \cup \mathcal{Q})^c$ is frozen. Equivalently, the decoder only needs to decode the bits in the set $\mathcal{P}$, which can be correctly decoded. The same procedure holds until the $(m-1)$th block. For the mth block, the information bits are only put in the set $\mathcal{P} \cap \mathcal{Q}$, and the remaining part has been determined. Hence, information bits can be reliably decoded.

Second, consider the case that channel Q is used. In this case, we decode from the last block. In the mth block, since the information bits are put in the set $\mathcal{Q}$, reliable decoding is guaranteed. For the $(m-1)$th block, due to the chaining process, the set $\mathcal{Q} \setminus \mathcal{P}$ in the mth block is chained to the set $\mathcal{P} \setminus \mathcal{Q}$ in the $(m-1)$th block, and note that the set $(\mathcal{P} \cup \mathcal{Q})^c$ is frozen. The decoder only needs to decode the information bits in the set $\mathcal{Q}$, thus correct decoding is ensured. This procedure is applied until the 2nd block. For the 1st block, information bits which have not

been determined fall in the set $\mathcal{P} \cap \mathcal{Q}$, thus the decoder can decode them correctly.

In summary, for a fixed m, if we let $n \to \infty$, we can achieve the rate in (7.30) with arbitrary small error probability, which also means that the rate $C(P) - \frac{1}{m}\frac{|\mathcal{P} \setminus \mathcal{Q}|}{n}$ can be achieved. Additionally, if we let $m \to \infty$, then the rate $C(P)$, which is the capacity of the compound channel consisting of channels P and Q, can be achieved.

7.3.3 Polar Coding for MAC Based on Monotone Chain Rules

Consider a two-user MAC $(\mathcal{X}_1 \times \mathcal{X}_2, P(y|x_1, x_2), \mathcal{Y})$ with binary input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$. The capacity region of this channel is the union of convex hull of all rate pairs satisfying

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2), \\ R_2 &\leq I(X_2; Y|X_1), \\ R_1 + R_2 &\leq I(X_1, X_2; Y), \end{aligned} \tag{7.31}$$

over the distributions of the form $P(x_1)P(x_2)$. The rate pairs satisfying $R_1 + R_2 = I(X_1, X_2; Y)$ are said to be on the *dominant face* of the rate region.

Reference [82] gives a polar coding scheme that achieves the entire dominant face based on the monotone chain rules. Consider $U_1^n = X_1^n G_n$ and $U_2^n = X_2^n G_n$. We call J^{2n} as a monotone permutation of $U_1^n U_2^n$ if the elements of both U_1^n and U_2^n appear in increasing order in J^{2n}. When we expand the mutual information term $I(U_1^n, U_2^n; Y^n)$ according to the monotone permutation, we say that it follows the

monotone chain rule

$$I(U_1^n, U_2^n; Y^n) = \sum_{i=1}^{2n} I(J_i; Y^n | J^{i-1}). \tag{7.32}$$

Moreover, define the rates as follows

$$\begin{aligned} R_x &= \frac{1}{n} \sum_{\{i \in [2n]: J_i \in U_1^n\}} I(J_i; Y^n | J^{i-1}), \\ R_y &= \frac{1}{n} \sum_{\{i \in [2n]: J_i \in U_2^n\}} I(J_i; Y^n | J^{i-1}). \end{aligned} \tag{7.33}$$

Reference [82] shows that the rate pair (R_x, R_y) in (7.33) can be set arbitrarily close to the rate pairs on the dominant face of (7.31) by the permutations of the form $J^{2n} = (U_1^i, U_2^n, U_1^{i+1:n})$, where $U_1^{i+1:n}$ denotes $U_{1,i+1}, \ldots, U_{1,n}$.

7.4 Polar Coding for the General Wiretap Channel

Assume now that we know the optimal distributions [105] to achieve the secrecy capacity C_s in (7.3), i.e., we know the optimal V and X. For illustration, we consider the case of a binary input channel, i.e., $|\mathcal{X}| = 2$. The cardinality bound for channel prefixing, V, is $|\mathcal{V}| \leq 2$. Although we focus on developing a coding scheme for binary inputs below, there is no difficulty to extend the work to q-ary inputs [106–109].

7.4.1 The Scheme

Let $U^n = V^n G_n$. Consider the following sets:

$$\begin{aligned}
\mathcal{H}_V &= \{i \in [n] : Z(U_i|U^{i-1}) \geq 1 - \delta_n\}, \\
\mathcal{L}_{V|Y} &= \{i \in [n] : Z(U_i|U^{i-1}, Y^n) \leq \delta_n\}, \\
\mathcal{L}_{V|Z} &= \{i \in [n] : Z(U_i|U^{i-1}, Z^n) \leq \delta_n\},
\end{aligned} \tag{7.34}$$

where $\delta_n = 2^{-n^\beta}$ and $\beta \in (0, 1/2)$.

The set $[n]$ can be partitioned into the following four sets:

$$\begin{aligned}
G_{Y \wedge Z} &= \mathcal{H}_V \cap \mathcal{L}_{V|Y} \cap \mathcal{L}_{V|Z}, \\
G_{Y \setminus Z} &= \mathcal{H}_V \cap \mathcal{L}_{V|Y} \cap \mathcal{L}^c_{V|Z}, \\
G_{Z \setminus Y} &= \mathcal{H}_V \cap \mathcal{L}^c_{V|Y} \cap \mathcal{L}_{V|Z}, \\
B_{Y \wedge Z} &= \mathcal{H}^c_V \cup (\mathcal{L}^c_{V|Y} \cap \mathcal{L}^c_{V|Z}).
\end{aligned} \tag{7.35}$$

From a successive decoding point of view, the sub-channels corresponding to the set $G_{Y \wedge Z}$ are simultaneously good for Bob and Eve. The sub-channels in the set $G_{Y \setminus Z}$ are good for Bob but bad for Eve. On the other hand, the sub-channels in the set $G_{Z \setminus Y}$ are good for Eve but bad for Bob. Last, the sub-channels in the set $B_{Y \wedge Z}$ are bad for both Bob and Eve.

Similar to (7.22)–(7.24), we have:

$$\begin{aligned}
\mathcal{I}_Y &= \mathcal{H}_V \cap \mathcal{L}_{V|Y}, \\
\mathcal{I}_Z &= \mathcal{H}_V \cap \mathcal{L}_{V|Z}, \\
\mathcal{F}_r^Y &= \mathcal{H}_V \cap \mathcal{L}_{V|Y}^c, \\
\mathcal{F}_r^Z &= \mathcal{H}_V \cap \mathcal{L}_{V|Z}^c, \\
\mathcal{F}_d &= \mathcal{H}_V^c.
\end{aligned} \tag{7.36}$$

By (7.26), we have

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}|\mathcal{I}_Y| &= I(V;Y), \\
\lim_{n\to\infty} \frac{1}{n}|\mathcal{I}_Z| &= I(V;Z).
\end{aligned} \tag{7.37}$$

For the *symmetric* and *degraded* wiretap channel [84–87], $G_{Z\setminus Y}$ is an empty set, since the degraded property of the channel causes $\mathcal{I}_Z \subset \mathcal{I}_Y$ [83]. However, for the general wiretap channel, $G_{Z\setminus Y}$ is no longer an empty set, and $|G_{Z\setminus Y}|$ cannot be negligible [88–90].

Here, we consider the positive secrecy capacity case, thus, we have $|G_{Y\setminus Z}| > |G_{Z\setminus Y}|$. Choose a set, $C_{Y\setminus Z}$, such that $C_{Y\setminus Z} \subset G_{Y\setminus Z}$ and $|C_{Y\setminus Z}| = |G_{Z\setminus Y}|$. Define the set S as:

$$S = G_{Y\setminus Z} \setminus C_{Y\setminus Z}. \tag{7.38}$$

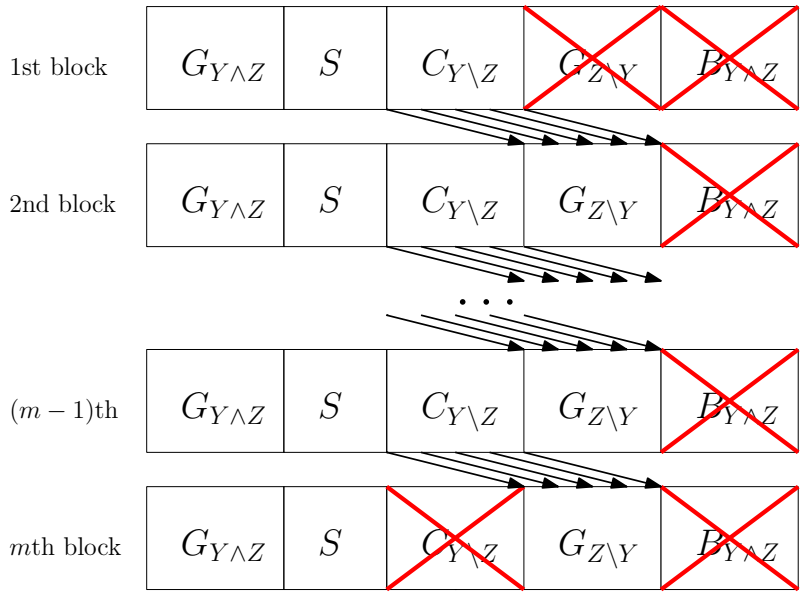

Figure 7.1: Chaining construction for the general wiretap channel.

From (7.37), we have

$$\lim_{n\to\infty}\frac{1}{n}|S| = I(V;Y) - I(V;Z). \tag{7.39}$$

We construct the code as follows. Consider an *m-chain* polar code in Definition 7.1. For $1 \leq i < m$, the set $C_{Y\setminus Z}$ in the ith block is chained to $G_{Z\setminus Y}$ in the $(i+1)$th block as in Fig. 7.1. For each of the m blocks, the set $B_{Y\wedge Z}$ is set to be frozen. Moreover, the set $G_{Z\setminus Y}$ in the 1st block is set to be frozen in the sense that $G_{Z\setminus Y} \subseteq \mathcal{F}_r^Y$, and the set $C_{Y\setminus Z}$ in the mth block is also set to be frozen in the sense that $C_{Y\setminus Z} \subseteq \mathcal{F}_r^Z$. In Fig. 7.1, we use a red cross to denote a frozen set.

We put the secret information bits in the set S in each block. Therefore, the set S is used for secret message transmission. For blocks $1 \leq i < m$, we put uniformly distributed random bits to $C_{Y\setminus Z}$ to serve as the confusion messages. Through the chaining construction, the confusion messages are also chained to the set $G_{Z\setminus Y}$ in

block $1 < i \leq m$. Moreover, the set $G_{Y \wedge Z}$ in each block are also filled with random bits to serve as confusion message. For the frozen sets, if the index belongs to $\mathcal{F}_r^Y$ or $\mathcal{F}_r^Z$, then we put uniformly distributed random bits and share the randomness with the decoder (Bob and Eve). Last, if the index belongs to $\mathcal{F}_d$, then we determine the value according to the mapping defined in (7.25). We summarize the encoding procedure as follows.

Encoding procedure:

For each block, put the secret information bits in U_S, and determine the bits in $U_{\mathcal{F}_d}$ by (7.25).

For the 1st block,

1. Put uniformly distributed random bits to $U_{G_{Y \wedge Z} \cup C_{Y \setminus Z}}$.
2. Put uniformly distributed random bits to $U_{\mathcal{F}_r^Y}$, and share the randomness with the decoder.

For the jth block, $2 \leq j < m$,

1. Put uniformly distributed random bits to $U_{G_{Y \wedge Z} \cup C_{Y \setminus Z}}$.
2. Chaining construction: repeat the bits in $C_{Y \setminus Z}$ of the $(j-1)$th block to the bits in $U_{G_{Z \setminus Y}}$.
3. Put uniformly distributed random bits to $U_{\mathcal{F}_r^Y \cap \mathcal{F}_r^Z}$, and share the randomness with the decoder.

For the mth block,

1. Put uniformly distributed random bits to $U_{G_{Y\wedge Z}}$.

2. Chaining construction: repeat the bits in $C_{Y\setminus Z}$ of the $(m-1)$th block to the bits in $U_{G_{Z\setminus Y}}$.

3. Put uniformly distributed random bits to $U_{\mathcal{F}_r^Z}$, and share the randomness with the decoder.

Note that in the chaining construction we require the bits in $U_{G_{Z\setminus Y}}$ equal the bits in $U_{C_{Y\setminus Z}}$. Since we fill uniformly distributed random bits to $U_{C_{Y\setminus Z}}$, we simultaneously fill random bits to $U_{G_{Z\setminus Y}}$. Due to the fact that $G_{Z\setminus Y}\cap\mathcal{F}_d=\emptyset$, we can freely choose the bits in this set.

Decoding procedure:

Bob decodes from the 1st block. If $i\in\mathcal{F}_d$, then $\hat{u}_i=\arg\max_{u\in\{0,1\}}P_{U_i|U^{i-1}}(u|\hat{u}^{i-1})$. For the 1st block,

$$\hat{u}_i=\begin{cases}u_i, & \text{if } i\in\mathcal{F}_r^Y,\\ \arg\max_{u\in\{0,1\}}P_{U_i|U^{i-1},Y^n}(u|\hat{u}^{i-1},y^n), & \text{if } i\in G_{Y\wedge Z}\cup C_{Y\setminus Z}\cup S.\end{cases}\tag{7.40}$$

For the jth block, $2\le j<m$,

$$\hat{u}_i=\begin{cases}u_i, & \text{if } i\in\mathcal{F}_r^Y\cap\mathcal{F}_r^Z,\\ \arg\max_{u\in\{0,1\}}P_{U_i|U^{i-1},Y^n}(u|\hat{u}^{i-1},y^n), & \text{if } i\in G_{Y\wedge Z}\cup C_{Y\setminus Z}\cup S,\\ \hat{u}_{i'} \text{ in the } (j-1)\text{th block, where } i'\in C_{Y\setminus Z}, & \text{if } i\in G_{Z\setminus Y}.\end{cases}\tag{7.41}$$

For the mth block,

$$\hat{u}_i = \begin{cases} u_i, & \text{if } i \in \mathcal{F}_r^Z, \\ \arg\max_{u\in\{0,1\}} P_{U_i|U^{i-1},Y^n}(u|\hat{u}^{i-1}, y^n), & \text{if } i \in G_{Y\wedge Z} \cup S, \\ \hat{u}_{i'} \text{ in the } (m-1)\text{th block, where } i' \in C_{Y\setminus Z}, & \text{if } i \in G_{Z\setminus Y}. \end{cases} \tag{7.42}$$

7.4.2 Reliability

From (7.39), we know as $n \to \infty$, our coding scheme can achieve the secrecy rate in (7.3). Moreover, when Bob applies the decoding procedure described in Section 7.4.1, according to (7.27), the block error probability of the whole *m-chain* block can be upper bounded by

$$P_e \leq (m-1) \sum_{i\in C_{Y\setminus Z}} Z(U_i|U^{i-1}, Y^n) + m \sum_{i\in G_{Y\wedge Z}\cup S} Z(U_i|U^{i-1}, Y^n) = O(2^{-n^\beta}) \tag{7.43}$$

for any $\beta \in (0, 1/2)$ with complexity $O(n \log n)$. Thus, the secrecy rate in (7.3) is achieved reliably.

7.4.3 Equivocation Calculation

We first introduce necessary notation for the calculation of the equivocation rate. In the encoding process, we consider m blocks each with block length n. Let Z^{mn} denote what Eve receives. For each block, we perform $U^n = V^n G_n$, therefore, for the total of m blocks, we have V^{mn} and U^{mn}.

Let W_s denote the secret message, and $\tilde{W}_s$ denote the confusion message. Let the subscript i of a set denote the set in the ith block. For example, S_i denotes the set S in the ith block, and $G_{Y\wedge Zj}$ denotes the set $G_{Y\wedge Z}$ in the jth block. Since secret message is put in S_i, $1 \leq i \leq m$, we have $W_s = \cup_{1\leq i\leq m} U_{S_i}$. Also, the confusion message is put in $G_{Y\wedge Zi},\ 1 \leq i \leq m$ and $C_{Y\setminus Zj},\ 1 \leq j < m$. Therefore, we have $\tilde{W}_s = \cup_{1\leq i\leq m, 1\leq j<m} U_{G_{Y\wedge Zi}} U_{C_{Y\setminus Zj}}$.

We can calculate the equivocation rate as follows:

$$H(W_s|Z^{mn}) = H(W_s, \tilde{W}_s|Z^{mn}) - H(\tilde{W}_s|W_s, Z^{mn}) \tag{7.44}$$

$$= H(W_s, \tilde{W}_s) - I(W_s, \tilde{W}_s; Z^{mn}) - H(\tilde{W}_s|W_s, Z^{mn}) \tag{7.45}$$

$$\geq H(W_s, \tilde{W}_s) - I(V^{mn}; Z^{mn}) - H(\tilde{W}_s|W_s, Z^{mn}) \tag{7.46}$$

$$= H(W_s) + H(\tilde{W}_s) - I(V^{mn}; Z^{mn}) - H(\tilde{W}_s|W_s, Z^{mn}) \tag{7.47}$$

which is equivalent to

$$\frac{1}{mn} I(W_s; Z^{mn}) \leq \frac{1}{mn} I(V^{mn}; Z^{mn}) + \frac{1}{mn} H(\tilde{W}_s|W_s, Z^{mn}) - \frac{1}{mn} H(\tilde{W}_s). \tag{7.48}$$

Note that in (7.44), to keep the notation concise we do not list the randomness shared with the decoder (see the encoding procedure in Section 7.4.1) in the expression of the conditional entropy. Here, (7.44) is due to the chain rule of conditional entropy, (7.45) is due to the definition of mutual information, (7.46) comes from the data processing inequality, (7.47) is due to the independence of the secret message and the confusion message. In (7.48), we bound each term on the right hand side as

follows:

For the first term, we have $I(V^{mn}; Z^{mn}) \leq \sum_{i=1}^{mn} I(V_i; Z_i) \leq mnI(V; Z)$. Therefore,

$\frac{1}{mn} I(V^{mn}; Z^{mn}) \leq I(V; Z)$.

To bound the second term, suppose Eve obtains W_s and Z^{mn}, and wants to decode $\tilde{W}_s$. By symmetry of chaining construction, Eve can apply similar decoding rule as described in Section 7.4.1. However, this time Eve decodes from the mth block, then the block error probability of the whole *m-chain* block can be upper bounded by

$$P_e \leq (m-1) \sum_{i \in G_{Z \setminus Y}} Z(U_i | U^{i-1}, Y^n) + m \sum_{i \in G_{Y \wedge Z}} Z(U_i | U^{i-1}, Y^n) = O(2^{-n^\beta}) \quad (7.49)$$

for $\beta \in (0, 1/2)$. Hence, by applying Fano's inequality, we have

$$H(\tilde{W}_s | W_s, Z^{mn}) \leq H(P_e) + P_e \log |\tilde{W}_s| < H(P_e) + P_e[mnI(V; Z)]. \quad (7.50)$$

Therefore, as $n \to \infty$, $\frac{1}{mn} H(\tilde{W}_s | W_s, Z^{mn}) \to 0$.

For the last term, as $n \to \infty$, by (7.30) and (7.37), we have $(m-1)nI(V; Z) < H(\tilde{W}_s) < mnI(V; Z)$. Hence, as $m \to \infty$, $\frac{1}{mn} H(\tilde{W}_s) \to I(V; Z)$.

From the above, we know as $n \to \infty$ and $m \to \infty$, $\frac{1}{mn} I(W_s; Z^{mn}) \to 0$. Thus, the weak secrecy constraint is achieved.

7.5 Polar Coding for the Multiple Access Wiretap Channel

In this section, instead of achieving the corner point of (7.5) through standard polar coding techniques [76], we show how to achieve the rate pairs on the dominant face of (7.5), since reference [110] shows the former scheme is strictly suboptimal. Here, we consider the positive rate case in (7.5), i.e., $R_1 > 0$, $R_2 > 0$ and $R_1 + R_2 > 0$. We first consider a constant T in (7.5). Following the method given in [81, Sec. III. B.], we can generalize the result to a T with arbitrary distribution. For $k \in 1, 2$, let $\mathcal{V}_k$ be the corresponding alphabet of the channel prefixing V_k. As in Section 7.4, we assume the cardinality for the channel prefixing V_k is $|\mathcal{V}_k| = 2$ for illustration.

7.5.1 The Scheme

For a fixed input distribution in (7.6), consider two different MACs, the first MAC, P, consisting of two users and Bob and the second MAC, Q, consisting of the two users and Eve. In Fig. 7.2, we use a solid line to show the achievable region for the first MAC, P, and a dotted line to represent the second MAC, Q. Consider two rate pairs on the dominant faces of the channels P and Q, which we use blue and red points to denote in Fig. 7.2.

Reference [82] shows that there exist monotone permutations J^{2n} and K^{2n} for channels P and Q to achieve the blue and red points in Fig. 7.2. Since the blue rate pair is greater than the red rate pair in the sense of both rate of user 1 and rate of user 2, we can also achieve the red rate pair for channel P by the same monotone chain J^{2n}. In the following, we present a polar coding scheme such that we set the

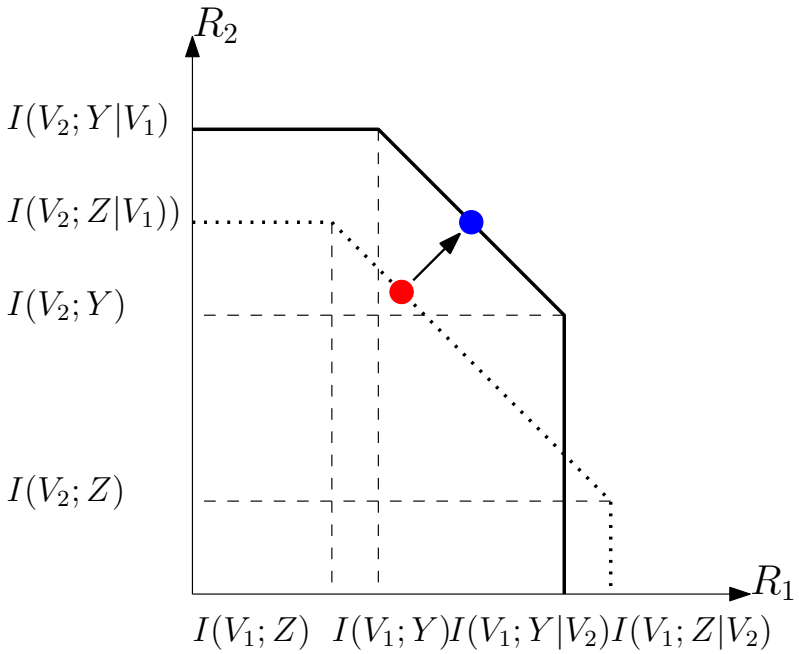

Figure 7.2: General MAC regions.

rate of the confusion message as the red rate pair and the rate of the secret message as the difference of the blue and red rate pairs.

For $k \in 1, 2$, let $U_k^n = V_k^n G_n$. Once we determine the distribution in (7.6), similar to (7.34), we can define $\mathcal{H}_{V_k}$. According to different monotone permutations, J^{2n}, we have different index sets for $\mathcal{L}_{V_k|Y,J}$. We define them as follows:

$$\mathcal{L}_{V_k|Y,J} = \{i \in [n] : Z(U_{k,i}|Y^n, J^{j-1}) \leq \delta_n,\ J_j = U_{k,i}\}, \tag{7.51}$$

where $\delta_n = 2^{-n^\beta}$ and $\beta \in (0, 1/2)$. Similarly, we can also define $\mathcal{L}_{V_k|Z,K}$ for another monotone permutation, K^{2n}.

The set $[n]$ for the user k can be partitioned into the following sets:

$$G_{Y\wedge Z}^{(k)} = \mathcal{H}_{V_k} \cap \mathcal{L}_{V_k|Y,J} \cap \mathcal{L}_{V_k|Z,K},$$

$$G_{Y\setminus Z}^{(k)} = \mathcal{H}_{V_k} \cap \mathcal{L}_{V_k|Y,J} \cap \mathcal{L}_{V_k|Z,K}^c,$$

$$G_{Z\setminus Y}^{(k)} = \mathcal{H}_{V_k} \cap \mathcal{L}_{V_k|Y,J}^c \cap \mathcal{L}_{V_k|Z,K},$$

$$B_{Y\wedge Z}^{(k)} = \mathcal{L}_{V_k} \cup (\mathcal{L}_{V_k|Y,J}^c \cap \mathcal{L}_{V_k|Z,K}^c). \tag{7.52}$$

Since we consider the positive rate case in (7.5), we have $|G_{Y\setminus Z}^{(k)}| > |G_{Z\setminus Y}^{(k)}|$. Pick $C_{Y\setminus Z}^{(k)} \subset G_{Y\setminus Z}^{(k)}$, such that $|C_{Y\setminus Z}^{(k)}| = |G_{Z\setminus Y}^{(k)}|$. Define the set $S^{(k)}$ as follows:

$$S^{(k)} = G_{Y\setminus Z}^{(k)} \setminus C_{Y\setminus Z}^{(k)}. \tag{7.53}$$

According to the result in [82], we have

$$\lim_{n\to\infty} \frac{1}{n}(|S^{(1)}| + |S^{(2)}|) = I(V_1, V_2; Y) - I(V_1, V_2; Z). \tag{7.54}$$

The encoding procedure for the two users are similar. We show the encoding procedure in Fig. 7.3 for user 1. For each user, we put the secret bits in the set $S^{(k)}$ and put random bits as the confusion message in the sets $G_{Y\wedge Z}^{(k)}$ and $C_{Y\setminus Z}^{(k)}$. Moreover, we chain the bits in the set $C_{Y\setminus Z}^{(k)}$ in the ith block to the set $G_{Z\setminus Y}^{(k)}$ in the $(i+1)$th block. To guarantee correct decoding, we freeze the sets $B_{Y\wedge Z}^{(k)}$ in each block, $G_{Z\setminus Y}^{(k)}$ in the 1st block, and $C_{Y\setminus Z}^{(k)}$ in the mth block. We use red crosses in Fig. 7.3 to denote the frozen sets.

The decoding procedure is from the 1st block to the mth block according to the monotone permutation J^{2n} for Bob. For the 1st block, since the bits Bob needs to decode are all in the sets $G_{Y\wedge Z}^{(k)}$ or $G_{Y\setminus Z}^{(k)}$, they all can be decoded reliably.

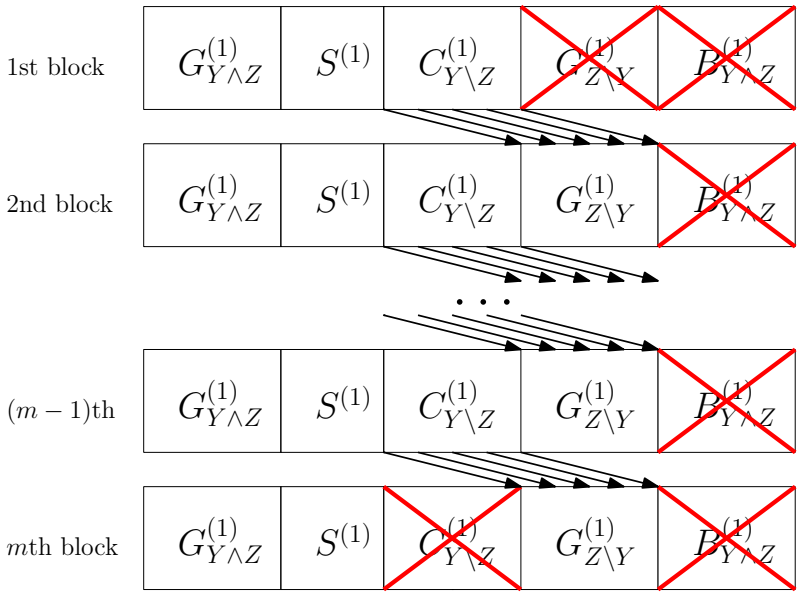

Figure 7.3: Chaining construction for the MAC-WTC for user 1.

For the 2nd block, due to the chaining construction in the encoding procedure, the remaining bits Bob needs to decode are also in the sets $G^{(k)}_{Y\wedge Z}$ or $G^{(k)}_{Y\setminus Z}$. Therefore, the correct decoding can also be guaranteed. The same procedure holds to the mth block. Since the confusion message and the secret message can be decoded reliably, we can guarantee that the rate in (7.54) can be achieved.

7.5.2 Equivocation Calculation

Following the notation given in Section 7.4.1, we show the equivocation rate calculation. For $k \in 1, 2$, let $W_s^{(k)}$ and $\tilde{W}_s^{(k)}$ denote the secret message and the confusion message sent by user k. Since we put the secret message in the set $S^{(k)}$ in each block, we have $W_s^{(k)} = \cup_{1\le i\le m} U_{k,S_i^{(k)}}$. For the confusion message, $\tilde{W}_s^{(k)}$, we have $\tilde{W}_s^{(k)} = \cup_{1\le i\le m, 1\le j\le (m-1)} U_{k,G^{(k)}_{Y\wedge Z i}} U_{k,C^{(k)}_{Y\setminus Z j}}$. For simplicity of notation, we let $W_s = W_s^{(1)} \cup W_s^{(2)}$ and $\tilde{W}_s = \tilde{W}_s^{(1)} \cup \tilde{W}_s^{(2)}$.

Similar to (7.44)–(7.47), we can calculate the equivocation rate as follows:

$$H(W_s|Z^{mn}) \geq H(W_s) + H(\tilde{W}_s) - I(V_1^{mn}, V_2^{mn}; Z^{mn}) - H(\tilde{W}_s|W_s, Z^{mn}), \quad (7.55)$$

which is equivalent to

$$\frac{1}{mn} I(W_s; Z^{mn}) \leq \frac{1}{mn} I(V_1^{mn}, V_2^{mn}; Z^{mn}) + \frac{1}{mn} H(\tilde{W}_s|W_s, Z^{mn}) - \frac{1}{mn} H(\tilde{W}_s). \quad (7.56)$$

To bound each term in (7.56), we only consider the second term since the first and third terms are similar to bounding in (7.48). These two terms can be upper bounded by ϵ, and $\epsilon \to 0$ as $n \to \infty$ and $m \to \infty$. For the second term, suppose Eve obtains W_s and Z^{mn}, and wants to decode $\tilde{W}_s$. This time Eve decodes from the mth block to the 1st block, and note that Eve decodes according to the monotone permutation K^{2n}. For the mth block, the bits that Eve needs to decode are in the set $G^{(k)}_{Y\wedge Z}$ and $G^{(k)}_{Z\setminus Y}$. Therefore, Eve can do the correct decoding. For the $(m-1)$th block, due to the chaining construction, the remaining bits that Eve needs to decode are also in the set $G^{(k)}_{Y\wedge Z}$ and $G^{(k)}_{Z\setminus Y}$. The same procedure holds to the 1st block. Since Eve can do the correct decoding, we can bound this term through Fano's inequality. Therefore, we can guarantee the conditions in (7.4).

7.6 Polar Coding for the Broadcast Channel with Confidential Messages

Before we show how to achieve the corner points of the rate region given in (7.8) by double chaining method, we briefly review the result in [80], which shows how to apply polar coding to achieve the rate pair $(R_1, R_2) = (I(V_1; Y_1), I(V_2; Y_2) - I(V_2; V_1))$ of the binning region. We first consider a constant T in (7.8). This result can be generalized to T with arbitrary distribution [81, Sec. III. B.]. Again, we consider binary code design for illustration.

7.6.1 Polar Coding for the Binning Region

Applying polar coding to achieve $R_1 = I(V_1; Y_1)$ is described in Section 7.3.1. Now, we discuss how to achieve $R_2 = I(V_2; Y_2) - I(V_2; V_1)$ following [80]. Let $U_2^n = V_2^n G_n$. Similar to (7.34), we can define $\mathcal{H}_{V_2}$ and $\mathcal{L}_{V_2|Y_2}$. Since V_1 and V_2 are dependent, by thinking of V_1 as the side information of V_2, we can further define the set $\mathcal{L}_{V_2|V_1}$. Similar to (7.35), the set $[n]$ can be partitioned into the following sets:

$$\begin{aligned} G_{Y_2 \wedge V_1} &= \mathcal{H}_{V_2} \cap \mathcal{L}_{V_2|Y_2} \cap \mathcal{L}_{V_2|V_1}, \\ G_{Y_2 \setminus V_1} &= \mathcal{H}_{V_2} \cap \mathcal{L}_{V_2|Y_2} \cap \mathcal{L}^c_{V_2|V_1}, \\ G_{V_1 \setminus Y_2} &= \mathcal{H}_{V_2} \cap \mathcal{L}^c_{V_2|Y_2} \cap \mathcal{L}_{V_2|V_1}, \\ B_{Y_2 \wedge V_1} &= \mathcal{H}^c_{V_2} \cup (\mathcal{L}^c_{V_2|Y_2} \cap \mathcal{L}^c_{V_2|V_1}). \end{aligned} \tag{7.57}$$

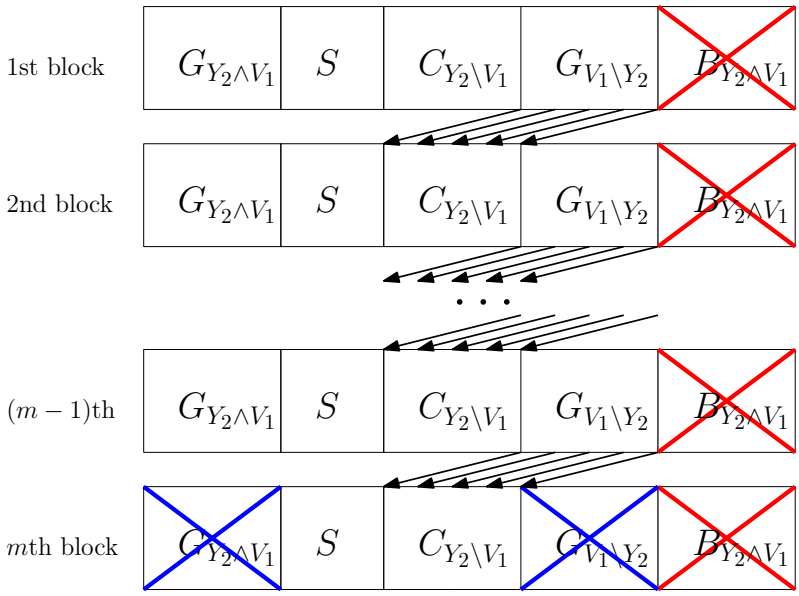

Figure 7.4: Chaining construction for the second user to achieve the binning region in a broadcast channel.

Roughly speaking, once the values for V_1 is known, the bits corresponding to the sets $G_{Y_2\wedge V_1}$ and $G_{V_1\setminus Y_2}$ can be determined. Since the second receiver observes Y_2, it can decode the set $G_{Y_2\wedge V_1}$ and $G_{Y_2\setminus V_1}$. To guarantee that the second receiver obtains the information bits in the set $G_{V_1\setminus Y_2}$, pick $C_{Y_2\setminus V_1} \subset G_{Y_2\setminus V_1}$ such that $|C_{Y_2\setminus V_1}| = |G_{V_1\setminus Y_2}|$ to serve the chaining purpose of repeating the information in the set $G_{V_1\setminus Y_2}$. Last, we put the information bits for the second user in the set $S = G_{Y_2\setminus V_1} \setminus C_{Y_2\setminus V_1}$. It can be verified that the rate of the second user is:

$$\lim_{n\to\infty} \frac{1}{n}|S| = I(V_2;Y_2) - I(V_2;V_1). \tag{7.58}$$

Consider the encoding procedure in Fig. 7.4. The information for the first receiver, V_1, is determined first. Since V_1 has been determined, the set $G_{Y_2\wedge V_1}$ and $G_{V_1\setminus Y_2}$ can also be determined from the 1st block to the mth block. It is important to

note that V_1 in the mth block is frozen and shared with the two receivers; therefore, the sets $G_{Y_2 \wedge V_1}$ and $G_{V_1 \setminus Y_2}$ can be decoded with the information of V_1 for the mth block, which we use blue crosses to denote in Fig. 7.4. Same as before, the red crosses denote the frozen sets in Fig. 7.4. By the chaining construction, for $1 \leq i < m$, we repeat the determined value in the set $G_{V_1 \setminus Y_2}$ in the ith block to the set $C_{Y_2 \setminus V_1}$ in the $(i+1)$th block. Last, we put the information bits for the second receiver in the set S in each block.

Decoding procedure for the second receiver starts from the mth block. For the mth block, the second user only needs to decode the information in the set S and $C_{Y_2 \setminus V_1}$. To decode the $(m-1)$th block, since the bits in the set $G_{V_1 \setminus Y_2}$ can be obtained from the mth block due to the chaining construction of the encoding process, the second user only needs to decode the bits in the set $G_{Y_2 \wedge V_1}$ and $G_{Y_2 \setminus V_1}$. The same procedure holds till the 1st block, and the information in the set S can be decoded reliably.

7.6.2 The Scheme

Here, we introduce a *double chaining* method to achieve the *double binning* rate pair $(R_1, R_2) = (I(V_1;Y_1) - I(V_1;V_2) - I(V_1;Y_2|V_2), I(V_2;Y_2) - I(V_2;V_1) - I(V_2;Y_1|V_1))$, which is the corner point of (7.8) when T is a constant. Let $U_2^n = V_2^n G_n$. Once we determine the distribution in (7.9), we can define $\mathcal{H}_{V_2}$, $\mathcal{L}_{V_2|Y_2}$ and $\mathcal{L}_{V_2|V_1}$. We can further define $\mathcal{L}_{V_2|Y_1,V_1}$ as in Section 7.6.1. The set $[n]$ can be partitioned into the

following sets:

$$
\begin{aligned}
A &= \mathcal{H}_{V_2} \cap \mathcal{L}_{V_2|Y_2} \cap \mathcal{L}_{V_2|V_1} \cap \mathcal{L}_{V_2|V_1,Y_1}, \\
B &= \mathcal{H}_{V_2} \cap \mathcal{L}_{V_2|Y_2} \cap \mathcal{L}^c_{V_2|V_1} \cap \mathcal{L}_{V_2|V_1,Y_1}, \\
C &= \mathcal{H}_{V_2} \cap \mathcal{L}_{V_2|Y_2} \cap \mathcal{L}^c_{V_2|V_1} \cap \mathcal{L}^c_{V_2|V_1,Y_1}, \\
D &= \mathcal{H}_{V_2} \cap \mathcal{L}^c_{V_2|Y_2} \cap \mathcal{L}_{V_2|V_1} \cap \mathcal{L}_{V_2|V_1,Y_1}, \\
E &= \mathcal{H}_{V_2} \cap \mathcal{L}^c_{V_2|Y_2} \cap \mathcal{L}^c_{V_2|V_1} \cap \mathcal{L}_{V_2|V_1,Y_1}, \\
F &= \mathcal{H}^c_{V_2} \cup (\mathcal{L}^c_{V_2|Y_2} \cap \mathcal{L}^c_{V_2|V_1} \cap \mathcal{L}^c_{V_2|V_1,Y_1}).
\end{aligned} \tag{7.59}
$$

Similarly, let $U_1^n = V_1^n G_n$. We can partition the set $[n]$ for user 1 as (7.59) by changing the subscript 2 to 1 and 1 to 2.

Similar to (7.36) and (7.37), we have

$$
\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}|A \cup B \cup C| &= I(V_2; Y_2), \\
\lim_{n\to\infty} \frac{1}{n}|A \cup D| &= I(V_2; V_1), \\
\lim_{n\to\infty} \frac{1}{n}|B \cup E| &= I(V_2; Y_1|V_1).
\end{aligned} \tag{7.60}
$$

Here, we consider the case $R_1 > 0$ and $R_2 > 0$. Therefore, we can pick $C_1 \subset C$ with $|C_1| = |D|$, $C_2 \subset C$ with $|C_2| = |E|$, and $C_1 \cap C_2 = \emptyset$. Define the set S as follows:

$$
S = C \setminus (C_1 \cup C_2). \tag{7.61}
$$

By (7.60), we also have

$$\lim_{n\to\infty}\frac{1}{n}|S| = I(V_2;Y_2) - I(V_2;V_1) - I(V_2;Y_1|V_1). \tag{7.62}$$

Now, we consider the encoding procedure. Assume we determine the information for the first receiver, V_1, at first. As described in Section 7.6.1, to guarantee the correct decoding of the second user, V_1 in the mth block is frozen and shared with the two receivers. As shown in Fig. 7.5, the red crosses denote the frozen sets. We put the secret message in the set S from the 1st block to the $(m-1)$th block. Later, we will show that the rate

$$R_1 = \left(\frac{m-1}{m}\right)[I(V_1;Y_1) - I(V_1;V_2) - I(V_1;Y_2|V_2)] \tag{7.63}$$

can be achieved. To guarantee the secrecy, we put the random bits in the set A, B, D and E in the 1st block. To ensure the reliability for the user 1, we chain the message in the sets D and E to the sets C_1 and C_2 in the 2nd block. The same procedure holds till the $(m-2)$th block. For the $(m-1)$th block, we still chain the sets D and E from the $(m-2)$th block to the sets C_1 and C_2; however, we freeze the set D and E in the $(m-1)$th block to guarantee correct decoding for user 1.

For the second user, we put the secret message to the set S from the 1st block to the mth block, and will show that the rate

$$R_2 = I(V_2;Y_2) - I(V_2;V_1) - I(V_2;Y_1|V_1) \tag{7.64}$$

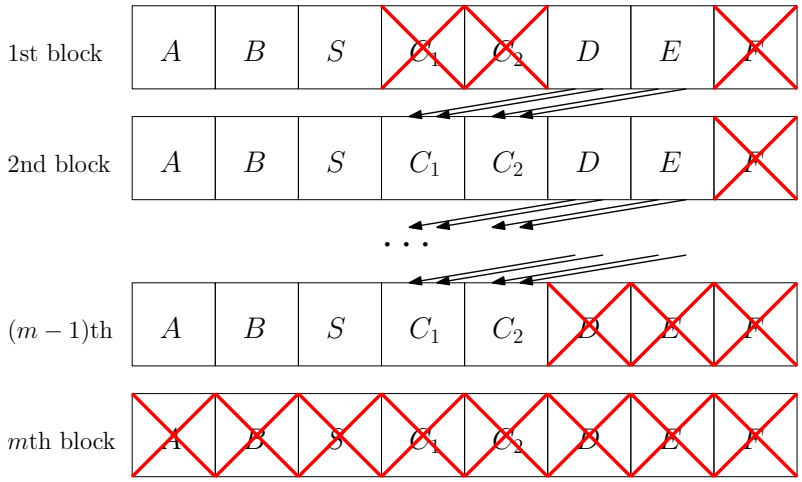

Figure 7.5: Chaining construction for the BC-CM for user 1.

can be achieved. To guarantee the secrecy, we put the random bits to the sets B and E as the confusion message from the 1st block to the $(m-1)$th block. Since V_1 has been determined, the sets A and D can also be determined with the knowledge of V_1. For the first chaining construction, for $1 \leq i < m$, we repeat the determined value in the set D in the ith block to the set C_1 in the $(i+1)$th block. For the second chaining construction, for $1 \leq i < m$, we repeat the determined value in the set E in the ith block to the set C_2 in the $(i+1)$th block. As described in Section 7.6.1, V_1 in the mth block is frozen and shared with the two receivers; thus, the sets A and D can be decoded with the information of V_1 for the mth block, which we use blue crosses to denote in Fig. 7.6. Same as before, the red crosses denote the frozen sets in Fig. 7.6. For the 1st block, we freeze the sets C_1 and C_2, and for the mth block, we freeze the set E, to guarantee the reliability.

The decoding procedure for the two users are similar. They both decode from the mth block to the 1st block. Let us use user 2 for illustration. For the mth block, since user 2 knows V_1, it can decode the sets A, B, C and D. Through the chaining

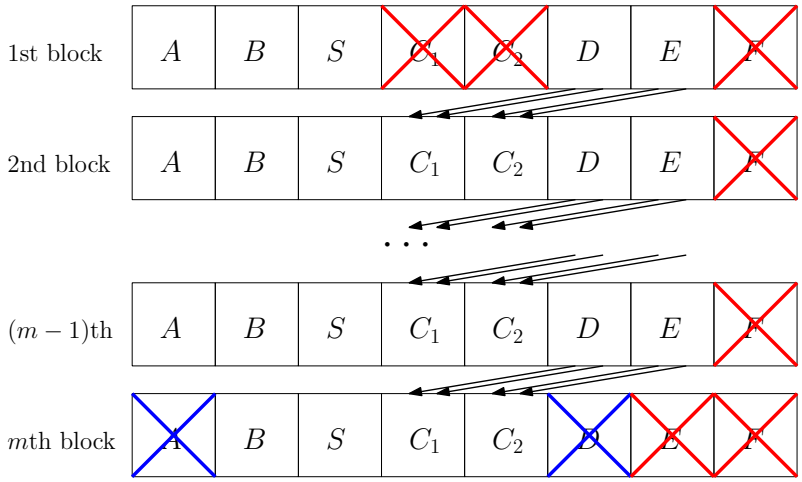

Figure 7.6: Chaining construction for the BC-CM for user 2.

construction, the decoder only needs to decode the sets A, B and C in the $(m-1)$th block. The same procedure holds till the 2nd block. For the 1st block, due to the chaining construction and the frozen sets, the decoder only needs to decode the sets A, B and S, which can be done reliably.

7.6.3 Reliability

The block error probability of the first and second user can be upper bounded by

$$\begin{aligned} P_{e,1} &\leq (m-2) \sum_{i \in A \cup B \cup C} Z(U_{1,i}|U_1^{i-1}, Y_1^n) + \sum_{i \in A \cup B \cup S} Z(U_{1,i}|U_1^{i-1}, Y_1^n) = O(2^{-n^\beta}), \\ P_{e,2} &\leq (m-2) \sum_{i \in A \cup B \cup C} Z(U_{2,i}|U_2^{i-1}, Y_2^n) + \sum_{i \in A \cup B \cup S} Z(U_{1,i}|U_2^{i-1}, Y_2^n) \\ &\quad + \sum_{i \in B \cup C} Z(U_{1,i}|U_2^{i-1}, Y_2^n) = O(2^{-n^\beta}) \end{aligned} \tag{7.65}$$

for any $\beta \in (0, 1/2)$ with complexity $O(n \log n)$. Therefore, the rate pair in (7.63) and (7.64) can be achieved reliably. Thus, as $m \to \infty$, we can achieve the rate pair

in (7.8).

7.6.4 Equivocation Calculation

Following the notation given in Section 7.4.3, we show the equivocation calculation for receiver 2, and this result can be extended to receiver 1 by symmetry. Since we put the secret message in the set S in each block, we have $W_{s,1} = \cup_{1\leq i<m} U_{1,S_i}$. For the confusion message, $\tilde{W}_{s,1}$, we have $\tilde{W}_{s,1} = \cup_{1\leq i<m, 1\leq j<(m-1)} U_{1,(A\cup B)_i} U_{1,(D\cup E)_j}$.

We can calculate the equivocation rate as follows:

$$
\begin{aligned}
H(W_{s,1}|Y_2^{mn}) &\geq H(W_{s,1}|Y_2^{mn}, V_2^{mn}, T^{mn}) && (7.66)\\
&= H(W_{s,1}, Y_2^{mn}|V_2^{mn}, T^{mn}) - H(Y_2^{mn}|V_2^{mn}, T^{mn}) && (7.67)\\
&= H(W_{s,1}, V_1^{mn}, Y_2^{mn}|V_2^{mn}, T^{mn}) - H(V_1^{mn}|Y_2^{mn}, V_2^{mn}, T^{mn}, W_{s,1})\\
&\quad - H(Y_2^{mn}|V_2^{mn}, T^{mn}) && (7.68)\\
&= H(W_{s,1}, V_1^{mn}|V_2^{mn}, T^{mn}) + H(Y_2^{mn}|V_1^{mn}, V_2^{mn}, T^{mn}, W_{s,1})\\
&\quad - H(V_1^{mn}|Y_2^{mn}, V_2^{mn}, T^{mn}, W_{s,1}) - H(Y_2^{mn}|V_2^{mn}, T^{mn}) && (7.69)\\
&= H(W_{s,1}, V_1^{mn}|V_2^{mn}, T^{mn}) + H(Y_2^{mn}|V_1^{mn}, V_2^{mn}, T^{mn})\\
&\quad - H(V_1^{mn}|Y_2^{mn}, V_2^{mn}, T^{mn}, W_{s,1}) - H(Y_2^{mn}|V_2^{mn}, T^{mn}) && (7.70)\\
&= H(W_{s,1}, V_1^{mn}|V_2^{mn}, T^{mn}) - H(V_1^{mn}|Y_2^{mn}, V_2^{mn}, T^{mn}, W_{s,1})\\
&\quad - I(V_1^{mn}; Y_2^{mn}|V_2^{mn}, T^{mn}) && (7.71)\\
&\geq H(V_1^{mn}|V_2^{mn}, T^{mn}) - H(V_1^{mn}|Y_2^{mn}, V_2^{mn}, T^{mn}, W_{s,1})\\
&\quad - I(V_1^{mn}; Y_2^{mn}|V_2^{mn}, T^{mn}) && (7.72)
\end{aligned}
$$

where (7.66) is due to conditioning reduces entropy, and (7.67), (7.68) and (7.69) are due to the chain rule of entropy. Due to the Markov chain $W_{s,1} \to (V_1^{mn}, V_2^{mn}, T^{mn}) \to Y_2^{mn}$, we have $I(W_{s,1}; Y_2^{mn} | V_1^{mn}, V_2^{mn}, T^{mn}) = 0$. Hence, (7.70) holds. (7.71) is due to the definition of conditional mutual information, and (7.72) is due to the chain rule of entropy.

Consider the first term in (7.72)

$$H(V_1^{mn} | V_2^{mn}, T^{mn}) = H(V_1^{mn} | T^{mn}) - I(V_1^{mn}; V_2^{mn} | T^{mn}). \tag{7.73}$$

Therefore, we can lower bound the sum of the first and the third term in (7.72) as

$$(m-2)nI(V_1; Y_1 | T) - mnI(V_1; V_2 | T) - mnI(V_1; Y_2 | V_2, T). \tag{7.74}$$

For the second term, $H(V_1^{mn} | Y_2^{mn}, V_2^{mn}, T^{mn}, W_{s,1}) = H(\tilde{W}_{s,1} | Y_2^{mn}, V_2^{mn}, T^{mn}, W_{s,1})$. Suppose receiver 2 knows Y_2^{mn}, V_2^{mn} and $W_{s,1}$, and tries to decode $\tilde{W}_{s,1}$. From Fig. 7.5, it can decode from the 1st block to the $(m-1)$th block, and the block error probability can be upper bounded by $O(2^{-n^\beta})$ for $\beta \in (0, 1/2)$. By applying Fano's inequality, we have $H(\tilde{W}_{s,1} | Y_2^{mn}, V_2^{mn}, T^{mn}, W_{s,1}) \leq mn\epsilon$. After we bound the three terms as above, we have

$$H(W_{s,1} | Y_2^{mn}) \geq mn \left[\left(1 - \frac{2}{m} \right) I(V_1; Y_1 | T) - I(V_1; V_2 | T) - I(V_1; Y_2 | V_2, T) - \epsilon \right]. \tag{7.75}$$

Therefore, as $n \to \infty$ and $m \to \infty$, the secrecy constraints in (7.7) hold.

7.7 Polar Coding for the Interference Channel with Confidential Messages

In the following, we show how to achieve the corner points of the rate region given in (7.11). By simple modification, this method can achieve the entire rate region. Note that given T, V_1 and V_2 are independent as in (7.12). Therefore, achieving (7.11) is also equivalent to achieving the rate pair $(R_1, R_2) = (I(V_1; Y_1) - I(V_1; Y_2, V_2), I(V_2; Y_2) - I(V_2; Y_1, V_1))$. We consider a constant T in (7.11), and binary code design for illustration.

7.7.1 The Scheme

Here, we discuss the code design for user 1 only, as the code design method for the two users is similar. Let $U_1^n = V_1^n G_n$. Once we determine the distribution in (7.12), similar to (7.34), we can define $\mathcal{H}_{V_1}$ and $\mathcal{L}_{V_1|Y_1}$. We can further define

$$\mathcal{L}_{V_1|Y_2,V_2} = \{i \in [n] : Z(U_{1,i}|U_1^{i-1}, Y_2^n, V_2^n) \leq \delta_n\}, \tag{7.76}$$

where $\delta_n = 2^{-n^\beta}$ and $\beta \in (0, 1/2)$.

By thinking of Y_1 as Y and $[Y_2, V_2]$ as Z in (7.35), we can partition the set $[n]$

into the following:

$$
\begin{aligned}
G_{Y_1 \wedge [Y_2,V_2]} &= \mathcal{H}_{V_1} \cap \mathcal{L}_{V_1|Y_1} \cap \mathcal{L}_{V_1|[Y_2,V_2]}, \\
G_{Y_1 \setminus [Y_2,V_2]} &= \mathcal{H}_{V_1} \cap \mathcal{L}_{V_1|Y_1} \cap \mathcal{L}^c_{V_1|[Y_2,V_2]}, \\
G_{[Y_2,V_2] \setminus Y_1} &= \mathcal{H}_{V_1} \cap \mathcal{L}^c_{V_1|Y_1} \cap \mathcal{L}_{V_1|[Y_2,V_2]}, \\
B_{Y_1 \wedge [Y_2,V_2]} &= \mathcal{H}^c_{V_1} \cup (\mathcal{L}^c_{V_1|Y_1} \cap \mathcal{L}^c_{V_1|[Y_2,V_2]}).
\end{aligned} \tag{7.77}
$$

Similar to (7.36), we also have

$$
\begin{aligned}
\mathcal{I}_{Y_1} &= \mathcal{H}_{V_1} \cap \mathcal{L}_{V_1|Y_1}, \\
\mathcal{I}_{[Y_2,V_2]} &= \mathcal{H}_{V_1} \cap \mathcal{L}_{V_1|[Y_2,V_2]}, \\
\mathcal{F}_r^{Y_1} &= \mathcal{H}_{V_1} \cap \mathcal{L}^c_{V_1|Y_1}, \\
\mathcal{F}_r^{[Y_2,V_2]} &= \mathcal{H}_{V_1} \cap \mathcal{L}^c_{V_1|[Y_2,V_2]}, \\
\mathcal{F}_d &= \mathcal{H}^c_{V_1}.
\end{aligned} \tag{7.78}
$$

Same as (7.37), we have

$$
\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} |\mathcal{I}_{Y_1}| &= I(V_1; Y_1), \\
\lim_{n\to\infty} \frac{1}{n} |\mathcal{I}_{[Y_2,V_2]}| &= I(V_1; Y_2, V_2).
\end{aligned} \tag{7.79}
$$

Here, we consider the case $R_1 > 0$; therefore, we have $|G_{Y_1 \setminus [Y_2,V_2]}| > |G_{[Y_2,V_2] \setminus Y_1}|$. Pick a set, $C_{Y_1 \setminus [Y_2,V_2]}$, such that $C_{Y_1 \setminus [Y_2,V_2]} \subset G_{Y_1 \setminus [Y_2,V_2]}$ and $|C_{Y_1 \setminus [Y_2,V_2]}| = |G_{[Y_2,V_2] \setminus Y_1}|$.

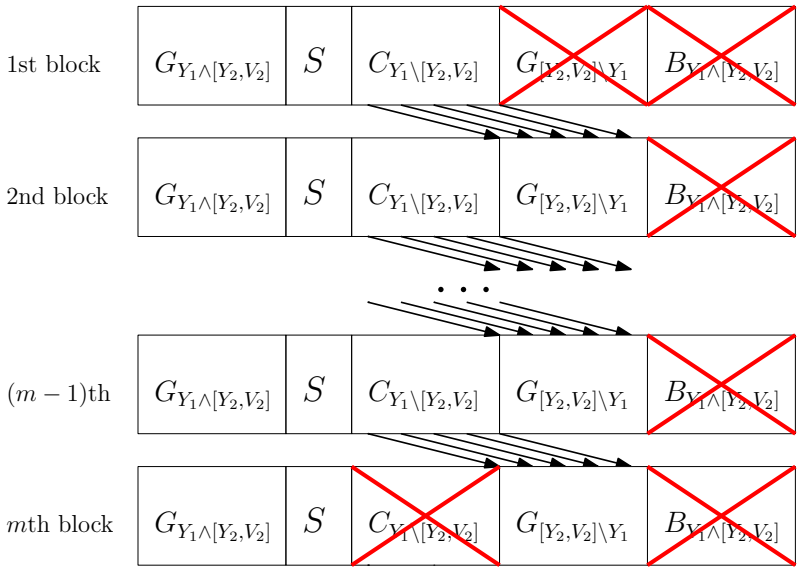

Figure 7.7: Chaining construction for the IC-CM for user 1.

Last, we define the set S similar to (7.38) as

$$S = G_{Y_1\setminus[Y_2,V_2]} \setminus C_{Y_1\setminus[Y_2,V_2]}. \tag{7.80}$$

From (7.79), we have

$$\lim_{n\to\infty} \frac{1}{n}|S| = I(V_1;Y_1) - I(V_1;Y_2,V_2). \tag{7.81}$$

The polar coding scheme construction for IC-CM is almost the same as the code design for the wiretap channel in Section 7.4.1. By replacing Y by Y_1 and Z by $[Y_2, V_2]$ in Section 7.4.1, we can construct the codebook for user 1 shown in Fig. 7.7, where the red crosses indicate that the sub-channels are frozen. Same as before, we put the secret message in the set S, and put the random bits in the sets $G_{Y_1\wedge[Y_2,V_2]}$ and $C_{Y_1\setminus[Y_2,V_2]}$ as the confusion message. By replacing U by U_1, $U_{\mathcal{F}_r^Y}$ by $U_{1,\mathcal{F}_r^{Y_1}}$, and

$U_{\mathcal{F}_r^Z}$ by $U_{1,\mathcal{F}_r^{[Y_2,V_2]}}$ as defined in (7.78), we can follow the same encoding and decoding procedures given in Section 7.4.1. The secrecy rate $R_1 = I(V_1;Y_1) - I(V_1;Y_2,V_2)$ can be achieved reliably since the secret message in the set S can be correctly decoded as described in Section 7.4.2, where the set S ensures the rate given in (7.81).

7.7.2 Equivocation Calculation

Following the notation given in Section 7.4.3, we show the equivocation calculation for receiver 2, and this result can be extended to receiver 1 by symmetry. Since we put the secret message in the set S in each block, we have $W_{s,1} = \cup_{1\le i\le m} U_{1,S_i}$. For the confusion message, $\tilde{W}_{s,1}$, we have $\tilde{W}_{s,1} = \cup_{1\le i\le m, 1\le j<m} U_{1,G_{Y_1\wedge[Y_2,V_2]^i}} U_{1,C_{Y_1\setminus[Y_2,V_2]^j}}$.

We can calculate the equivocation rate as follows (see (7.66)–(7.72)):

$$\begin{aligned} H(W_{s,1}|Y_2^{mn}) \ge & H(V_1^{mn}|V_2^{mn},T^{mn}) - H(V_1^{mn}|Y_2^{mn},V_2^{mn},T^{mn},W_{s,1}) \\ & - I(V_1^{mn};Y_2^{mn}|V_2^{mn},T^{mn}). \end{aligned} \tag{7.82}$$

Now, we discuss each term in (7.82). Since given $T^{mn} = t^{mn}$, V_1^{mn} and V_2^{mn} are independent, we have $H(V_1^{mn}|V_2^{mn},T^{mn}) = H(V_1^{mn}|T^{mn})$, and $I(V_1^{mn};Y_2^{mn}|V_2^{mn},T^{mn}) = I(V_1^{mn};Y_2^{mn},V_2^{mn}|T^{mn})$. Then, we can lower bound the sum of the first and third term as

$$(m-1)nI(V_1;Y_1|T) - mnI(V_1;Y_2,V_2|T). \tag{7.83}$$

For the second term, $H(V_1^{mn}|Y_2^{mn},V_2^{mn},T^{mn},W_{s,1}) = H(\tilde{W}_{s,1}|Y_2^{mn},V_2^{mn},T^{mn},W_{s,1})$. Suppose receiver 2 knows Y_2^{mn}, V_2^{mn} and $W_{s,1}$, and tries to decode $\tilde{W}_{s,1}$. From

Fig. 7.7, it can decode from the mth block to the 1st block, and the block error probability can be upper bounded by

$$P_e \leq (m-1) \sum_{i \in G_{[Y_2,V_2]\setminus Y_1}} Z(U_{1,i}|U_1^{i-1}, Y_2^n) + m \sum_{i \in G_{Y_1 \wedge [Y_2,V_2]}} Z(U_{1,i}|U_1^{i-1}, Y_2^n) = O(2^{-n^\beta}) \tag{7.84}$$

for $\beta \in (0, 1/2)$. Hence, by applying Fano's inequality, we have

$$H(\tilde{W}_{s,1}|Y_2^{mn}, V_2^{mn}, T^{mn}, W_{s,1}) \leq H(P_e) + P_e \log |\tilde{W}_s| < H(P_e) + P_e[mnI(V_1; Y_2, V_2|T)]. \tag{7.85}$$

Therefore, as $n \to \infty$, $H(\tilde{W}_{s,1}|Y_2^{mn}, V_2^{mn}, T^{mn}, W_{s,1}) \to 0$.

Finally, considering (7.83) and (7.85), we know that as $n \to \infty$ and $m \to \infty$, the secrecy constraints in (7.10) hold.

7.8 Conclusion

We proposed practical coding schemes based on polar coding for the general wiretap channel, multiple access wiretap channel (MAC-WTC), broadcast channel with confidential messages (BC-CM), and interference channel with confidential messages (IC-CM). By applying the chaining construction and polar coding for asymmetric channels, we proposed a polar coding scheme to achieve the secrecy capacity of the general wiretap channel. Compared to the previous work, our construction has better decoding error probability and it can be constructed more efficiently. For the MAC-WTC, we combined our coding scheme for the general wiretap channel

with the technique of monotone chain rule. For the BC-CM, we introduced double chaining construction to guarantee the secrecy and achieve the binning rate. For the IC-CM, we viewed the output of the channel as the actual output and the intended message carrying signal, and applied our coding scheme for the general wiretap channel.

CHAPTER 8

Conclusions

In this dissertation, we studied the privacy problem via the PIR problem with a focus on its interactions with available side information. We studied the security problem via the wiretap channel with a focus on the design of practical coding schemes to achieve information-theoretically achievable random-coding based secrecy rates.

In Chapter 2, we studied the cache-aided PIR problem from N non-communicating and replicated databases, when the cache stores uncoded bits that are unknown to the databases. We determined inner and outer bounds for the optimal normalized download cost $D^*(r)$ as a function of the total number of messages K, the number of databases N, and the caching ratio r. Both inner and outer bounds are piece-wise linear functions in r (for fixed N, K) that consist of K line segments. For general K, N, and r, we showed that the largest gap between the achievability and the converse bounds is $\frac{1}{6}$. The outer bound shows significant reduction in the download cost with respect to the case when the cache content is fully known at all databases [28].

In Chapter 3, we studied the cache-aided PIR problem when the cache stores uncoded bits that are partially known to the databases. We determined inner and

outer bounds for the optimal normalized download cost $D^*(r)$. Both inner and outer bounds are piece-wise linear functions in r that consist of K line segments. The achievable scheme extends the greedy scheme in [12] so that it starts with exploiting the cache bits as side information. For fixed K, N, there are $K-1$ non-degenerate corner points. These points differ in the number of cached bits that contribute in generating one side information equation. The achievability for the remaining caching ratios is done by memory-sharing between the two adjacent corner points that enclose that caching ratio r. For the converse, we extend the induction-based techniques in [12] and Chapter 2 to account for the availability of uncoded and partially known side information at the retriever. The converse proof hinges on developing K lower bounds on the length of the undesired portion of the answer string. By applying induction on each bound separately, we obtain the piece-wise linear inner bound. For general K, N, and r, we showed that the largest additive gap between the achievability and the converse bounds is $\frac{5}{32}$. We observed that the achievable download cost here is larger than that in the previous case due to the partial knowledge of the databases regarding the cache content.

In Chapter 4, we have introduced PIR with partially known private side information as a natural model for studying practical PIR problems with cached side information. In this model, the nth database provides the user with m_n side information messages in the prefetching phase such that $\sum_{n=1}^{N} m_n \leq M$, hence, each database has *partial knowledge* about the side information. Based on this side information, the user designs a retrieval scheme that does not reveal the identity of the desired message or the identities of the remaining $M-m_n$ messages to the nth

database. For this model, we determined the exact capacity to be $C = \frac{1-\frac{1}{N}}{1-(\frac{1}{N})^{K-M}}$. The capacity is attained for any prefetching strategy that satisfies the cache memory size constraint with equality. The achievable scheme in [33] can also be used for this model. We further proposed another PIR scheme which requires smaller subpacketization and field size for the case of uniform prefetching. Uniform prefetching, when feasible, is optimal. Interestingly, the capacity expression we derive for this problem is exactly the same as the capacity expression for the PIR problem with completely unknown side information [33]. Therefore, our result implies that there is no loss in employing the same databases for prefetching and retrieval purposes.

In Chapter 5, we studied PIR-PSI under a storage constraint. In this model, the user randomly chooses M messages and caches the first r_i portion of the chosen messages for $i = 1, \ldots, M$ subject to the memory size constraint $\sum_{i=1}^{M} r_i = S$. In the retrieval phase, the user wishes to retrieve a message such that no individual database can learn the identity of the desired message and the identities of the cached messages. For each caching scheme, i.e., $(r_1, \ldots, r_M)$, we characterized the optimal normalized download cost to be $D^* = 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1-M}} + \frac{1-r_M}{N^{K-M}} + \frac{1-r_{M-1}}{N^{K-M+1}} + \cdots + \frac{1-r_1}{N^{K-1}}$. We conclude that for a fixed memory size S, the uniform caching scheme caching K messages achieves the lowest normalized download cost.

In Chapter 6, we considered the PIR problem from decentralized uncoded caching databases. We showed that uniform and random decentralized caching scheme, originally proposed in [54] for the problem of decentralized coded caching, results in the lowest expected normalized download cost in the PIR phase. We char-

acterized the expected normalized download cost to be $\frac{D}{L} = \sum_{n=1}^{N+1} \binom{N}{n-1} \mu^{n-1}(1-\mu)^{N+1-n} \left(1 + \frac{1}{n} + \cdots + \frac{1}{n^{K-1}}\right)$. For the achievability, we applied the PIR scheme in [12] for all subfiles. For the converse, we first applied the lower bound derived in [43], and to compare different probability distributions in the caching phase, we focused on the marginal distributions on individual bits. By using the nature of decentralization and uncoded caching, we further lower bounded the normalized download cost. Finally, we showed the matching converse for the expected normalized download cost, obtaining the exact capacity of the resulting PIR problem.

In Chapter 7, we proposed practical coding schemes based on polar coding for the general wiretap channel, multiple access wiretap channel (MAC-WTC), broadcast channel with confidential messages (BC-CM), and interference channel with confidential messages (IC-CM). By applying the chaining construction and polar coding for asymmetric channels, we proposed a polar coding scheme to achieve the secrecy capacity of the general wiretap channel. For the MAC-WTC, we combined our coding scheme for the general wiretap channel with the technique of monotone chain rule. For the BC-CM, we introduced double chaining construction to guarantee the secrecy and achieve the binning rate. For the IC-CM, we viewed the output of the channel as the actual output and the intended message carrying signal, and applied our coding scheme for the general wiretap channel.

The contents of Chapter 2 are published in [111, 112], Chapter 3 in [113, 114], Chapter 4 in [115, 116], Chapter 5 in [117, 118], Chapter 6 in [119] and Chapter 7 in [120, 121].

Bibliography

[1] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. Private information retrieval. *Journal of the ACM*, 45(6):965–981, 1998.

[2] W. Gasarch. A survey on private information retrieval. In *Bulletin of the EATCS*, 2004.

[3] C. Cachin, S. Micali, and M. Stadler. Computationally private information retrieval with polylogarithmic communication. In *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 1999.

[4] R. Ostrovsky and W. Skeith III. A survey of single-database private information retrieval: Techniques and applications. In *International Workshop on Public Key Cryptography*, pages 393–411. Springer, 2007.

[5] S. Yekhanin. Private information retrieval. *Communications of the ACM*, 53(4):68–73, 2010.

[6] N. B. Shah, K. V. Rashmi, and K. Ramchandran. One extra bit of download ensures perfectly private information retrieval. In *IEEE ISIT*, June 2014.

[7] G. Fanti and K. Ramchandran. Efficient private information retrieval over unsynchronized databases. *IEEE Journal of Selected Topics in Signal Processing*, 9(7):1229–1239, October 2015.

[8] T. Chan, S. Ho, and H. Yamamoto. Private information retrieval for coded storage. In *IEEE ISIT*, June 2015.

[9] A. Fazeli, A. Vardy, and E. Yaakobi. Codes for distributed PIR with low storage overhead. In *IEEE ISIT*, June 2015.

[10] R. Tajeddine and S. El Rouayheb. Private information retrieval from MDS coded data in distributed storage systems. In *IEEE ISIT*, July 2016.

[11] H. Sun and S. A. Jafar. The capacity of private information retrieval. In *IEEE Globecom*, December 2016.

[12] H. Sun and S. A. Jafar. The capacity of private information retrieval. *IEEE Transactions on Information Theory*, 63(7):4075–4088, July 2017.

[13] S. A. Jafar. Blind interference alignment. *IEEE Journal of Selected Topics in Signal Processing*, 6(3):216–227, June 2012.

[14] H. Sun and S. A. Jafar. Blind interference alignment for private information retrieval. 2016. Available at arXiv:1601.07885.

[15] H. Sun and S. A. Jafar. The capacity of robust private information retrieval with colluding databases. *IEEE Transactions on Information Theory*, 64(4):2361–2370, April 2018.

[16] H. Sun and S. A. Jafar. The capacity of symmetric private information retrieval. *IEEE Transactions on Information Theory*, 2018.

[17] K. Banawan and S. Ulukus. The capacity of private information retrieval from coded databases. *IEEE Transactions on Information Theory*, 64(3):1945–1956, March 2018.

[18] H. Sun and S. A. Jafar. Optimal download cost of private information retrieval for arbitrary message length. *IEEE Transactions on Information Forensics and Security*, 12(12):2920–2932, December 2017.

[19] Q. Wang and M. Skoglund. Symmetric private information retrieval for MDS coded distributed storage. 2016. Available at arXiv:1610.04530.

[20] H. Sun and S. A. Jafar. Multiround private information retrieval: Capacity and storage overhead. *IEEE Transactions on Information Theory*, 64(8):5743 – 5754, August 2018.

[21] R. Freij-Hollanti, O. Gnilke, C. Hollanti, and D. Karpuk. Private information retrieval from coded databases with colluding servers. *SIAM Journal on Applied Algebra and Geometry*, 1(1):647–664, 2017.

[22] H. Sun and S. A. Jafar. Private information retrieval from MDS coded data with colluding servers: Settling a conjecture by Freij-Hollanti et al. *IEEE Transactions on Information Theory*, 64(2):1000–1022, February 2018.

[23] R. Tajeddine, O. W. Gnilke, D. Karpuk, R. Freij-Hollanti, C. Hollanti, and S. El Rouayheb. Private information retrieval schemes for coded data with arbitrary collusion patterns. In *IEEE ISIT*, June 2017.

[24] K. Banawan and S. Ulukus. Multi-message private information retrieval: Capacity results and near-optimal schemes. *IEEE Transactions on Information Theory*, 64(10):6842–6862, October 2018.

[25] Y. Zhang and G. Ge. A general private information retrieval scheme for MDS coded databases with colluding servers. 2017. Available at arXiv: 1704.06785.

[26] Y. Zhang and G. Ge. Multi-file private information retrieval from MDS coded databases with colluding servers. 2017. Available at arXiv: 1705.03186.

[27] K. Banawan and S. Ulukus. The capacity of private information retrieval from Byzantine and colluding databases. *IEEE Transactions on Information Theory*. To appear. Also available at arXiv:1706.01442.

[28] R. Tandon. The capacity of cache aided private information retrieval. In *IEEE Allerton*, September 2017.

[29] Q. Wang and M. Skoglund. Secure symmetric private information retrieval from colluding databases with adversaries. 2017. Available at arXiv:1707.02152.

[30] R. Tajeddine and S. El Rouayheb. Robust private information retrieval on coded data. In *IEEE ISIT*, June 2017.

[31] Q. Wang and M. Skoglund. Linear symmetric private information retrieval for MDS coded distributed storage with colluding servers. 2017. Available at arXiv:1708.05673.

[32] S. Kadhe, B. Garcia, A. Heidarzadeh, S. El Rouayheb, and A. Sprintson. Private information retrieval with side information. 2017. Available at arXiv:1709.00112.

[33] Z. Chen, Z. Wang, and S. A. Jafar. The capacity of private information retrieval with private side information. 2017. Available at arXiv:1709.03022.

[34] Q. Wang and M. Skoglund. Secure private information retrieval from colluding databases with eavesdroppers. 2017. Available at arXiv:1710.01190.

[35] H. Sun and S. A. Jafar. The capacity of private computation. 2017. Available at arXiv:1710.11098.

[36] M. Kim, H. Yang, and J. Lee. Cache-aided private information retrieval. In *IEEE Asilomar*, October 2017.

[37] M. Mirmohseni and M. A. Maddah-Ali. Private function retrieval. 2017. Available at arXiv:1711.04677.

[38] M. Abdul-Wahid, F. Almoualem, D. Kumar, and R. Tandon. Private information retrieval from storage constrained databases – coded caching meets PIR. 2017. Available at arXiv:1711.05244.

[39] K. Banawan and S. Ulukus. Asymmetry hurts: Private information retrieval under asymmetric traffic constraints. 2018. Available at arXiv:1801.03079.

[40] Z. Chen, Z. Wang, and S. A. Jafar. The asymptotic capacity of private search. 2018. Available at arXiv:1801.05768.

[41] K. Banawan and S. Ulukus. Private information retrieval through wiretap channel II: Privacy meets security. 2018. Available at arXiv:1801.06171.

[42] Q. Wang, H. Sun, and M. Skoglund. The capacity of private information retrieval with eavesdroppers. 2018. Available at arXiv:1804.10189.

[43] M. A. Attia, D. Kumar, and R. Tandon. The capacity of private information retrieval from uncoded storage constrained databases. 2018. Available at arXiv:1805.04104.

[44] R. Tajeddine, O. W. Gnilke, D. Karpuk, R. Freij-Hollanti, and C. Hollanti. Private information retrieval from coded storage systems with colluding, Byzantine, and unresponsive servers. 2018. Available at arXiv:1806.08006.

[45] K. Banawan and S. Ulukus. Noisy private information retrieval: On separability of channel coding and information retrieval. 2018. Available at arXiv:1807.05997.

[46] Z. Jia, H. Sun, and S. A. Jafar. Cross subspace alignment and the asymptotic capacity of X-secure T-private information retrieval. 2018. Available at arXiv:1808.07457.

[47] C. Tian, H. Sun, and J. Chen. Capacity-achieving private information retrieval codes with optimal message size and upload cost. 2018. Available at arXiv:1808.07536.

[48] S. Kumar, A. G. i Amat, E. Rosnes, and L. Senigagliesi. Private information retrieval from a cellular network with caching at the edge. 2018. Available at arXiv:1809.00872.

[49] R. Bitar and S. El Rouayheb. Staircase-PIR: Universally robust private information retrieval. *Available at arXiv:1806.08825*, 2018.

[50] S. Li and M. Gastpar. Converse for multi-server single-message PIR with side information. 2018. Available at arXiv:1809.09861.

[51] R. G.L. D'Oliveira and S. El Rouayheb. One-shot PIR: Refinement and lifting. 2018. Available at arXiv:1810.05719.

[52] R. Tajeddine, A. Wachter-Zeh, and C. Hollanti. Private information retrieval over networks. 2018. Available at arXiv:1810.08941.

[53] M. A. Maddah-Ali and U. Niesen. Fundamental limits of caching. *IEEE Transactions on Information Theory*, 60(5):2856–2867, May 2014.

[54] M. A. Maddah-Ali and U. Niesen. Decentralized coded caching attains order-optimal memory-rate tradeoff. *IEEE/ACM Transactions on Networking*, 23(4):1029–1040, August 2015.

[55] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr. The exact rate-memory tradeoff for caching with uncoded prefetching. *IEEE Transactions on Information Theory*, 64(2):1281–1296, February 2018.

[56] M. Ji, G. Caire, and A. F. Molisch. Fundamental limits of caching in wireless D2D networks. *IEEE Transactions on Information Theory*, 62(2):849–869, February 2016.

[57] R. Pedarsani, M. A. Maddah-Ali, and U. Niesen. Online coded caching. *IEEE/ACM Transactions on Networking*, 24(2):836–845, April 2016.

[58] H. Ghasemi and A. Ramamoorthy. Improved lower bounds for coded caching. *IEEE Transactions on Information Theory*, 63(7):4388–4413, July 2017.

[59] K. Shanmugam, M. Ji, A. M. Tulino, J. Llorca, and A. G. Dimakis. Finite-length analysis of caching-aided coded multicasting. *IEEE Transactions on Information Theory*, 62(10):5524–5537, October 2016.

[60] A. Sengupta, R. Tandon, and T. C. Clancy. Fundamental limits of caching with secure delivery. *IEEE Transactions on Information Forensics and Security*, 10(2):355–370, February 2015.

[61] J. Zhang and P. Elia. Fundamental limits of cache-aided wireless BC: Interplay of coded-caching and CSIT feedback. *IEEE Transactions on Information Theory*, 63(5):3142–3160, May 2017.

[62] F. Xu, M. Tao, and K. Liu. Fundamental tradeoff between storage and latency in cache-aided wireless interference networks. *IEEE Transactions on Information Theory*, 63(11):7464–7491, November 2017.

[63] C. Tian and J. Chen. Caching and delivery via interference elimination. *IEEE Transactions on Information Theory*, 64(3):1548–1560, March 2018.

[64] S. S. Bidokhti, M. Wigger, and R. Timo. Noisy broadcast networks with receiver caching. *IEEE Transactions on Information Theory*, November 2018.

[65] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr. Characterizing the rate-memory tradeoff in cache networks within a factor of 2. *IEEE Transactions on Information Theory*, 2018.

[66] A. M. Ibrahim, A. A. Zewail, and A. Yener. Coded caching for heterogeneous systems: An optimization perspective. 2018. Available at arXiv:1810.08187.

[67] P. Hassanzadeh, A. M. Tulino, J. Llorca, and E. Erkip. Rate-memory trade-off for caching and delivery of correlated sources. 2018. Available at arXiv:1806.07333.

[68] K. Wan, D. Tuninetti, and P. Piantanida. On the optimality of uncoded cache placement. In *IEEE ITW*, September 2016.

[69] Q. Yang and D. Gündüz. Coded caching and content delivery with heterogeneous distortion requirements. *IEEE Transactions on Information Theory*, 64(6):4347–4364, June 2018.

[70] A. A. Zewail and A. Yener. Combination networks with or without secrecy constraints: The impact of caching relays. *IEEE Journal on Selected Areas in Communications*, June 2018.

[71] A. D. Wyner. The wire-tap channel. *Bell System Tech. J.*, 54(8):1355–1387, October 1975.

[72] I. Csiszár and J. Körner. Broadcast channels with confidential messages. *IEEE Transactions on Information Theory*, 24(3):339–348, May 1978.

[73] E. Arıkan. Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory*, 55(7):3051–3073, July 2009.

[74] E. Arıkan. Source polarization. In *IEEE ISIT*, June 2010.

[75] S. Korada and R. Urbanke. Polar codes are optimal for lossy source coding. *IEEE Transactions on Information Theory*, 56(4):1751–1768, April 2010.

[76] E. Şaşoğlu, İ. E. Telatar, and E. Yeh. Polar codes for the two-user multiple-access channel. *IEEE Transactions on Information Theory*, 59(10):6583–6592, October 2013.

[77] E. Abbe and İ. E. Telatar. Polar codes for the m-user multiple access channel. *IEEE Transactions on Information Theory*, 58(8):5437–5448, August 2012.

[78] S. Önay. Successive cancellation decoding of polar codes for the two-user binary-input MAC. In *IEEE ISIT*, July 2013.

[79] N. Goela, E. Abbe, and M. Gastpar. Polar codes for broadcast channels. *IEEE Transactions on Information Theory*, 61(2):758–782, 2015.

[80] M. Mondelli, S. H. Hassani, I. Sason, and R. Urbanke. Achieving Marton's region for broadcast channels using polar codes. *IEEE Transactions on Information Theory*, 61(2):783–800, November 2014.

[81] L. Wang and E. Şaşoğlu. Polar coding for interference networks. *http://arxiv.org/abs/1401.7293*, January 2014.

[82] E. Arıkan. Polar coding for the Slepian-Wolf problem based on monotone chain rules. In *IEEE ISIT*, July 2012.

[83] S. B. Korada. *Polar codes for channel and source coding*. PhD thesis, EPFL, May 2009.

[84] H. Mahdavifar and A. Vardy. Achieving the secrecy capacity of wiretap channels using polar codes. *IEEE Transactions on Information Theory*, 57(10):6428–6443, October 2011.

[85] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund. Nested polar codes for wiretap and relay channels. *IEEE Comm. Letters*, 14(8):752–754, August 2010.

[86] O. O. Koyluoglu and H. E. Gamal. Polar coding for secure transmission and key agreement. In *IEEE PIMRC*, September 2010.

[87] E. Hof and S. Shamai. Secrecy-achieving polar-coding. In *IEEE ITW*, August 2010.

[88] S. H. Hassani, S. Korada, and R. Urbanke. The compound capacity of polar codes. In *IEEE Allerton*, September 2009.

[89] I. Tal and A. Vardy. How to construct polar codes. *IEEE Transactions on Information Theory*, 59(10):6562–6582, October 2013.

[90] D. Sutter and J. M. Renes. Universal polar codes for more capable and less noisy channels and sources. In *IEEE ISIT*, June 2014.

[91] S. H. Hassani and R. Urbanke. Universal polar codes. In *IEEE ISIT*, June 2014.

[92] E. Şaşoğlu and L. Wang. Universal polarization. In *IEEE ISIT*, June 2014.

[93] J. Honda and H. Yamamoto. Polar coding without alphabet extension for asymmetric models. *IEEE Transactions on Information Theory*, 59(12):7829–7838, December 2013.

[94] J. M. Renes, R. Renner, and D. Sutter. Efficient one-way secret-key agreement and private channel coding via polarization. In *Advances in Cryptology-ASIACRYPT 2013*, pages 194–213. Springer, 2013.

[95] E. Tekin and A. Yener. The Gaussian multiple access wire-tap channel. *IEEE Transactions on Information Theory*, 54(12):5747–5755, December 2008.

[96] E. Tekin and A. Yener. The general Gaussian multiple-access and two-way wiretap channels: Achievable rates and cooperative jamming. *IEEE Transactions on Information Theory*, 54(6):2735–2751, June 2008.

[97] R. Liu, I. Maric, P. Spasojevic, and R. D. Yates. Discrete memoryless interference and broadcast channels with confidential messages: Secrecy rate regions. *IEEE Transactions on Information Theory*, 54(6):2493–2507, June 2008.

[98] T. C. Gulcu and A. Barg. Achieving secrecy capacity of the wiretap channel and broadcast channel with a confidential component. *IEEE Transactions on Information Theory*, 63(2):1311–1324, 2017.

[99] R. A. Chou and M. R. Bloch. Polar coding for the broadcast channel with confidential messages: A random binning analogy. *IEEE Transactions on Information Theory*, 62(5):2410–2429, 2016.

[100] E. Şaşoğlu and A. Vardy. A new polar coding scheme for strong security on wiretap channels. In *IEEE ISIT*, July 2013.

[101] E. Ekrem and S. Ulukus. On the secrecy of multiple access wiretap channel. In *IEEE Allerton*, September 2008.

[102] E. Ekrem and S. Ulukus. Cooperative secrecy in wireless communications. In *Securing Wireless Communications at the Physical Layer*, pages 143–172. W. Trappe and R. Liu, Eds. Springer, 2009.

[103] R. Bassily and S. Ulukus. Ergodic secret alignment. *IEEE Transactions on Information Theory*, 58(3):1594–1611, March 2012.

[104] M. Ye and A. Barg. Polar codes for distributed hierarchical source coding. *Advances in Mathematics of Communications*, 9(1):87–103, February 2015.

[105] O. Ozel and S. Ulukus. Wiretap channels: Implications of the more capable condition and cyclic shift symmetry. *IEEE Transactions on Information Theory*, 59(4):2153–2164, April 2013.

[106] E. Şaşoğlu and İ.E. Telatar. Polarization for arbitrary discrete memoryless channels. In *IEEE ITW*, October 2009.

[107] R. Mori and T. Tanaka. Channel polarization on q-ary discrete memoryless channels by arbitrary kernel. In *IEEE ISIT*, June 2010.

[108] E. Şaşoğlu. Polar codes for discrete alphabets. In *IEEE ISIT*, July 2012.

[109] W. Park and A. Barg. Polar codes for q-ary channels, $q = 2^r$. *IEEE Transactions on Information Theory*, 59(2):955–969, February 2013.

[110] L. Wang, E. Şaşoğlu, and Y.-H. Kim. Sliding-window superposition coding for interference networks. In *IEEE ISIT*, June 2014.

[111] Y.-P. Wei, K. Banawan, and S. Ulukus. Cache-aided private information retrieval with unknown and uncoded prefetching. In *IEEE ISIT*, June 2018.

[112] Y.-P. Wei, K. Banawan, and S. Ulukus. Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching. *IEEE Transactions on Information Theory*. To appear. Also available at arXiv:1709.01056.

[113] Y.-P. Wei, K. Banawan, and S. Ulukus. Cache-aided private information retrieval with partially known uncoded prefetching. In *IEEE ICC*, May 2018.

[114] Y.-P. Wei, K. Banawan, and S. Ulukus. Cache-aided private information retrieval with partially known uncoded prefetching: Fundamental limits. *IEEE Jour. on Selected Areas in Communications*, 36(6):1126–1139, June 2018.

[115] Y.-P. Wei, K. Banawan, and S. Ulukus. Private information retrieval with partially known private side information. In *IEEE CISS*, March 2018.

[116] Y.-P. Wei, K. Banawan, and S. Ulukus. The capacity of private information retrieval with partially known private side information. 2017. Available at arXiv:1710.00809.

[117] Y.-P. Wei and S. Ulukus. Private information retrieval with private side information under storage constraints. In *IEEE ITW*, November 2018.

[118] Y.-P. Wei and S. Ulukus. The capacity of private information retrieval with private side information under storage constraints. 2018. Available at arXiv:1806.01253.

[119] Y.-P. Wei, B. Arasli, K. Banawan, and S. Ulukus. The capacity of private information retrieval from decentralized uncoded caching databases. 2018. Available at arXiv:1811.11160.

[120] Y.-P. Wei and S. Ulukus. Polar coding for the general wiretap channel. In *IEEE ITW*, April 2015.

[121] Y.-P. Wei and S. Ulukus. Polar coding for the general wiretap channel with extensions to multiuser scenarios. *IEEE Journal on Selected Areas in Communications*, 34(2):278–291, 2016.