ABSTRACT

Title of dissertation: PRIVATE INFORMATION RETRIEVAL AND SECURITY IN NETWORKS

Karim Banawan, Doctor of Philosophy, 2018

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

This dissertation focuses on privacy and security issues in networks from an information-theoretic perspective. Protecting privacy requires protecting the identity of the desired message from the data source. This is highly desirable in next-generation networks, where data-mining techniques are present everywhere. Ensuring security requires that the data content is not interpretable by non-authorized nodes. This is critical in wireless networks, which are inherently open.

We first focus on the privacy issue via the private information retrieval (PIR) problem. PIR is a canonical problem to study the privacy of the downloaded content from public databases. In PIR, a user wishes to retrieve a file from distributed databases, in such a way that no database can know the identity of the user's desired file. PIR schemes need to be designed to be more efficient than the trivial scheme of downloading all the files stored in the databases. Fundamentally, PIR lies at the intersection of computer science, information theory, coding theory, and signal processing.

The classical PIR formulation makes the following assumptions: The con-

tent is *exactly replicated* across the databases; the user wishes to retrieve a *single file* privately; the databases do *not collude*; the databases answer the user queries *truthfully*; the database answers go through *noiseless orthogonal* channels; there are *no external security* threats; and the answer strings have unconstrained *symmetric* lengths. These assumptions are too idealistic to be practical in modern systems. In this thesis, we introduce extended versions of the classical PIR problem to be relevant to modern applications, namely: PIR from coded databases, multi-message PIR, PIR from colluding and Byzantine databases, PIR under asymmetric traffic constraints, noisy PIR, and PIR from wiretap channel II. We characterize the *fundamental limits* of such problems from an information-theoretic point of view. This involves two parts for each setting: first, we devise a practical scheme that retrieves the desired file(s) correctly and privately; second, we mathematically prove that no other retrieval scheme can achieve any higher PIR rate than the proposed scheme. The optimal PIR rate is called the *PIR capacity* reminiscent of the *capacity* of communication channels.

First, we consider *PIR from MDS-coded databases*. Due to node failures and erasures that arise naturally in any storage system, redundancy should be introduced. However, replicating the content across the databases incurs high storage cost. This motivates the content of the databases to be coded instead of merely being replicated. We investigate the PIR problem from MDS-coded databases. We determine the optimal retrieval scheme for this problem, and characterize the exact PIR capacity. The result implies a fundamental tradeoff between the retrieval cost and the storage cost.

Second, we consider *multi-message PIR*. In this problem, the user is interested in retrieving multiple files from the databases without revealing the identities of these files. We show that multiple messages can be retrieved more efficiently than retrieving them one-by-one in a sequence. When the user wishes to retrieve at least half of the files stored in the databases, we characterize the exact capacity of the problem by proposing a novel scheme that downloads MDS-coded mixtures of all messages. For all other cases, we develop a near-optimal scheme which is optimal if the ratio between the total number of files and the number of desired files is an integer.

Third, we consider *PIR from colluding and Byzantine databases*. In this problem, a subset of the databases, called Byzantine databases, can return arbitrarily corrupted answers. In addition, a subset of the databases can collude by exchanging user queries. The errors introduced by the Byzantine databases can be unintentional (if databases store outdated message sets), or even worse, can be intentional (as in the case of maliciously controlled databases). We propose a Byzantine and collusion resilient retrieval scheme, and determine the exact PIR capacity for this problem. The capacity expression reveals that the effect of the Byzantine databases is equivalent to removing twice the number of Byzantine databases from the system.

Fourth, we consider *PIR under asymmetric traffic constraints*. A common property of the schemes constructed for the existing PIR settings is that they exhibit a symmetric structure across the databases. In practice, this may be infeasible, for instance, when the links from the databases have different capacities. To that end, we develop a novel upper bound for the PIR capacity that incorporates the traffic

asymmetry. We propose explicit achievability schemes for specific traffic ratios. For any other traffic ratio, we employ time-sharing. Our results show that asymmetry fundamentally hurts the retrieval rate.

Fifth, we consider *noisy PIR*, where the returned answers reach the user via noisy channel(s). This is motivated by practical applications, such as, random packet dropping, random packet corruption, and PIR over wireless networks. We consider two variations of the problem, namely: noisy PIR with orthogonal links, and PIR from multiple access channels. For noisy PIR with orthogonal links, we show that channel coding and retrieval scheme are *almost separable* in the sense that the noisy channels affect only the traffic ratios. For the PIR problem from multiple access channels, the output of the channel is a mixture of all the answers returned by the databases. In this case, we show explicit examples, where the channel coding and the retrieval scheme are *inseparable*, and the privacy may be achieved for free.

Sixth, we consider *PIR from wiretap channel II*. In this problem, there is an external eavesdropper who wishes to learn the contents of the databases by observing portions of the traffic exchanged between the user and the databases during the PIR process. The databases must (information theoretically) encrypt their responses such that the eavesdropper learns nothing from its observation. We design a retrieval code that satisfies the combined privacy and security constraints. We show the necessity of using asymmetric retrieval schemes which build on our work on PIR under asymmetric traffic constraints.

Next, we focus on the security problem in multi-user networks via physical layer techniques. Physical layer security enables secure transmission of information

without need for encryption keys, thereby mitigating the problems associated with exchanging encryption keys across open wireless networks. Existing work in physical layer security makes the following assumptions: All nodes are *altruistic* and follow a prescribed transmission policy to maximize the secure rate of the entire system; the channel inputs to Gaussian channels are constrained by a total *transmitter-side* power constraint; and in the secure degrees of freedom studies for the interference channel, users have a *single antenna* each. We address these issues by investigating the MIMO interference channel with confidential messages, security in networks with user misbehavior, and MIMO wiretap channel under receiver-side power constraints. We characterize the optimal secure transmission strategies in terms of the *secrecy capacity* and its high-SNR approximation, the secure degrees of freedom (s.d.o.f.).

First, we determine the exact s.d.o.f. region of the two-user *MIMO interference channel* with confidential messages (ICCM). To that end, we propose a novel achievable scheme for the 2×2 ICCM system, which is a building block for any other ICCM system. We show that the s.d.o.f. region starts as a square region, then it takes the shape of an irregular polytope until it returns back to a square region when the number of transmit antennas is at least twice the number of receiving antennas.

Second, we investigate the security problem in the presence of *user misbehavior*. We consider the following multi-user scenarios: Multiple access wiretap channel with deviating users who do not follow agreed-upon optimum protocols, where we quantify the effect of user deviations and propose counter-strategies for the honest users; the broadcast channel with confidential messages in the presence of combating

helpers, where we show that the malicious intentions of the helpers are neutralized and the full s.d.o.f. is retained; and interference channel with confidential messages when the users are selfish and have conflicting interests, where we show that selfishness precludes secure communication and no s.d.o.f. is achieved.

Third, we consider the MIMO wiretap channel with a *receiver-side minimum power constraint* in addition to the usual transmitter-side maximum power constraint. This problem is motivated by energy harvesting communications with wireless energy transfer, where an added goal is to deliver a minimum amount of energy to a receiver in addition to delivering secure data to another receiver. We prove that the problem is equivalent to solving a secrecy capacity problem with a *double-sided correlation matrix constraint* on the channel input. We extend the channel enhancement technique to our setting. We propose two optimum schemes that achieve the optimum rate: Gaussian signaling with a fixed mean and Gaussian signaling with Gaussian artificial noise. We extend our techniques to other related multi-user settings.

Private Information Retrieval and Security in Networks

by

Karim Banawan

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 Charalampos Papamanthou
Professor Adrian Papamarcou
Professor Lawrence C. Washington

Dedication

To my family.

Acknowledgments

First and foremost, all praise is due to Allah, the lord of the worlds, the all-knowing, the most merciful, and the most generous. I express my sincere gratitude to Allah for his countless blessings. His grace provided me with strength, purpose, and patience to reach this stage of my life.

I gratefully thank my advisor Prof. Sennur Ulukus for her support and guidance throughout my Ph.D. program. Her dedication to her students and passion for research is truly impressive. She genuinely cares about the future of her students. On an academic level, I learned from her how to formulate a meaningful and interesting research problems, how to present ideas in a simple yet elegant fashion whether in an oral presentation, a lecture, or on a written document, how to intuitively assess the soundness of an argument, and how to maintain high quality research program. On a personal level, I learned from her the importance of hard work, dedication, and giving your best at every situation. Her door was always open to me to discuss any ideas that occur to me without showing any sign of boredom. I learned a lot from our discussions, whether in our one-to-one meeting or in a group meeting.

I would like to thank Professors Alexander Barg, Charalampos Papamanthou, Adrian Papamarcou, and Lawrence Washington, for being in my dissertation committee and for their insightful discussions. I am thankful to all the professors I engaged with, either through courses or personally, over the past five years at UMD. I specifically want to thank Prof. Andre Tits, whom I was honored to co-teach the

signals and systems course with him. I learned a lot from him either technically or personally. I also enjoyed the classes of Professors Adrian Papamarcou, Prakash Narayan, Piya Pal, and Leonid Koralov. These courses were challenging yet very interesting to me.

I am thankful to all my colleagues at UMD. I would like to specifically thank Ahmed Arafa, Abdulrahman Baknina, and Yi-Peng Wei. They were great companions of this long journey. Ahmed always provided a great help in every stage of Ph.D. and on a personal level. Abdulrahman was my office mate in Alexandria and in Maryland for almost 8 years, in which he was a dependable ally. I, Abdulrahman and Yi-Peng started the program together and we had a great time together. I have many interesting discussions with Yi-Peng which result in many joint research works. I would also like to thank all CSPL lab mates: Jianwei Xie, Omur Ozel, Berk Gurakan, Pritam Mukherjee, Praneeth Boda, Ajaykrishnan Nageswaran, Baturalp Buyukates, Melih Bastopcu, Brian Kim, Batuhan Arasli, and Bibhusa Rawal. Special thanks to Pritam Mukherjee, whom I had many fruitful discussions with about physical layer security.

I must express my profound gratitude to my family. Their constant support is endless. I would not reach this far without their guidance and prayers. Their endurance of my absence was a true sacrifice that I cannot repay.

I am deeply grateful to my wife and friend, Eman. Saying that she was supportive would be an understatement. She never lost faith in me even in my darkest hours. Eman took care of countless responsibilities, that could distract me from my work, willingly and with sincere love. In addition, our daily discussions about

nearly everything provide me with warmth and comfort. She sacrificed a lot for me and regarded this Ph.D. degree as a project of her own. I would always be indebted to her. I am thankful for my son, Noureldeen, who brought joy and a new purpose to my life. He always manages to make me smile even on a gloomy day. I wholeheartedly wish I can be a good father for him.

I am thankful to all my friends in Maryland and in particular the Arab and Egyptian community. They give me a sense of unity and belonging. They were always there for me when I needed them. Our gathering reduced my home-sickness and give me a pleasant sense of community.

Finally, I would like to thank the ECE staff members, who ease the Ph.D. experience for all the students by their utmost dedication. I would thank Melanie Prange in particular, who offered her sincere help in many occasions, even before laying a foot in UMD.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

Privacy and security are challenging, yet crucial issues in the design of next generation networks. Preserving *privacy* entails protecting the *identity* of the desired messages (files) from the content generator (e.g., a database). This is highly relevant in the era of big data, where efficient data-mining techniques are present everywhere, from social media to online-shopping to search history. On the other hand, ensuring *security* entails guaranteeing that the *data contents* are not interpretable by non-authorized nodes (e.g., external eavesdroppers). This is particularly vital in wireless networks, where the openness of the wireless medium imposes a security risk on the wireless transmission. Although the privacy and security problems are seemingly different, they share a deeper connection. Both problems require the user/transmitter to create a form of *confusion* (i.e., obfuscation) in the queries/messages to satisfy privacy/security constraints. In this dissertation, we investigate the privacy and security problems through the lens of information theory. Our goal is to characterize the fundamental limits of retrieval/communication in networks subject to various practical considerations and devise optimal schemes to achieve such limits.

In Chapters 2-6, we focus solely on the privacy problem. Protecting the privacy of downloaded information from curious publicly accessible databases has been the focus of considerable research within the computer science community [1–5]. Private information retrieval (PIR), introduced by Chor et al. [1], is a canonical problem to study the privacy of the downloaded content from public databases. In the classical PIR setting, a user requests to download a certain message (or file) out of M distinct messages from N non-communicating (non-colluding) databases without leaking the identity of the message to any individual database. The contents of these databases are identical. To that end, the user prepares N queries, one for each database, such that the queries do not reveal the user's interest in the desired message. Upon receiving these queries, each database responds truthfully with an answering string. The user needs to be able to reconstruct the entire message by decoding the answer strings from all databases. A straightforward solution for this seemingly challenging task is to download all of the contents of the databases. However, this solution is highly impractical, in particular for large number of messages which is the case in modern storage systems. The aim of the PIR problem is to design efficient retrieval schemes. The efficiency of PIR systems is assessed by the PIR rate, which is the ratio between the desired message size and the total downloaded symbols. Many practical applications are related to PIR, such as: protecting the identity of stock market records reviewed by an investor, as showing interest in a specific record may affect its value; ensuring the privacy of an inventor as they look up existing patents in a database, since revealing what they are looking at leaks some information about the current invention they are working on; and protecting the

nature of content browsed by activists on the internet in oppressive regimes. From a technical standpoint, PIR lies at the intersection of computer science, information theory, coding theory, and signal processing.

In the original formulation of the problem in the computer science literature [1], the messages are assumed to have a size of one bit. The computer science formulation considers optimizing two performance metrics, namely, the download cost, which is the sum of the lengths of the answer strings, and the upload cost, which is the sum of the lengths of the queries. Most of this work adopts computational guarantees for the privacy constraint, where it is assumed that the databases cannot infer any information about the identity of the desired message unless they solve certain computationally hard problems [3, 5]. Recently, there has been a growing interest in the PIR problem in the information-theory society, with early examples [6–11]. The information-theoretic reformulation of the problem assumes that the messages are of arbitrarily large size and hence the upload cost can be neglected with respect to the download cost [8]. This formulation provides an absolute, i.e., information-theoretic, guarantee that no server participating in the protocol gets any information about the user intent irrespective of their computational powers.

In the pioneering paper [12], Sun and Jafar introduce the *PIR capacity* notion to characterize the fundamental limits of the PIR problem. The PIR capacity is defined as the supremum of PIR rates over all achievable retrieval schemes (optimal retrieval rate) reminiscent of the capacity of communication channels. [12] determines the exact capacity of the classical PIR model to be $C = (1+\frac{1}{N}+\frac{1}{N^2}+\cdots+\frac{1}{N^{M-1}})^{-1} =$

$\frac{1-\frac{1}{N}}{1-(\frac{1}{N})^M}$. The achievability scheme is a greedy algorithm that employs a symmetric query structure, which is based on three principles: message symmetry, database symmetry, and exploitation of side information. The achievable scheme hinges on an interesting relationship between PIR and blind interference alignment introduced for wireless networks in [13] as observed in [11].

Following [12], the fundamental limits of many interesting variants of the classical PIR problem have been considered, such as: PIR with T colluding databases (TPIR) [14, 15], where any T of N databases might collude; robust PIR (RPIR) [14, 16], where some databases may fail to respond; symmetric PIR (SPIR) [17], which adds the constraint that the user should learn only the desired message; 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 [19]; MDS-coded symmetric PIR [20]; symmetric PIR from Byzantine databases [21]; MDS-coded PIR with colluding databases [22–24], and its multi-message [25], Byzantine [26], and symmetric [27] versions; cache-aided PIR where additional side information is present [28–35]; private computation [36, 37], where the user is interested in retrieving a function of the database contents as opposed to direct database content; private search [38], where the user searches for all records that match a privately chosen value without revealing the chosen value; PIR from storage constrained databases [39, 40], where each database stores a fraction of the messages instead of the complete copy of the content; secure PIR [41, 42], where E of the databases are captured and observed by an external eavesdropper and its sym-

metric version [21]; PIR from secure distributed storage [43, 44], where the contents of the databases need to be secured against X database collusion.

The classical PIR model imposes the following assumptions: First, it assumes that the content is *exactly replicated* across the databases. Second, it assumes that the user wishes to retrieve only a *single file* privately. Third, it assumes that the databases do *not collude* and answer the user queries *truthfully*. Fourth, it assumes that the database answers are received through *noiseless orthogonal* bit pipes (channels). Fifth, it *ignores the security* of the retrieved bits against external eavesdroppers. Sixth, the answer strings have unconstrained lengths, which typically exhibit a *symmetric structure* in most known PIR schemes. These assumptions are too idealistic from a practical point of view. Consequently, in this thesis, we introduce and investigate practically relevant extensions of the classical PIR problem, namely: PIR from coded databases, multi-message PIR, PIR from colluding and Byzantine databases, PIR under asymmetric traffic constraints, noisy PIR, and PIR from wiretap channel II. We aim at characterizing the *fundamental limits* of such problems from an information-theoretic point of view. This task is two-fold, first, one should devise a practical scheme that retrieves the desired file(s) correctly and privately, then, one should mathematically prove that no other retrieval scheme can achieve any higher rate than the proposed scheme by constructing a matching converse.

In Chapter 2, we consider the problem of *MDS-coded PIR (CPIR)*. Due to node failures and erasures that arise naturally in any storage system, redundancy should be introduced [45]. The simplest form of redundancy is repetition coding. Although repetition coding across databases offers the highest immunity against erasures and

simplicity in designing PIR schemes, it results in extremely large storage cost. This motivates the use of erasure coding techniques that achieve the same level of reliability with less storage cost. A common erasure coding technique is the MDS code that achieves the optimal redundancy-reliability tradeoff. An (N, K) MDS code maps K sub-packets of data into N sub-packets of coded data. This code tolerates upto $N-K$ node failures (or erasures). Despite the ubiquity of work on the classical PIR problem, little research exists for coded PIR with a few exceptions: [6] which has initiated the work on coded databases and has designed an explicit erasure code and PIR algorithm that requires only one extra bit of download to provide perfect privacy. The result is achieved at the expense of having the number of storage nodes N grow with the message size. [8] considers a general formulation for the coded PIR problem, and obtains a tradeoff between storage and retrieval costs based on certain sufficient conditions. [10] presents the best-known achievable scheme for the MDS-coded PIR problem, which achieves a retrieval rate of $R = 1 - R_c$, where R_c is the code rate of the storage system. The scheme is universal in that it depends only on the code rate. Finally, [46] investigates the problem from the storage overhead perspective and shows that information-theoretic PIR can be achieved with storage overhead arbitrarily close to the optimal value of 1 by proposing new binary linear codes called the k-server PIR codes.

In this chapter, we consider the PIR problem for non-colluding and MDS-coded databases. We assume that the contents of the databases are coded using an (N, K) MDS storage code. This formulation includes the models of [12] and [10] as special cases. We show that the exact PIR capacity in this case is given by

$C = \left(1 + \frac{K}{N} + \frac{K^2}{N^2} + \cdots + \frac{K^{M-1}}{N^{M-1}}\right)^{-1} = (1 + R_c + R_c^2 + \cdots + R_c^{M-1})^{-1} = \frac{1-R_c}{1-R_c^M}$. The PIR capacity depends only on the code rate R_c and the number of messages M irrespective of the generator matrix structure or the number of nodes. Surprisingly, the result implies the optimality of separation between the design of the PIR scheme and the MDS storage code for a fixed code rate. The result outperforms the best-known lower bound in [10], and reduces to the repetition-coded case (which is a special case of MDS codes) in [12] by observing that $R_c = \frac{1}{N}$ in that case. The achievable scheme is similar to the scheme in [12] with extra steps that entail decoding of the interference and the desired message by solving K linearly independent equations. The converse proof hinges on the fact that the contents of any K storage nodes are independent and hence the answer strings in turn are independent. We present two lemmas that capture the essence of the converse proof, namely, interference lower bound lemma and induction lemma. The proof of the induction lemma uses Han's inequality to lower bound the entropy of any K answer strings. These lemmas generalize the converse technique in [12, Lemmas 5, 6] to account for MDS coding.

In Chapter 3, we consider the problem of *multi-message PIR (MPIR)*. In some applications, the user may be interested in retrieving multiple messages from the databases without revealing the identities of these messages. Returning to the examples presented earlier: The investor may be interested in comparing the values of multiple records at the same time, and the inventor may be looking up several patents that are closely related to their work. One possible solution to this problem is to use single-message retrieval scheme in [12] successively. We show in this work

that multiple messages can be retrieved more efficiently than retrieving them one-by-one in a sequence. This resembles superiority of joint decoding in multiple access channels over multiple simultaneous single-user transmissions [47]. A few works exist in MPIR in the computer science literature, such as: Reference [48] proposes a multi-block (multi-message) scheme and observes that if the user requests multiple blocks (messages), it is possible to reuse randomly mixed data blocks (answer strings) across multiple requests (queries). Reference [49] develops a multi-block scheme which further reduces the communication overhead. An achievable scheme for the multi-block PIR by designing k-safe binary matrices that uses XOR operations is developed in [50]. Reference [50] extends the scheme in [1] to multiple blocks. Reference [51] designs an efficient non-trivial multi-query computational PIR protocol and gives a lower bound on the communication of any multi-query information retrieval protocol. Reference [52] suggests using batch codes to allow a single client to retrieve multiple records simultaneously while allowing the server computation to scale sublinearly with the number of records fetched. This idea is extended further in [53] to design a PIR server algorithm that achieves sublinear scaling in the number of records fetched, even when they are requested by distinct, non-collaborating clients. These works do not consider determining the information-theoretic capacity.

In this chapter, we formulate the MPIR problem with non-colluding replicated databases. Our goal is to characterize the sum capacity of the MPIR problem C_s^P, which is defined as the maximum ratio of the number of retrieved symbols from the P desired messages to the number of total downloaded symbols. When the number of desired messages P is at least half of the total number of messages M,

i.e., $P \geq \frac{M}{2}$, we determine the exact sum capacity of MPIR as $C_s^P = \frac{1}{1+\frac{M-P}{PN}}$. We use a novel achievable scheme which downloads MDS-coded mixtures of all messages. For the case of $P \leq \frac{M}{2}$, we derive lower and upper bounds that match if the total number of messages M is an integer multiple of the number of desired messages P, i.e., $\frac{M}{P} \in \mathbb{N}$. In this case, the sum capacity is $C_s^P = \frac{1-\frac{1}{N}}{1-(\frac{1}{N})^{M/P}}$. The result resembles the single-message capacity with $\frac{M}{P}$ messages. In other cases, we show numerically that the gap between the lower and upper bounds is monotonically decreasing in N and is upper bounded by 0.0082. The achievable scheme when $P \leq \frac{M}{2}$ is inspired by the greedy algorithm in [12], which retrieves all possible combinations of messages. The main difference of our scheme from the scheme in [12] is the number of stages required in each download round. Interestingly, the number of stages for each round is related to the output of a P-order IIR filter [54]. This intriguing connection to IIR filtering is a result of constructing the scheme in [12] *backwards* and observing the required side information needed in previous rounds. Our converse proof generalizes the proof in [12] for $P \geq 1$. The essence of the proof is captured in two lemmas: the first lemma lower bounds the uncertainty of the interference for the case $P \geq \frac{M}{2}$, and the second lemma upper bounds the remaining uncertainty after conditioning on P interfering messages.

In Chapter 4, we consider the problem of *PIR from Byzantine databases (BPIR)*. A common assumption in the literature is that the databases respond truthfully with the correct answer strings. Using the correct answers, the user can use the undesired symbols downloaded from one database as side information at other databases, and distribute the requests for the desired symbols among the

N databases. In this chapter, we investigate how we can reconstruct the desired message even if B databases (called Byzantine databases) respond with incorrect answer strings. Returning to the examples presented earlier: The databases storing the stock market records may not be updated simultaneously, therefore some of the databases may store outdated versions of the messages and can introduce unintentional errors to the answering strings. This scenario is referred to in the literature as the *unsynchronized PIR* problem [7]. For the oppressive regime example, some databases may be controlled by the regime, and these databases may return incorrect answer strings on purpose to confuse the user. This scenario is referred to in the literature as the *PIR with adversarial databases* problem [55, 56]. In both cases, the user needs to be able to reconstruct the desired message with no error, irrespective of the actions performed by the Byzantine databases. The BPIR problem was introduced in [55], which proposes a generic transformation from schemes of RPIR to robust protocols that tolerate Byzantine servers, and gives an explicit Byzantine robust scheme when $B \leq T \leq \frac{N}{3}$. [57] presents a fault-tolerant PIR scheme that can cope with malicious failures for $B \leq T \leq \frac{N}{2}$. [56] observes that allowing for list decoding instead of unique decoding enlarges the feasible set up to $B < N - T - 1$. Their achievable scheme allows for a small failure probability. The scheme depends on Shamir's secret sharing algorithm [58] and Guruswami-Sudan decoding algorithm [59]. The unsynchronized PIR problem is investigated in [7], where they propose a two-round retrieval scheme. The scheme returns the desired record by first identifying which records are mis-synchronized, and then by constructing a PIR scheme that avoids these problematic records.

In this chapter, we consider the single-round BPIR problem from N replicated databases in the presence of B Byzantine databases. The remaining storage nodes store the exact copy of the message set, and respond truthfully with the correct answer strings. We consider the T-privacy constraint, which permits colluding between any T databases to exchange the queries submitted by the user. Our goal is to characterize the single-round capacity of the BPIR problem under the zero-error reliability constraint and the T-privacy constraint. To that end, we propose an achievable scheme that is resilient to the worst-case errors that result from the Byzantine databases. Our achievable scheme extends the optimal scheme for the RPIR problem [14] to correct the *errors* resulted from the Byzantine databases, in contrast to the *erasures* introduced by the unresponsive databases in RPIR. The new ingredients to the achievable scheme are: encoding the undesired symbols via a punctured MDS code, successive interference cancellation of the side information, and encoding the desired symbols by an outer-layer MDS code. For the converse, we extend the converse arguments developed for the network coding problem in [60] and distributed storage systems in [61] to the PIR problem. This cut-set upper bound can be thought of as a network version of the Singleton bound [62]. We determine the exact capacity of the BPIR problem to be $C = \frac{N-2B}{N} \cdot \frac{1-\frac{T}{N-2B}}{1-\left(\frac{T}{N-2B}\right)^M}$, if $2B+T < N$. The capacity expression shows the severe degradation of the retrieval rate due to the presence of Byzantine databases. The capacity expression is equivalent the TPIR capacity with $N-2B$ databases with a multiplicative factor of $\frac{N-2B}{N}$, which signifies the ignorance of the user as to which $N-2B$ databases are honest. Our formulation includes the special case of the single-round unsynchronized PIR problem, if the

user has no knowledge about the number of mis-synchronized messages, and only knows that the entirety of some B databases may be unsynchronized in contrast to [7]. Under our assumptions, the single-round capacity of the unsynchronized PIR problem and the BPIR problem are the same.

In Chapter 5, we introduce the problem of *PIR under asymmetric traffic constraints*. A common property of the achievability schemes in the PIR literature is that they exhibit a symmetric structure across the databases. This enables the user to balance the load of retrieval of the desired message equally among the databases, and re-use the side information generated from one database equally in all the remaining databases. Now, consider the following scenarios that render symmetry assumption unworkable: *Varying database availability:* Certain databases are available only a fraction of the time other databases are available for downloads. *Different capacities:* The capacities of the links (bit pipes) from the databases to the user have different capacities. This may be due to different physical locations of the databases, e.g., the user may be able to access physically closer databases more often than physically distant databases, or it may be due to the quality of the physical layer communication channel, e.g., the bandwidths (rates) of the download channels may be different for different databases. In these cases, the user is forced to deal with each database differently, i.e., the user should utilize the databases which have better quality links more often than the other databases. This breaks the database symmetry assumption, makes load balancing of desired message and side information more challenging, and poses the following interesting questions: Can we perform efficient PIR without applying database symmetry? Is there a fundamental

PIR rate loss due to not being able to use symmetric schemes? Motivated by these practical scenarios, we consider the classical PIR problem under *asymmetric traffic constraints*. Formally, we assume that the nth database responds with a t_n-length answer string. We constrain the lengths of the answer strings such that $t_n = \lambda_n t_1$ for $n \in \{2, \cdots, N\}$. This, in turn, forces the ratios between the traffic from the databases to be $1 : \lambda_2 : \lambda_3 : \cdots : \lambda_N$. We denote the traffic ratio with respect to the total download by a vector $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_N)$, where $\tau_n = \frac{\lambda_n}{\sum_{j=1}^{N} \lambda_j}$. We aim at characterizing the capacity of this PIR problem, $C(\boldsymbol{\tau})$, as a function of the given traffic ratio vector $\boldsymbol{\tau}$ for arbitrary M and N.

In this chapter, we investigate the fundamental limits of the PIR problem under asymmetric traffic constraints. To that end, we develop a novel upper bound for the capacity $\bar{C}(\boldsymbol{\tau})$. This generalizes the converse proof of [12] to incorporate the asymmetric traffic constraints. Originally, the proof in [12] exploits the database symmetry. The rationale is that even if the optimal scheme is not symmetric, we can transform it into a symmetric scheme without changing the retrieval rate by means of time-sharing [12]. In our case, we cannot use this technique as we must deal with the databases differently. We characterize the upper bound as a piece-wise affine function in $\boldsymbol{\tau}$. The upper bound implies that asymmetry fundamentally hurts the retrieval rate. Then, we propose explicit achievability schemes for $\binom{M+N-1}{M}$ corner points. Each corner point corresponds to a specific partitioning of the databases according to the number of side information symbols that are used simultaneously within the initial round of the download. We describe the achievability scheme via a system of difference equations in the number of stages at each round of the

download (in parallel to Chapter 3). For any other traffic ratio vector $\boldsymbol{\tau}$, we employ time-sharing between the corner points that enclose $\boldsymbol{\tau}$. We provide an explicit rate expression for the case of $N = 2$ for arbitrary M. We show that the upper bound and the lower bound exactly match for the cases of $M = 2$ and $M = 3$ messages for any N and any $\boldsymbol{\tau}$, leading to the exact capacity $C(\boldsymbol{\tau})$ for these cases.

In Chapter 6, we introduce the problems of *noisy PIR with orthogonal links (NPIR)* and *PIR from multiple access channels (MAC-PIR)*. In all previous works, the links from the databases to the user are assumed to be noiseless. Further, these works assume that the answer strings are returned via orthogonal links, i.e., the user receives N separate answer strings, which are not mixed. There are many practical settings where these assumptions may not be valid. For instance, while browsing the internet, some packets may be dropped randomly. This scenario can be abstracted out as passing the answer strings through an erasure channel. Alternatively, the data packets may be randomly corrupted, which can be modeled as a binary symmetric channel that flips randomly some symbols in the answer strings. Hence, a more realistic retrieval model may be to assume that the databases return their answer strings through noisy channels with known transition probabilities. Yet, in other applications, the answer strings may be mixed before reaching the user. For example: if the user is retrieving the desired file from wireless base stations, the answer strings would be combined in the air before reaching the user. Another example is retrieval from a cloud, where the returned packets may collide and superimpose each other. These practical settings can be represented with another abstract model, which is the cooperative multiple access channel (MAC) model, where the databases cooperate to

convey the desired message to the user, while the user receives a stochastic mapping from the database responses in general. These two cases pose many interesting questions, such as: How can we devise schemes that mitigate the errors introduced by the channel with a small sacrifice from the private retrieval rate? Is there a separation between the channel coding needed for reliable transmission over noisy channels and the private retrieval scheme, or is there a necessity for joint processing? How do the statistical properties of the noisy channels fundamentally affect the private retrieval rate?

In this chapter, we first focus on the NPIR problem. In NPIR, the nth database is connected to the user via a discrete memoryless channel with known transition probability distribution $p(y_n|x_n)$. Intuitively, since a channel with worse channel condition needs a lower code rate to combat the channel errors, we do not expect the lengths of the answer strings to be the same from all the databases. Therefore, in this work, we allow the traffic from each database to be asymmetric as in Chapter 5. We first derive a general upper bound for the retrieval rate in the form of a max-min problem. The converse proof is inspired by the converse proof in Chapter 5, in particular in the way the asymmetry is handled. We show the achievability proof by random coding arguments and enforcing the uncoded responses to operate at one of the corner points of the PIR problem under asymmetric traffic constraints. The upper and lower bounds match for $M = 2$ and $M = 3$ messages, for arbitrary N databases, and any noisy channel. Our results show that the channel coding needed to mitigate the channel errors and the retrieval scheme are *almost separable* in the sense that the noisy channels affect only the traffic ratio requested from each

database and not the explicit coding technique. Interestingly, the upper and lower bounds depend only on the capacity of the noisy channels and not on the explicit transition probability of the channels. Next, we consider the MAC-PIR problem, where the responses of the databases reach the user through a discrete memoryless MAC with a known transition probability $p(y|x_1, \cdots, x_N)$. In this case, the output of the channel is a mixture (possibly noisy mixture) of all database responses. Interestingly, for this model, we show that channel coding and retrieval strategy are *inseparable* unlike in the NPIR problem. We show this fact by deriving the PIR capacity of two simple MACs, namely: additive MAC, and logical conjunction/disjunction MAC. In these two cases, we show that privacy for free can be attained by designing retrieval strategies that exploit the properties of the channel to maximize the retrieval rate. We show that for the additive MAC, the optimal PIR scheme is linear, while for the logical conjunction/disjunction MAC, we show that a non-linear PIR scheme that requires $N \geq 2^{M-1}$ is needed to achieve $C = 1$. We conclude this discussion by showing that full unconstrained capacity may not be attainable for all MACs by giving a counterexample, which is the selection MAC.

Throughout Chapters 2-6, we have confined ourselves to protecting the privacy of the desired message from the databases in addition to satisfying the reliability constraint. In Chapter 7, we tackle the problem of *secure PIR*. We impose an extra constraint to the PIR problem, namely, the secrecy constraint in addition to the usual privacy constraint. This ties together the two focuses of this dissertation. The secrecy constraint ensures that the queries and the answer strings do not leak any information about the contents of the databases to an external eavesdropper. Such

systems are relevant in practice, for example, in the stock market example, consider the case when the contents of the records themselves are confidential except for a small subset of authorized investors. Thus, the queries and the answer strings should be designed such that unauthorized entities who wiretap the retrieval process learn absolutely nothing about the contents of these confidential records. A few works exist on secure PIR: [63] considers the more general problem of information storage and retrieval, guaranteeing that also the process of storing the information is secure in the presence of failing servers. [21] considers a symmetric PIR setting where there is a passive eavesdropper who can tap in on the incoming and outgoing transmissions of any E servers. [21] derives the PIR capacity in this setting. Interestingly, the secret key needed for the symmetric retrieval process is used as an encryption key to secure the contents of the databases from the eavesdropper. This requires, as in the underlying symmetric PIR, that databases exchange a secret key of at least a certain size. This problem is investigated further in [41] for the classical PIR problem under T-privacy constraint for the case of $E \leq T$. [41] derives inner and outer bounds for this problem in addition to the minimum amount of common randomness required, which is shared between the databases.

In Chapter 7, we study the secure PIR problem from a different angle than [21, 41, 63] by investigating the problem of *PIR through wiretap channel II (PIR-WTC-II)*. Ozarow and Wyner [64] introduced the wiretap channel II (WTC-II) model, which considers a noiseless main channel and a binary erasure channel to the wiretapper, where the wiretapper is able to select the positions of erasures. In PIR-WTC-II, the user observes the t_n-length answer strings through a noiseless

channel from the nth database. The eavesdropper can observe a fraction μ_n from the nth answer string. The databases should encode the answer strings such that the eavesdropper learns nothing from observing any μ_n fraction of the traffic from the nth database. Naturally, the nth database dedicates $\mu_n t_n$ portion of the answer string to confuse the eavesdropper, constraining the meaningful portion of the answer to be $(1-\mu_n)t_n$. This fundamentally relates PIR-WTC-II to the PIR problem under asymmetric traffic constraints in Chapter 5, as the lengths of the answer strings can no longer be symmetric. We raise the following questions: How can we design a retrieval code that satisfies the combined privacy and security constraints for the PIR-WTC-II problem? Does PIR-WTC-II problem necessitate the existence of common randomness between the databases as in [41]? Should the databases share any common randomness with the user (retriever)?

In this chapter, we obtain a general upper bound for the PIR-WTC-II problem, when the eavesdropper can wiretap $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)$ fractions from the traffic outgoing from every database. We show that this upper bound can be expressed as a max-min problem. The inner minimization problem extends the converse techniques for the PIR problem under asymmetric traffic constraints in Chapter 5 to the PIR-WTC-II problem. The outer problem maximizes the retrieval rate over all possible traffic ratio vectors. For the achievability, we extend the achievable scheme used in Chapter 5 to achieve the corner points for the meaningful portions of the queries. In the extension, to satisfy the security constraint, each database generates a secret key with $\mu_n t_n$ length and encodes it into an artificial noise vector using a $(t_n, \mu_n t_n)$ MDS code and encrypts the returned answer strings with the artificial noise vector.

Interestingly, our achievable rate does not need any shared randomness among the databases or between the databases and the user. The keys used by the databases are unknown to the user, but are decodable and canceled at the retriever; however, the same keys are not extractable at the wiretapper due to the MDS code used. Our upper and lower bounds match for $M = 2$ and $M = 3$, for any N, and any $\boldsymbol{\mu}$.

In Chapters 8-10, we shift our focus to security problems in multi-user networks by means of physical layer (information-theoretic) security techniques. Physical layer security provides unconditional and provable security schemes that are quantifiable in terms of information-theoretic quantities and rates [65]. Physical layer security techniques allow secure transmission of information (in absolute sense) without the need for encryption keys. Consequently, the problems of exchanging encryption keys across open wireless networks are mitigated. The wiretap channel was first considered by Wyner in [66]. The canonical wiretap channel model consists of a transmitter, a legitimate receiver and an eavesdropper. Wyner showed the feasibility of attaining a positive secrecy rate in his degraded channel model, and determined the rate-equivocation region of a degraded wiretap channel. This model was generalized to arbitrary, not necessarily degraded, channels by Csiszar and Korner in [67], where they determined the rate-equivocation region of the most general wiretap channel. Leung-Yan-Cheong and Hellman considered SISO Gaussian wiretap channel, which is degraded, under a transmitter-side power constraint in [68]. They showed that Gaussian signalling is optimal. The physical layer security framework is then extended to various multiuser settings such as: the multiple access wiretap channel (MAC-WT) [69], broadcast channel with confidential mes-

sages (BCCM) [70–74], interference channel with confidential messages (ICCM) [70], multi-receiver wiretap channels [75, 76], and relay-eavesdropper channels [77]. The secure degrees of freedom (s.d.o.f.) have been considered in the literature as a first order approximation of the secure rates (the pre-log factor of the secure rate) in many multiuser channel models, such as: helper wiretap channel [78, 79], multiple-access wiretap channel [78,80–82], interference channel [78,82–86], X-channel [87,88], half-duplex relay channel [89], compound wiretap channel [90], diamond channel [91], multiuser channel models under imperfect CSIT [92–96].

The most relevant works to this dissertation are the s.d.o.f. characterization of SISO one-hop networks [78, 82, 86] and the MIMO Gaussian wiretap channel in [97–100]. [78] determines the exact s.d.o.f. of several SISO networks, including the wiretap channel with helpers, MAC-WT, BCCM, and ICCM. For achievability, [78] proposes real interference alignment [101] based achievable schemes that use structured codes in the form of pulse amplitude modulation (PAM). For the converse, [78] develops two converse lemmas, the secrecy penalty lemma and the role of a helper lemma, which prove the optimality of the proposed achievable schemes. Reference [86] generalizes the sum s.d.o.f. result of ICCM in [78] to the case of K-users. The work in [86] shows that in order to achieve real interference alignment at multiple receivers as in the case of the K-user interference channel, asymptotic real interference alignment is needed. [82] generalizes [78], [86] to determine the entire s.d.o.f. region; [82] shows that the s.d.o.f. region has a general polytope structure. Reference [79] extends the result for the wiretap channel with helpers in [78] to the case of MIMO nodes for the special case of a single helper. To that end, [79]

extends the role of a helper lemma of [78] for the MIMO case and provides multiple achievable schemes for different regimes including spatial precoding/alignment, transmission in the null space, and projecting onto a SISO dimension where real interference alignment of [78] is used. The MIMO Gaussian wiretap channel was considered in [97–99], under a transmitter-side power constraint. These references show that channel prefixing is not needed, and Gaussian signalling is optimal. An interesting alternative proof is given in [100] based on the channel enhancement technique developed in [102]. Reference [100] considers the MIMO wiretap channel under a transmitter-side covariance matrix constraint which is more general than a transmitter-side power constraint.

In all these works, the following assumptions are imposed: First, they assume that all nodes are *altruistic* and follow a prescribed transmission policy in order to maximize the sum secure rate of the entire system, even if that obliges the transmitters to jam their own receivers as in the case of ICCM. Second, the channel inputs to Gaussian channels are usually constrained by a total *transmitter-side* power constraint. Third, the transmitters and receivers have a *single antenna* in the secure degrees of freedom studies of interference channels. That is, the optimal interplay between interference, security, and multiple antennas is not fully understood even in high SNR regimes. In this thesis, we address these issues in Chapters 8-10, namely, the MIMO interference channel with confidential messages, security in networks with user misbehavior, and MIMO wiretap channel under receiver-side power constraints. The goal is to characterize the optimal secure transmission strategy in terms of the *secrecy capacity* or its high-SNR approximation, the secure degrees of

freedom (s.d.o.f.).

In Chapter 8, we consider the *two-user MIMO ICCM* [70], where two users wish to send messages to their respective receivers reliably, while keeping them secure from the unintended receiver in the information-theoretic sense. The secrecy capacity region of the ICCM is unknown today. In fact, the capacity region of the IC without secrecy constraints is known only within a constant gap [103]. Most of the current work concentrates on the asymptotic behavior of the secrecy capacity at high SNR in terms of s.d.o.f. The exact sum s.d.o.f. [78, 86] and the entire s.d.o.f. region [82] of the single-input single-output (SISO) ICCM are known for an arbitrary number of transmitters and receivers. In this chapter, we extend the s.d.o.f. results for the ICCM in [78,82,86] to the case of MIMO nodes, for the special case of a two-user system with an equal number of antennas at both transmitters (M) and both receivers (N).

We first focus on the optimal achievability schemes for the sum s.d.o.f. point. We propose a novel achievable scheme for the 2×2 ICCM system. The 2×2 achievable scheme is central in this chapter, since for the ICCM, the final sum s.d.o.f. numbers are multiples of 1/3. The required achievable scheme depends on the value of the fractional (non-integer) part of the sum s.d.o.f. If it is 1/3, a projection onto a single SISO dimension as in [79] is sufficient. In this SISO dimension, we use real interference alignment scheme of [78] for ICCM. However, if it is 2/3, the projection strategy results in a 2×2 ICCM system. In this case, we use a combination of spatial interference alignment scheme [104], which ensures security, and an asymptotic real interference alignment scheme [86], which ensures efficient decodability. Any other

antenna configuration (any M and N) can be reduced to either a 1×1 ICCM (i.e., SISO) or a 2×2 ICCM system after proper vector space operations for the integer part of the sum s.d.o.f. These operations include transmission in the null space of the cross-links and spatial alignment.

Next, we develop a matching converse by using three distinct outer bounds. The first upper bound is the cooperative bound, in which we allow cooperative stochastic encoding among the two users. The second upper bound uses the vectorized version of the upper bounding technique developed in [78] using the secrecy penalty and role of a helper lemmas. The third upper bound is the decodability upper bound developed in [105] for the IC without secrecy constraints. The intersection of these upper bounds gives a tight upper bound for any number of antennas.

Then, we characterize the complete s.d.o.f. region. We prove that the region is a four-vertex polytope in general as in [82]. The non-trivial extreme points are the sum s.d.o.f. point and the two symmetric maximum individual s.d.o.f. points. We note that the s.d.o.f. region becomes a square if $\frac{N}{2} \leq M \leq \frac{2N}{3}$ or $M \geq 2N$, which implies the feasibility of simultaneous secure transmission with a full s.d.o.f. in these regimes. For other regimes, the s.d.o.f. region is a non-square polytope, since the sum s.d.o.f. point and the maximum individual s.d.o.f. points evolve differently with the number of transmit antennas M. After establishing the achievable schemes for the non-trivial points of the polytope, the rest of the region is obtained via time-sharing. Finally, we specialize the problem to the case of time-varying ICCM. We develop simpler achievable schemes that depend on repeating the transmitted symbols over multiple channel uses, which replaces the complicated asymptotic real interference

alignment scheme, and exploits the time-diversity inherent in time-varying ICCM.

In Chapter 9, we investigate BCCM, ICCM, and MAC-WTC channel models in the case of selfish and malicious behaviour, where the users/helpers do not perform the system-wide-optimal altruistic behaviour but apply a selfish strategy and/or take sides by aiming to help one user and potentially hurt the other. These new models are extensions of the ones studied in [69, 70, 78] and are a step forward in studying channel models with active adversaries. We use s.d.o.f. metric to quantify the effects of these malicious behaviours. For BCCM and ICCM channel models, we note a self-enforcing property: Even with the excessive capabilities of the helpers/users (infinite power and all-knowing entities), these capabilities are naturally restricted in these channel models due to the users/helpers' interest in reliable communication to/with their own receivers. That is, no entity can use infinite powered Gaussian jamming signals which would wipe out the communication for everybody. This self-enforcing property necessitates users to apply selective jamming via interference alignment. This motivates studying such jamming techniques and analyzing their effect on the s.d.o.f. of the users. In addition, a careful look at the achievable scheme for the MAC-WTC in [78] reveals that the cooperative jamming signal of each user protects parts of the message-carrying signals of the other users; and that no user can protect its own signals. This creates an interesting ecosystem where each user strictly depends on the rest of the users for its own security. The fact that a user's cooperative jamming transmission does not contribute to its own security, but at the same time uses up its own transmit power, may motivate some selfish users not to send cooperative jamming signals. In this chapter, we investigate the effects of such

(and worse) deviations from the optimum signalling scheme on the system s.d.o.f., and the actions that the rest of the users can take to compensate for such behavior.

In the first model, which is the *BCCM with combating helpers*, there are two helpers, where each helper takes the side of one of the receivers and at the same time aims to hurt the secure communication to the other receiver. The two helpers have contradicting objectives and hence are combating. Helpers in this model do not coordinate with the transmitter as in [78]. We use a stringent objective function for each helper: Each helper minimizes the s.d.o.f. of the other receiver, while not decreasing the s.d.o.f. of its own receiver by its action. We formulate the problem as an extensive-form game [106], which is a sequential strategic game, where every player (node) acts according to its information about the other nodes' actions in previous transmission frames. We investigate achievable schemes that use real interference alignment [101] in a recursive way. We prove that under this stringent objective function and recursive real interference alignment, the malicious behaviours of the two combating helpers are neutralized, and the s.d.o.f. for each user converges to the optimal s.d.o.f. of $1/2$ per user [78], as if both helpers are altruistic.

In the second model, which is the *ICCM with selfish users*, there is an external system helper. In this model, the users do not coordinate as in the optimal strategy in [78] instructs. The users are selfish and want to hurt the other receiver; each transmitter's goal is to maximize the difference of the s.d.o.f. between the two receivers. This permits each user to jam its own receiver if this hurts the other receiver more, making self-jamming more natural here than the optimum scheme in [78]. There is a neutral helper in this system which aims to maximize the s.d.o.f. of the system. Us-

ing the extensive-form game formulation and recursive real interference alignment, we show that the selfishness of the users precludes any secure communication, and drives the s.d.o.f. of both users to zero, despite the existence of a mediating helper.

In the third model, which is the *MAC-WTC with deviating users*, we first consider the case where M out of K users deviate by not transmitting cooperative jamming signals. We start by evaluating the achievable sum s.d.o.f. when the remaining users do not change their original optimum strategies. We show that the sum s.d.o.f. of the system decreases, and deviating users do not benefit from their actions. Then, we consider two possible counter-strategies by the remaining users: In the first strategy, all users decrease their rates to ensure that all message-carrying signals are protected by the remaining cooperative jamming signals, and leakage s.d.o.f. is zero. We show that, in this case, the individual s.d.o.f. of the deviating users increase. Hence, deviating users gain at the expense of well-behaving users. In the second strategy, we allow the leakage s.d.o.f. to be non-zero, but constrain leakage in a single dimension. We show that, although the sum s.d.o.f. of the system is lower than the case of the first counter-strategy, this strategy decreases the individual s.d.o.f. of the deviating users and increases the s.d.o.f. of well-behaving users. Next, we consider a more severe form of deviation by considering one user turning malicious and sending intentional jamming signals. As this deviating user has infinite power, it can wipe out all communication, secure or otherwise, if it sends Gaussian signals. For the sake of a meaningful formulation, we restrict the strategy set of this deviating user to be of structured signalling and alignment type. Under this restriction, we formulate the problem as an extensive-form game [106]. We show

that this deviating user can drive the s.d.o.f. of the system to zero. We then show that, interestingly, the remaining users can utilize these intentional (malicious) jamming signals to protect more message-carrying signals at the eavesdropper, achieving a sum s.d.o.f. of $\frac{(K-1)^2}{(K-1)^2+1}$. We prove that this sum s.d.o.f. matches the sum s.d.o.f. of a $K-1$ user MAC-WTC with 1 external altruistic helper, thereby, show that the system turns a malicious jammer into an altruistic helper, i.e., the deviating user benefits the system against its intensions.

Finally, motivated by the emerging applications of wireless energy transfer and cognitive radio, we investigate the *MIMO wiretap channel under receiver-side power constraints* in Chapter 10. Most existing literature on Gaussian channels is based on a transmitter-side average power constraint. This constraint models the maximum allowable power at the transmitter-side. Gastpar [107] was the first to consider a receiver-side power constraint. In [107], he considered a maximum receiver-side power constraint motivated by the desire to limit the received interference in a cognitive radio application. He observed that the solution changes significantly for a MIMO channel. Subsequently, Varshney [108] considered a minimum receiver-side power constraint motivated by the desire to transport both information and energy simultaneously over a wireless channel. Varshney as well observed that the solution changes significantly with respect to a classical transmitter-side amplitude constrained SISO channel [109]. In this chapter, we consider a multi-user and multi-objective version of the problem considered by Gastpar and Varshney. In particular, we consider a MIMO wiretap channel where the transmitter wishes to have secure communication with a legitimate receiver in the presence of an eavesdropper. In this

model, messages need to be sent to the legitimate receiver with perfect secrecy from the eavesdropper. We impose the usual transmitter-side power constraint in addition to a receiver-side power constraint. Therefore, our model generalizes [107, 108] from a single-user setting of two nodes to a multi-user scenario of a wiretap channel with three nodes, and also to a multi-objective setting where we have both reliability and security constraints.

We first characterize the secrecy capacity of the MIMO wiretap channel under a minimum receiver-side power constraint at the eavesdropper only. To this end, we first show that, solving the secrecy capacity of the MIMO wiretap channel under a transmitter-side maximum power constraint and a receiver-side minimum power constraint is equivalent to solving the secrecy capacity of a MIMO wiretap channel under a *double-sided correlation matrix constraint* on the channel input at the transmitter. This is a generalization of the approach of [100, 102]. We then generalize the channel enhancement technique of [100, 102] to the case of double-sided correlation matrix constraint. This gives us the converse. We next show that the rates given in the converse can be achieved by two different achievable schemes: a mean based scheme where the transmitter uses a Gaussian codebook with a fixed mean, and an artificial noise [110] (or cooperative jamming [111]) based scheme, which uses Gaussian channel prefixing with a Gaussian codebook. The role of the mean or the artificial noise is to enable energy transfer without sacrificing from the secure rate. This is the first instance of a channel model where either the use of a mean signal or the use of channel prefixing via artificial noise is strictly necessary for the canonical MIMO wiretap channel. We note that, in a related work, refer-

ences [112, 113] consider simultaneous information and energy transfer in a MISO wiretap channel, and focus on optimizing the performance of a specific artificial noise based achievable scheme with no claim of optimality. We also note a similar set-up in [114, 115], where the authors consider the case of statistical channel state information only at the transmitter and focus on optimizing asymptotic transmit covariance matrix of Gaussian codebooks without artificial noise for the case of a large number of transmit antennas.

We then extend the developed methodology to find the capacities of the following related channels: We first consider the case that both receivers have minimum receiver-side power constraints. Next, we impose maximum power constraints, which corresponds to a cognitive radio setting where we control the received interference power at users. In this case, we show that ordinary Gaussian signalling is sufficient, and there is no need for mean or artificial noise signalling. Next, we drop the secrecy constraint and consider the classical MIMO broadcast channel (BC) with minimum receiver-side power constraints. We prove that dirty paper coding (DPC) used in [102] is optimal to achieve the capacity. This result intuitively verifies that neither mean nor artificial noise transmission is needed, because the freedom afforded by the design of the covariance matrices of the DPC scheme suffices to achieve all desired feasible receiver-side powers. Finally, we put back the secrecy constraints for both users and consider the BC with confidential messages BCCM [72]. We show that secure DPC (S-DPC) is optimal for the BCCM as in [72] without the need for mean or artificial noise signalling.

In Chapter 11, we provide conclusions to this dissertation.

CHAPTER 2

Private Information Retrieval from Coded Databases

2.1 Introduction

In this chapter, we consider the PIR problem over a distributed storage system. Due to node failures and erasures that arise naturally in any storage system, redundancy should be introduced. However, replicating the content across the databases incurs high storage cost. This motivates the content of the databases to be coded instead of merely being replicated. In this chapter, the storage system consists of N non-colluding databases, each storing an MDS-coded version of M messages. We derive the information-theoretic capacity of the MDS-coded PIR problem to be $C = \left(1 + \frac{K}{N} + \frac{K^2}{N^2} + \cdots + \frac{K^{M-1}}{N^{M-1}}\right)^{-1} = (1 + R_c + R_c^2 + \cdots + R_c^{M-1})^{-1} = \frac{1-R_c}{1-R_c^M}$, where R_c is the rate of the (N, K) MDS code used. The capacity is a function of the code rate and the number of messages only regardless of the explicit structure of the storage code. The result implies a fundamental tradeoff between the optimal retrieval cost and the storage cost when the storage code is restricted to the class of MDS codes.

2.2 System Model

Consider an (N, K) MDS-coded distributed storage system storing M messages (or files). The messages are independent and identically distributed with

$$H(W_i) = L, \quad i \in \{1, \cdots, M\} \tag{2.1}$$

$$H(W_1, W_2, \cdots, W_M) = ML \tag{2.2}$$

The message W_i, $i \in \{1, \cdots, M\}$ is a $\mathbb{F}_q^{\tilde{L} \times K}$ matrix with sufficiently large field $\mathbb{F}_q$, such that $\tilde{L} \times K = L$. The elements of W_i are picked uniformly and independently from $\mathbb{F}_q$. We denote the jth row of message W_i by $\mathbf{w}_j^{[i]} \in \mathbb{F}_q^K$. The generator matrix of the (N, K) storage code $\mathbf{H}$ is a $\mathbb{F}_q^{K \times N}$ matrix such that

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_N \end{bmatrix}_{K \times N} \tag{2.3}$$

where $\mathbf{h}_i \in \mathbb{F}_q^K$, $i \in \{1, \cdots, N\}$. For an MDS code, any set $\mathcal{K}$ of columns of $\mathbf{H}$ such that $|\mathcal{K}| \leq K$ are linearly independent.[1] We do not assume any specific structure on the distributed storage code other than that the encoding is performed

[1]For the converse proof, the linear independence requirement of every K columns in $\mathbf{H}$ is not strictly needed. In fact, from the converse point of view, any storage code that enforces the contents of every K databases to be statistically independent leads to the same upper bound even if the code is not linear. In this chapter, the linear independence assumption, which is equivalent to having an MDS code, is important for the construction of the achievable scheme (see Section 2.4) that relies on solving K linear equations, in addition to creating an instance of statistical independence that is needed in the converse proof.

independently over the rows, i.e., the rows/messages are not mixed[2,3]. Hence, the storage code $f_n : \mathbf{w}_j^{[i]} \to y_{n,j}^{[i]}$ on the nth database maps each row[4] of W_i separately into coded bit $y_{n,j}^{[i]}$, see Fig. 2.1,

$$y_{n,j}^{[i]} = \mathbf{h}_n^T \mathbf{w}_j^{[i]} \tag{2.4}$$

Consequently, the stored bits $\mathbf{y}_n \in \mathbb{F}_q^{M\tilde{L}}$ on the nth database, $n \in \{1, \cdots, N\}$ are concatenated projections of all messages $\{W_1, \cdots, W_M\}$ and are given by

$$\mathbf{y}_n = \begin{bmatrix} W_1 \\ \vdots \\ W_M \end{bmatrix} \mathbf{h}_n \tag{2.5}$$

[2]By *non-mixing* MDS code, we mean that each message is encoded separately. Furthermore, we assume that each row within each message is encoded separately as well. This assumption is made to enable the MDS code to be flexible enough so that the code structure makes sense for every message size L, which is needed to characterize the capacity in the Shannon sense (i.e., as $L \to \infty$). Here we give a concrete example: if $W_1 = (a_1, \cdots, a_4)$, and $W_2 = (b_1, \cdots, b_4)$ and they are encoded via a $(3, 2)$ non-mixing MDS code, then each message is arranged in 2 rows. Each row is encoded separately, for example, row 1 is encoded as $(a_1, a_2, a_1 + a_2)$, and row 2 is encoded as $(a_3, a_4, a_3 + a_4)$, and similarly for W_2. Note that this example MDS code neither mixes messages, nor the rows of each message. The results of this chapter are restricted to such non-mixing code structures and hence the qualifier "non-mixing" is dropped.

[3]We note that in [8, Example 2], an example for a mixing $(3, 2)$ MDS code for $M = 2$ is presented. In this case, letting $W_1 = (a_1, a_2)$, $W_2 = (b_1, b_2)$, database 1 stores (a_1, a_2), database 2 stores (b_1, b_2) and database 3 stores $(a_1 + b_1, a_2 + b_2)$. This code mixes W_1, W_2 in database 3. [8] provides a retrieval scheme for this specific code that achieves a retrieval rate of $\frac{2}{3}$, which is higher than the capacity of non-mixing $(3, 2)$ MDS codes ($C = \frac{3}{5}$). The characterization of the capacity of mixing MDS codes is an interesting open problem.

[4]We note that the assumption of encoding each row with the same generator matrix is indeed without loss of generality and is made to simplify the presentation. If each row is encoded via a different MDS generator matrix, i.e., the jth row of message i is encoded via $\mathbf{H}_j^{[i]}$, the capacity is still given by Theorem 2.1. For the achievable scheme, we note that the scheme downloads K coded symbols directly from the databases with no further processing. This suffices to decode the entire row because the MDS property is still valid for each row. The converse proof still holds since the contents of every K databases are statistically independent and hence Lemma 2.1 is still valid.

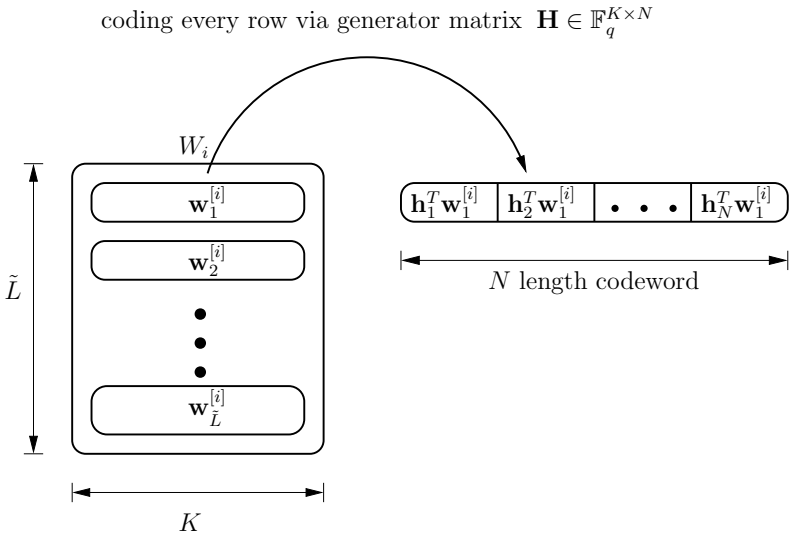


Figure 2.1: Coding process for message W_i.

$$= \left[\mathbf{h}_n^T\mathbf{w}_1^{[1]} \quad \dots \quad \mathbf{h}_n^T\mathbf{w}_{\tilde{L}}^{[1]} \quad \mathbf{h}_n^T\mathbf{w}_1^{[2]} \quad \dots \quad \mathbf{h}_n^T\mathbf{w}_{\tilde{L}}^{[2]} \quad \dots \quad \mathbf{h}_n^T\mathbf{w}_1^{[M]} \quad \dots \quad \mathbf{h}_n^T\mathbf{w}_{\tilde{L}}^{[M]} \right]^T \tag{2.6}$$

The explicit structure of the coded storage system is illustrated in Table 2.1. The described storage code can tolerate up to $N - K$ errors by connecting to any K databases. Thus, we have for any set $\mathcal{K}$ such that $|\mathcal{K}| \geq K$,

$$H(\mathbf{y}_{\bar{\mathcal{K}}}|\mathbf{y}_{\mathcal{K}}) = 0 \tag{2.7}$$

where $\mathbf{y}_{\mathcal{K}}$ are the stored bits on databases indexed by $\mathcal{K}$, and $\bar{\mathcal{K}}$ is the complement of the set $\mathcal{K}$. The code rate of this distributed storage system R_c is given by

$$R_c = \frac{K}{N} \tag{2.8}$$

Table 2.1: Explicit structure of (N, K) code for distributed databases with M messages.

	DB1 ($\mathbf{y}_1$)	DB2 ($\mathbf{y}_2$)	$\cdots$	DBN ($\mathbf{y}_N$)
message 1	$\mathbf{h}_1^T\mathbf{w}_1^{[1]}$ $\mathbf{h}_1^T\mathbf{w}_2^{[1]}$ $\vdots$ $\mathbf{h}_1^T\mathbf{w}_{\tilde{L}}^{[1]}$	$\mathbf{h}_2^T\mathbf{w}_1^{[1]}$ $\mathbf{h}_2^T\mathbf{w}_2^{[1]}$ $\vdots$ $\mathbf{h}_2^T\mathbf{w}_{\tilde{L}}^{[1]}$	$\cdots$ $\cdots$ $\cdots$ $\cdots$	$\mathbf{h}_N^T\mathbf{w}_1^{[1]}$ $\mathbf{h}_N^T\mathbf{w}_2^{[1]}$ $\vdots$ $\mathbf{h}_N^T\mathbf{w}_{\tilde{L}}^{[1]}$
message 2	$\mathbf{h}_1^T\mathbf{w}_1^{[2]}$ $\mathbf{h}_1^T\mathbf{w}_2^{[2]}$ $\vdots$ $\mathbf{h}_1^T\mathbf{w}_{\tilde{L}}^{[2]}$	$\mathbf{h}_2^T\mathbf{w}_1^{[2]}$ $\mathbf{h}_2^T\mathbf{w}_2^{[2]}$ $\vdots$ $\mathbf{h}_2^T\mathbf{w}_{\tilde{L}}^{[2]}$	$\cdots$ $\cdots$ $\cdots$ $\cdots$	$\mathbf{h}_N^T\mathbf{w}_1^{[2]}$ $\mathbf{h}_N^T\mathbf{w}_2^{[2]}$ $\vdots$ $\mathbf{h}_N^T\mathbf{w}_{\tilde{L}}^{[2]}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
message M	$\mathbf{h}_1^T\mathbf{w}_1^{[M]}$ $\mathbf{h}_1^T\mathbf{w}_2^{[M]}$ $\vdots$ $\mathbf{h}_1^T\mathbf{w}_{\tilde{L}}^{[M]}$	$\mathbf{h}_2^T\mathbf{w}_1^{[M]}$ $\mathbf{h}_2^T\mathbf{w}_2^{[M]}$ $\vdots$ $\mathbf{h}_2^T\mathbf{w}_{\tilde{L}}^{[M]}$	$\cdots$ $\cdots$ $\cdots$ $\cdots$	$\mathbf{h}_N^T\mathbf{w}_1^{[M]}$ $\mathbf{h}_N^T\mathbf{w}_2^{[M]}$ $\vdots$ $\mathbf{h}_N^T\mathbf{w}_{\tilde{L}}^{[M]}$

The retrieval process over MDS-coded databases is illustrated in Fig. 2.2. To retrieve W_i, the user generates a query $Q_n^{[i]}$ and sends it to the nth database. Since the user does not have knowledge about the messages in advance, the queries are independent of the messages,

$$I(Q_1^{[i]}, \cdots, Q_N^{[i]}; W_1, \cdots, W_M) = 0 \tag{2.9}$$

In order to ensure privacy, the retrieval strategy for the ith message should be indistinguishable from the retrieval strategy of W_1, hence, for $i \in \{1, \cdots, M\}$, $n \in$

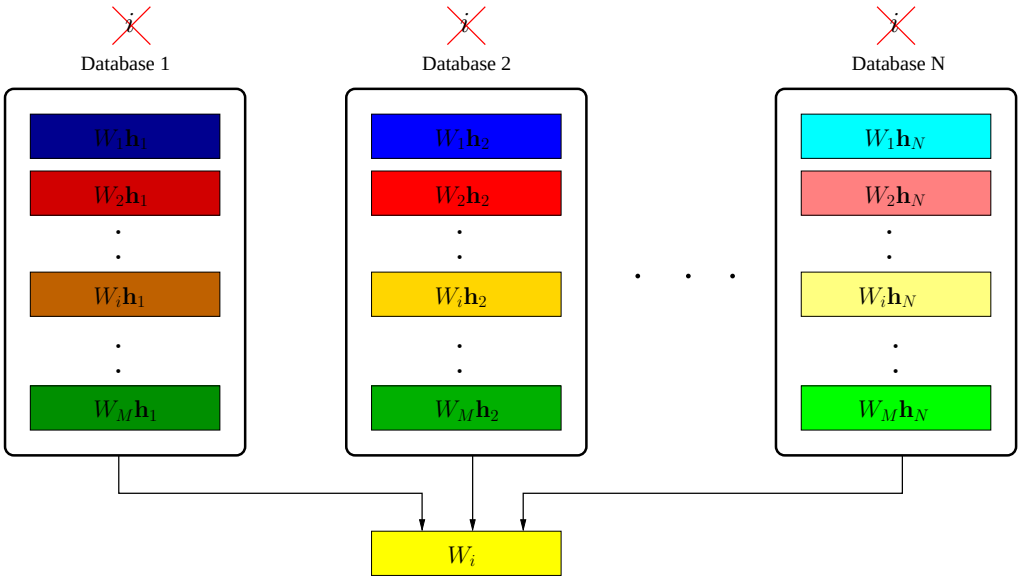


Figure 2.2: The MDS-coded PIR problem.

$\{1, \cdots N\}$

$$(Q_n^{[i]}, A_n^{[i]}, W_1, \cdots, W_M) \sim (Q_n^{[1]}, A_n^{[1]}, W_1, \cdots, W_M) \tag{2.10}$$

which implies that the queries and answers should be independent of the desired message index i, i.e., the privacy constraint is,

$$I(Q_n^{[i]}, A_n^{[i]}, W_1, \cdots, W_M; i) = 0, \quad n \in \{1, \cdots, N\} \tag{2.11}$$

Each database responds with an answer string $A_n^{[i]}$, which is a deterministic function[5] of the received query and the stored coded bits in the nth database. Hence, by the

[5]We note that the assumption that the answer strings are deterministic functions of the queries and the stored information is indeed without loss of generality and is kept for the simplicity of presentation. The converse proof can be extended to the case of allowing the databases to use randomized strategies. In this case, a common randomness should be shared between the user and the databases. More specifically, we can assume that there exists a random variable $\mathbb{G}$ that is shared between the user and the databases such that $\mathbb{G}$ is independent of $(i, W_{1:M})$, and $H(A_n^{[i]}|Q_n^{[i]}, \mathbf{y}_n, \mathbb{G}) = 0$. This does not change the converse lemmas except for conditioning all inequalities on $\mathbb{G}$. A similar formulation of this idea can be found in [19].

data processing inequality,

$$H(A_n^{[i]}|Q_n^{[i]}, \mathbf{y}_n) = H(A_n^{[i]}|Q_n^{[i]}, W_1, \cdots, W_M) = 0 \tag{2.12}$$

In addition, the user should be able to decode W_i reliably from all the answer strings collected from the N databases with a small probability of error. Consequently, from Fano's inequality, we have the following reliability constraint,

$$H(W_i|A_1^{[i]}, \cdots, A_N^{[i]}, Q_1^{[i]}, \cdots, Q_N^{[i]}) = o(L) \tag{2.13}$$

where $\frac{o(L)}{L} \to 0$ as $L \to \infty$. The retrieval rate R for the PIR problem is the ratio of the size of the desired message to the total download cost under the reliability constraint (2.13) and the privacy constraint (2.10) for some $L \in \mathbb{N}$, i.e.,

$$R = \frac{H(W_i)}{\sum_{n=1}^{N} H(A_n^{[i]})}, \quad \text{subject to } (2.10), (2.13) \tag{2.14}$$

The PIR capacity C is the supremum of R over all retrieval schemes as $L \to \infty$.

In this chapter, as in [12], we follow a Shannon theoretic formulation by assuming that the message size can be arbitrarily large. Also, we neglect the upload cost with respect to the download cost as in [12].

We note that the described storage code is a generalization of the repetition-coded problem in [12]. If $K = 1$ and $h_n = 1$, $n \in \{1, \cdots, N\}$, then the problem reduces to the classical PIR in [12]. In addition, the systematic MDS-coded instance[6]

[6]We note that although the code structure presented in [10] is assumed to be systematic,

presented in [10] is a special case of this setting with $\mathbf{h}_n = \mathbf{e}_n$, $n \in 1, \cdots, K$, where $\mathbf{e}_n$ is the nth standard basis vector.

2.3 Main Result

Theorem 2.1 *For an (N, K) MDS-coded distributed database system with coding rate $R_c = \frac{K}{N}$ and M messages, the PIR capacity is given by*

$$C = \frac{1 - R_c}{1 - R_c^M} \tag{2.15}$$

$$= \frac{1}{1 + R_c + \cdots + R_c^{M-1}} \tag{2.16}$$

$$= \left(1 + \frac{K}{N} + \frac{K^2}{N^2} + \cdots + \frac{K^{M-1}}{N^{M-1}}\right)^{-1} \tag{2.17}$$

We have the following remarks about the main result. We first note that the PIR capacity in (2.15) is a function of the coding rate R_c and the number of messages M only, and does not depend on the explicit structure of the coding scheme (i.e., the generator matrix) or the number of databases. This observation implies the universality of the scheme over any MDS-coded database system with the same coding rate and number of messages. The result also entails the optimality of separation between distributed storage code design and PIR scheme design for a fixed R_c. We also note that the capacity C decreases as R_c increases. As $R_c \to 0$, the PIR capacity approaches $C = 1$. On the other hand, as $R_c \to 1$, the PIR capacity approaches $\frac{1}{M}$ which is the trivial retrieval rate obtained by downloading

this assumption is indeed without loss of generality. The scheme in [10] is universal and can be applied for any (N, K) MDS code and was presented for systematic MDS codes for sake of simpler exposition of the retrieval scheme.

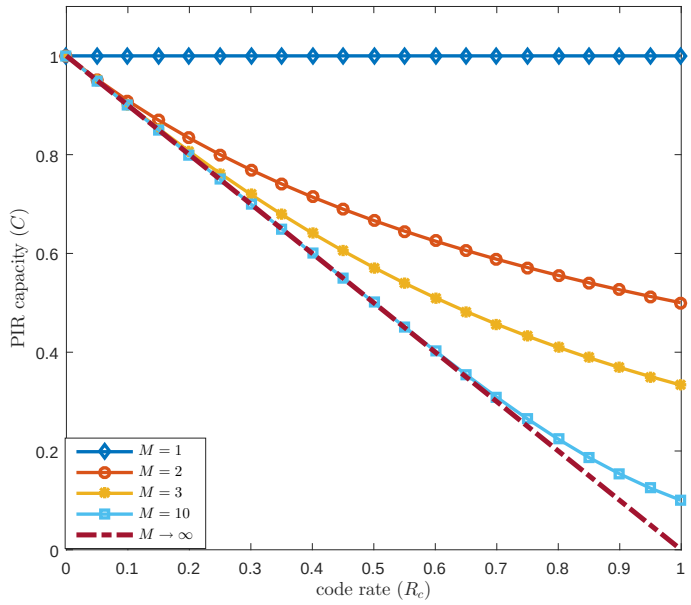


Figure 2.3: PIR capacity versus R_c.

the contents of all databases. This observation implies that a fundamental tradeoff exists between storage cost and the retrieval download cost when the storage code is restricted to the class of MDS codes. This tradeoff conforms with the result of [8]. The capacity expression in Theorem 1 is plotted in Fig. 2.3 as a function of the code rate R_c for various numbers of messages M.

The capacity in (2.15) is strictly larger than the best-known achievable rate in [10], where $R = 1 - R_c$ for any finite number of messages. We observe also that the PIR capacity for a given fixed code rate R_c is monotonically decreasing in M. The rate in (2.15) converges to $1 - R_c$ as $M \to \infty$. Intuitively, as the number of messages increases, the undesired download rate must increase to hide the identity of the desired message; eventually, the gain from applying the greedy algorithm in Section 2.4 over the scheme in [10] diminishes. This confirms that the achievable

scheme in [10] is asymptotically optimal. Our capacity here generalizes the capacity in [12] where $R_c = \frac{1}{N}$. That is, the classical PIR problem may be viewed as a special case of the MDS-coded PIR problem with a specific code structure which is repetition coding.

2.4 Achievability Proof

In this section, we present the general achievable scheme for Theorem 2.1. We give a few specific examples in Section 2.5. Our achievable scheme generalizes the achievable scheme in [12] which induces symmetry across databases and symmetry across messages, and exploits the side information. The achievable scheme here includes two extra steps due to the presence of coding: decoding of the interference and decoding of the desired rows which are not present in [12].

2.4.1 Achievable Scheme

The scheme requires $\tilde{L} = N^M$, which implies that the size of message $H(W_i) = L = KN^M$. The scheme is completed in M rounds, each corresponding to the sum of i terms, $i \in \{1, \cdots, M\}$, and is repeated K times to decode the desired message; see Tables 2.2 and 2.3 for examples.

1. *Index preparation:* The user interleaves the indices of rows for all messages randomly and independently from each other, i.e., for any message W_ℓ,

$$\mathbf{x}_i^{[m]} = \mathbf{w}_{\pi_\ell(i)}^{[\ell]}, \quad i \in \{1, \cdots, \tilde{L}\} \tag{2.18}$$

where $\pi_\ell(\cdot)$ is a random interleaver used for message ℓ and known privately to the user only. In this case the rows chosen at any database appear to be chosen at random and independent from the desired message index.

2. *Initialization:* The user downloads K^{M-1} desired coded bits from different rows of the desired message W_m from database 1 (DB1) and sets round index $i = 1$, i.e., the user starts by downloading the symbols $\mathbf{h}_1^T\mathbf{x}_1^{[m]}, \cdots, \mathbf{h}_1^T\mathbf{x}_{K^{M-1}}^{[m]}$ from database 1.

3. *Symmetry across databases:* The user downloads K^{M-1} desired bits each from a different row from each database, i.e., the user downloads from database 2 the symbols $\mathbf{h}_2^T\mathbf{x}_{K^{M-1}+1}^{[m]}, \cdots, \mathbf{h}_2^T\mathbf{x}_{2K^{M-1}}^{[m]}$, from database 3 the symbols $\mathbf{h}_3^T\mathbf{x}_{2K^{M-1}+1}^{[m]}, \cdots, \mathbf{h}_3^T\mathbf{x}_{3K^{M-1}}^{[m]}$, $\cdots$, similarly until the user downloads $\mathbf{h}_N^T\mathbf{x}_{(N-1)K^{M-1}+1}^{[m]}, \cdots, \mathbf{h}_N^T\mathbf{x}_{NK^{M-1}}^{[m]}$ from database N. Then, the total number of desired bits in the ith round is NK^{M-1}.

4. *Message symmetry:* To satisfy the privacy constraint, the user needs to download an equal amount of coded bits from all other messages. Consequently, the user downloads $\binom{M-1}{i}K^{M-i}(N-K)^{i-1}$ bits from each database. The undesired equation is a sum of i terms picked from the remaining undesired messages. To be more specific, the user downloads the sum $\mathbf{h}_n^T\left(\mathbf{x}_{j_1}^{[\ell_1]} + \mathbf{x}_{j_2}^{[\ell_2]} + \cdots + \mathbf{x}_{j_i}^{[\ell_i]}\right)$ from the rows $j_1, \cdots, j_i \in \{1, \cdots, \tilde{L}\}$ of messages $\ell_1, \cdots, \ell_i \in \{1, \cdots, M\} \setminus m$ from the nth database. The specification of rows will become clear in step 5. Hence, the number of undesired equations downloaded in the ith round is $N\binom{M-1}{i}K^{M-i}(N-K)^{i-1}$.

5. *Decoding the interference:* The main difference of the coded problem from the uncoded PIR (i.e., repetition-coded counterpart) is that in order to exploit the undesired coded bits in the form of side information, the interference needs to be decoded first. Note that we are not interested in decoding the individual components of each term of the sum, but rather the components of the *aligned sum*. To perform this, we group each K undesired equations to be from the same rows, i.e., the user downloads the same sum from the rows $j_1, \cdots, j_i \in \{1, \cdots, \tilde{L}\}$ of messages $\ell_1, \cdots, \ell_i \in \{1, \cdots, M\} \setminus m$ as $\mathbf{h}_{n \mod N}^T \left(\mathbf{x}_{j_1}^{[\ell_1]} + \mathbf{x}_{j_2}^{[\ell_2]} + \cdots + \mathbf{x}_{j_i}^{[\ell_i]}\right), \cdots, \mathbf{h}_{n+K-1 \mod N}^T \left(\mathbf{x}_{j_1}^{[\ell_1]} + \mathbf{x}_{j_2}^{[\ell_2]} + \cdots + \mathbf{x}_{j_i}^{[\ell_i]}\right)$. The rows are chosen in order starting from row 1, and the index of the row is incremented whenever K symbols from the same row is downloaded. For example: the user downloads $\mathbf{h}_1^T \mathbf{x}_1^{[\ell]}$ from the undesired message ℓ from database 1, then the user downloads $\mathbf{h}_2^T \mathbf{x}_1^{[\ell]}$ from database 2, $\cdots$, until the user downloads $\mathbf{h}_K^T \mathbf{x}_1^{[\ell]}$ from database K. Starting from this point the user increments the index of the row to 2 and downloads $\mathbf{h}_{K+1}^T \mathbf{x}_2^{[\ell]}$ from database $K+1$, and so on. In this case, we have K linearly independent equations that can be uniquely solved, and hence the corresponding row of the interfering messages is decoded due to (2.7). Therefore, this generates $N\binom{M-1}{i} K^{M-(i+1)}(N-K)^{i-1}$ side information equations in the form of i term sums.

6. *Exploiting side information:* The side information generated in the previous step can be exploited in the $(i+1)$th round within the remaining $N-K$

databases that did not participate in generating them. The side information is used in $i+1$ term sum that includes the desired message as one of the terms. Since side information is successfully decoded, it can be canceled from these equations to leave desired coded bits. Hence, we can download $N\binom{M-1}{i}K^{M-(i+1)}(N-K)^i$ extra desired coded bits. More specifically, the user downloads the sums $\mathbf{h}_{n_1(n)}^T\left(\mathbf{x}_{\theta_1}^{[m]}+\mathbf{x}_{j_1}^{[\ell_1]}+\mathbf{x}_{j_2}^{[\ell_2]}+\cdots+\mathbf{x}_{j_i}^{[\ell_i]}\right),\cdots,$ $\mathbf{h}_{n_{N-K}(n)}^T\left(\mathbf{x}_{\theta_{N-K}}^{[m]}+\mathbf{x}_{j_1}^{[\ell_1]}+\mathbf{x}_{j_2}^{[\ell_2]}+\cdots+\mathbf{x}_{j_i}^{[\ell_i]}\right)$ from databases $n_1(n)=n+K$ mod $N,\cdots,n_{N-K}(n)=n+N-1 \mod N$ in the $(i+1)$th round, where $\mathbf{x}_{\theta_l}^{[m]}$ is the row θ_l from the desired message W_m, i.e., the user downloads the sum of the row from the desired message to the side information generated in the ith round.

7. Repeat steps 4, 5, 6 after setting $i=i+1$ until $i=M-1$.

8. *Decoding the desired message:* Till this point the scheme has downloaded one bit from each row of the desired message. To reliably decode the desired message, the scheme (precisely steps 2-7) is repeated K times. We repeat the scheme exactly except for shifting the order of databases circularly at each repetition for the desired coded bits. Note that the chosen indices for the desired message is the same up to circular shift at each repetition, however we download new undesired coded bits at each repetition. This creates K different equations for each row of the message and hence decodable.

9. *Shuffling the order of queries:* Since all databases know the retrieval scheme, every database can identify the desired message by observing the first query

only. 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 Decodability, Privacy, and Calculation of the Achievable Rate

Decodability: The decodability follows from the MDS property of the storage code, which states that in a $K \times N$ MDS generator matrix, any $K \times K$ submatrix is invertible. To show decodability formally, let W_m be the desired message without loss of generality. In each repetition, at the ith round, the user downloads $\binom{M-1}{i} K^{M-i}(N-K)^{i-1}$ symbols from the undesired messages from every database. These coded symbols are constructed as the sums of i coded symbols from some rows, i.e., the user downloads the sum $\mathbf{h}_n^T \left(\mathbf{x}_{j_1}^{[\ell_1]} + \mathbf{x}_{j_2}^{[\ell_2]} + \cdots + \mathbf{x}_{j_i}^{[\ell_i]} \right)$ from the rows $j_1, \cdots, j_i \in \{1, \cdots, \tilde{L}\}$ of messages $\ell_1, \cdots, \ell_i \in \{1, \cdots, M\} \backslash m$ from the nth database. The same sum is downloaded from K different databases, i.e., the user downloads the same sum from the rows $j_1, \cdots, j_i \in \{1, \cdots, \tilde{L}\}$ of messages $\ell_1, \cdots, \ell_i \in \{1, \cdots, M\} \setminus m$ as $\mathbf{h}_{n+1 \mod N}^T \left(\mathbf{x}_{j_1}^{[\ell_1]} + \mathbf{x}_{j_2}^{[\ell_2]} + \cdots + \mathbf{x}_{j_i}^{[\ell_i]} \right), \cdots, \mathbf{h}_{n+K-1 \mod N}^T \left(\mathbf{x}_{j_1}^{[\ell_1]} + \mathbf{x}_{j_2}^{[\ell_2]} + \cdots + \mathbf{x}_{j_i}^{[\ell_i]} \right)$. Since the submatrix $[\mathbf{h}_n \quad \mathbf{h}_{n+1 \mod N} \quad \cdots \quad \mathbf{h}_{n+K-1 \mod N}]$ is an invertible matrix by the MDS property, the sum of rows of $\mathbf{x}_{j_1}^{[\ell_1]} + \mathbf{x}_{j_2}^{[\ell_2]} + \cdots + \mathbf{x}_{j_i}^{[\ell_i]}$ is decodable. Note that there are a total of $N\binom{M-1}{i} K^{M-i}(N-K)^{i-1}$ of such symbols in the ith round, therefore $N\binom{M-1}{i} K^{M-i-1}(N-K)^{i-1}$ rows can be decoded as every K sums must be derived from the same set of rows.

These rows are used as side information in the $(i+1)$th round at the remaining

$N-K$ databases that do not contribute to the process of creating these side information. The user downloads from databases $n_1(n) = n + K \mod N, \cdots, n_{N-K}(n) = n + N - 1 \mod N$ the sums $\mathbf{h}_{n_1(n)}^T \left(\mathbf{x}_{\theta_1}^{[m]} + \mathbf{x}_{j_1}^{[\ell_1]} + \mathbf{x}_{j_2}^{[\ell_2]} + \cdots + \mathbf{x}_{j_i}^{[\ell_i]}\right)$, $\cdots$, $\mathbf{h}_{n_{N-K}(n)}^T \left(\mathbf{x}_{\theta_{N-K}}^{[m]} + \mathbf{x}_{j_1}^{[\ell_1]} + \mathbf{x}_{j_2}^{[\ell_2]} + \cdots + \mathbf{x}_{j_i}^{[\ell_i]}\right)$ in the $(i+1)$th round, where $\mathbf{x}_{\theta_l}^{[m]}$ is the row θ_l from the desired message W_m, i.e., the user downloads the sum of rows from the desired message with the side information generated in the ith round. Since the user has decoded the sum $\mathbf{x}_{j_1}^{[\ell_1]} + \mathbf{x}_{j_2}^{[\ell_2]} + \cdots + \mathbf{x}_{j_i}^{[\ell_i]}$, all undesired symbols can be canceled, and the user is left with the desired symbols only.

Now, for the desired symbols, we note that the user downloads from different rows within each repetition. Since the scheme repeats itself K times with the starting database shifted circularly, the user is left with $\mathbf{h}_n^T \mathbf{x}_\theta^{[m]}, \mathbf{h}_{n+1 \mod N}^T \mathbf{x}_\theta^{[m]}, \cdots, \mathbf{h}_{n+K-1 \mod N}^T \mathbf{x}_\theta^{[m]}$ for $\theta \in \{1, \cdots, \tilde{L}\}$. This creates K linearly independent equations for each row from W_m by the MDS property. Therefore, all rows can be decoded reliably.

Privacy: The scheme downloads all combinations of the sums containing i terms in the ith round from each database. Therefore, the same number of symbols from each message is queried from each database (specifically, KN^{M-1} coded symbols). Note that due to the fact that the user downloads the symbols (desired/undesired) from K databases in a circular shift pattern, each row is queried once within the same database. Thus, the user downloads from KN^{M-1} distinct rows from each database from every message. Since the indices of these rows are chosen randomly and uniformly, and the order of the queries is shuffled randomly and uniformly, the

privacy constraint $(Q_n^{[i]}, A_n^{[i]}, W_1, \cdots, W_M) \sim (Q_n^{[1]}, A_n^{[1]}, W_1, \cdots, W_M)$ is satisfied as all the query realizations are equally likely regardless of the message index i.

Achievable Rate Calculation: In each repetition, at the ith round, the user downloads the K coded symbols from $N\binom{M-1}{i}K^{M-i-1}(N-K)^{i-1}$ different rows of each message distributed among the N databases. From the described scheme, we note that other than the initial download of NK^{M-1} coded desired bits, at each round the scheme downloads $N\binom{M-1}{i}K^{M-(i+1)}(N-K)^{i}$ desired equations and $N\binom{M-1}{i}K^{M-i}(N-K)^{i-1}$ undesired equations. Hence, the total number of desired equations is $KN\sum_{i=0}^{M-1}\binom{M-1}{i}K^{M-1-i}(N-K)^{i}$, and the total number of undesired equations is $KN\sum_{i=1}^{M-1}\binom{M-1}{i}K^{M-i}(N-K)^{i-1}$ along the K repetitions of the scheme. Therefore, the achievable rate is,

$$\frac{1}{R} = 1 + \frac{\text{total undesired equations}}{\text{total desired equations}} \tag{2.19}$$

$$= 1 + \frac{\sum_{i=1}^{M-1}\binom{M-1}{i}K^{M-i}(N-K)^{i-1}}{\sum_{i=0}^{M-1}\binom{M-1}{i}K^{M-1-i}(N-K)^{i}} \tag{2.20}$$

$$= 1 + \frac{\frac{K}{N-K}\sum_{i=1}^{M-1}\binom{M-1}{i}K^{M-1-i}(N-K)^{i}}{N^{M-1}} \tag{2.21}$$

$$= 1 + \frac{\frac{K}{N-K}\left(\sum_{i=0}^{M-1}\binom{M-1}{i}K^{M-1-i}(N-K)^{i} - K^{M-1}\right)}{N^{M-1}} \tag{2.22}$$

$$= 1 + \frac{\frac{K}{N-K}\left(N^{M-1} - K^{M-1}\right)}{N^{M-1}} \tag{2.23}$$

$$= 1 + \frac{K}{N-K}\left(1 - R_c^{M-1}\right) \tag{2.24}$$

$$= \frac{N - KR_c^{M-1}}{N-K} \tag{2.25}$$

$$= \frac{1-R_c^M}{1-R_c} \tag{2.26}$$

Hence, $R = \frac{1-R_c}{1-R_c^M}$. Note that if $K = 1$, our achievable scheme reduces to the one presented in [12]. We note that our scheme inherits all the properties of the scheme in [12], in particular, its optimality over any subset of messages.

2.5 Examples

In this section, we give two explicit examples for our scheme. Without loss of generality, we assume that the desired message is W_1.

2.5.1 (5,3) Code with $M = 2$

Initially, sub-indices of all messages are randomly and independently interleaved. For this case, we will have $M = 2$ rounds and then $K = 3$ repetitions; see Table 2.2. We begin round one by downloading $K^{M-1} = 3$ coded bits for the desired message (message W_1) from every database, e.g., we download $\mathbf{h}_1^T\mathbf{x}_1^{[1]}, \mathbf{h}_1^T\mathbf{x}_2^{[1]}, \mathbf{h}_1^T\mathbf{x}_3^{[1]}$ from database 1, and similarly for databases 2-5 by database symmetry. By message symmetry, we download another 3 coded bits from W_2 from each database. Note that for the undesired message, we group every $K = 3$ databases to download from the same row, e.g., we download $\mathbf{h}_1^T\mathbf{x}_1^{[2]}, \mathbf{h}_2^T\mathbf{x}_1^{[2]}, \mathbf{h}_3^T\mathbf{x}_1^{[2]}$ from databases 1-3, $\mathbf{h}_4^T\mathbf{x}_2^{[2]}, \mathbf{h}_5^T\mathbf{x}_2^{[2]}, \mathbf{h}_1^T\mathbf{x}_2^{[2]}$ from databases 4,5,1, and similarly for the remaining databases. By downloading 3 linearly independent equations for every row, we solve for the interference generated by W_2 and create 5 useful side information rows for round two, which are rows $\mathbf{x}_1^{[2]}$ to $\mathbf{x}_5^{[2]}$ from W_2.

In round two, we download sums of the coded bits from W_1, W_2. Since each of

the rows $\mathbf{x}_1^{[2]}$ to $\mathbf{x}_5^{[2]}$ is decoded from 3 databases, we can exploit these side information to download further coded bits from W_1 in the remaining $N-K=2$ databases that do not participate in decoding this row. For example, we use $\mathbf{x}_1^{[2]}$ in databases 4,5 by downloading the sums $\mathbf{h}_4^T(\mathbf{x}_{19}^{[1]}+\mathbf{x}_1^{[2]})$, and $\mathbf{h}_5^T(\mathbf{x}_{20}^{[1]}+\mathbf{x}_1^{[2]})$ and similarly for the rows $\mathbf{x}_2^{[2]}$ to $\mathbf{x}_5^{[2]}$. This creates extra 10 decodable equations in round two in the form of a sum of the two messages. At this point symmetry exists across databases and within messages, and all the interference from the undesired message W_2 is decoded and exploited. However, until this point, we downloaded one equation from every row of W_1. To reliably decode W_1, we need to repeat the previous steps a total of $K=3$ times by shifting the starting database in a circular pattern, e.g., in repetition 2, we download new equations for the rows $\mathbf{x}_1^{[1]},\mathbf{x}_2^{[1]},\mathbf{x}_3^{[1]}$ from database 2 instead of database 1 in repetition 1, and $\mathbf{x}_4^{[1]},\mathbf{x}_5^{[1]},\mathbf{x}_6^{[1]}$ from database 3 instead of database 2, etc. As a final step, we shuffle the order of the queries to preclude the databases from identifying the message index from the index of the first downloaded bit.

Since we download symmetric amount of W_1, W_2 from each database and their indices are randomly chosen, privacy constraint is satisfied. Since vectors $\mathbf{x}_i^{[2]}, i \in \{1,\cdots,5\}$ are downloaded from K databases, their interference is completely decoded. Hence, they can be canceled from round two. Finally, we repeat the scheme 3 times with circular shifts, every desired row is received from K different databases and hence reliably decoded. The explicit query table is shown in Table 2.2. The retrieval rate in this case is $R=\frac{75}{120}=\frac{5}{8}=\frac{1-\frac{3}{5}}{1-(\frac{3}{5})^2}$.

Table 2.2: PIR for code (5,3) and $M = 2$

		DB1	DB2	DB3	DB4	DB5
repetition 1	round 1	$\mathbf{h}_1^T\mathbf{x}_1^{[1]}$ $\mathbf{h}_1^T\mathbf{x}_2^{[1]}$ $\mathbf{h}_1^T\mathbf{x}_3^{[1]}$ $\mathbf{h}_1^T\mathbf{x}_1^{[2]}$ $\mathbf{h}_1^T\mathbf{x}_2^{[2]}$ $\mathbf{h}_1^T\mathbf{x}_4^{[2]}$	$\mathbf{h}_2^T\mathbf{x}_4^{[1]}$ $\mathbf{h}_2^T\mathbf{x}_5^{[1]}$ $\mathbf{h}_2^T\mathbf{x}_6^{[1]}$ $\mathbf{h}_2^T\mathbf{x}_1^{[2]}$ $\mathbf{h}_2^T\mathbf{x}_3^{[2]}$ $\mathbf{h}_2^T\mathbf{x}_4^{[2]}$	$\mathbf{h}_3^T\mathbf{x}_7^{[1]}$ $\mathbf{h}_3^T\mathbf{x}_8^{[1]}$ $\mathbf{h}_3^T\mathbf{x}_9^{[1]}$ $\mathbf{h}_3^T\mathbf{x}_1^{[2]}$ $\mathbf{h}_3^T\mathbf{x}_3^{[2]}$ $\mathbf{h}_3^T\mathbf{x}_5^{[2]}$	$\mathbf{h}_4^T\mathbf{x}_{10}^{[1]}$ $\mathbf{h}_4^T\mathbf{x}_{11}^{[1]}$ $\mathbf{h}_4^T\mathbf{x}_{12}^{[1]}$ $\mathbf{h}_4^T\mathbf{x}_2^{[2]}$ $\mathbf{h}_4^T\mathbf{x}_3^{[2]}$ $\mathbf{h}_4^T\mathbf{x}_5^{[2]}$	$\mathbf{h}_5^T\mathbf{x}_{13}^{[1]}$ $\mathbf{h}_5^T\mathbf{x}_{14}^{[1]}$ $\mathbf{h}_5^T\mathbf{x}_{15}^{[1]}$ $\mathbf{h}_5^T\mathbf{x}_2^{[2]}$ $\mathbf{h}_5^T\mathbf{x}_4^{[2]}$ $\mathbf{h}_5^T\mathbf{x}_5^{[2]}$
	round 2	$\mathbf{h}_1^T(\mathbf{x}_{16}^{[1]} + \mathbf{x}_3^{[2]})$ $\mathbf{h}_1^T(\mathbf{x}_{21}^{[1]} + \mathbf{x}_5^{[2]})$	$\mathbf{h}_2^T(\mathbf{x}_{17}^{[1]} + \mathbf{x}_2^{[2]})$ $\mathbf{h}_2^T(\mathbf{x}_{22}^{[1]} + \mathbf{x}_5^{[2]})$	$\mathbf{h}_3^T(\mathbf{x}_{18}^{[1]} + \mathbf{x}_2^{[2]})$ $\mathbf{h}_3^T(\mathbf{x}_{23}^{[1]} + \mathbf{x}_4^{[2]})$	$\mathbf{h}_4^T(\mathbf{x}_{19}^{[1]} + \mathbf{x}_1^{[2]})$ $\mathbf{h}_4^T(\mathbf{x}_{24}^{[1]} + \mathbf{x}_4^{[2]})$	$\mathbf{h}_5^T(\mathbf{x}_{20}^{[1]} + \mathbf{x}_1^{[2]})$ $\mathbf{h}_5^T(\mathbf{x}_{25}^{[1]} + \mathbf{x}_3^{[2]})$
repetition 2	round 1	$\mathbf{h}_1^T\mathbf{x}_{13}^{[1]}$ $\mathbf{h}_1^T\mathbf{x}_{14}^{[1]}$ $\mathbf{h}_1^T\mathbf{x}_{15}^{[1]}$ $\mathbf{h}_1^T\mathbf{x}_6^{[2]}$ $\mathbf{h}_1^T\mathbf{x}_7^{[2]}$ $\mathbf{h}_1^T\mathbf{x}_9^{[2]}$	$\mathbf{h}_2^T\mathbf{x}_1^{[1]}$ $\mathbf{h}_2^T\mathbf{x}_2^{[1]}$ $\mathbf{h}_2^T\mathbf{x}_3^{[1]}$ $\mathbf{h}_2^T\mathbf{x}_6^{[2]}$ $\mathbf{h}_2^T\mathbf{x}_8^{[2]}$ $\mathbf{h}_2^T\mathbf{x}_9^{[2]}$	$\mathbf{h}_3^T\mathbf{x}_4^{[1]}$ $\mathbf{h}_3^T\mathbf{x}_5^{[1]}$ $\mathbf{h}_3^T\mathbf{x}_6^{[1]}$ $\mathbf{h}_3^T\mathbf{x}_6^{[2]}$ $\mathbf{h}_3^T\mathbf{x}_8^{[2]}$ $\mathbf{h}_3^T\mathbf{x}_{10}^{[2]}$	$\mathbf{h}_4^T\mathbf{x}_7^{[1]}$ $\mathbf{h}_4^T\mathbf{x}_8^{[1]}$ $\mathbf{h}_4^T\mathbf{x}_9^{[1]}$ $\mathbf{h}_4^T\mathbf{x}_7^{[2]}$ $\mathbf{h}_4^T\mathbf{x}_8^{[2]}$ $\mathbf{h}_4^T\mathbf{x}_{10}^{[2]}$	$\mathbf{h}_5^T\mathbf{x}_{10}^{[1]}$ $\mathbf{h}_5^T\mathbf{x}_{11}^{[1]}$ $\mathbf{h}_5^T\mathbf{x}_{12}^{[1]}$ $\mathbf{h}_5^T\mathbf{x}_7^{[2]}$ $\mathbf{h}_5^T\mathbf{x}_9^{[2]}$ $\mathbf{h}_5^T\mathbf{x}_{10}^{[2]}$
	round 2	$\mathbf{h}_1^T(\mathbf{x}_{20}^{[1]} + \mathbf{x}_8^{[2]})$ $\mathbf{h}_1^T(\mathbf{x}_{25}^{[1]} + \mathbf{x}_{10}^{[2]})$	$\mathbf{h}_2^T(\mathbf{x}_{16}^{[1]} + \mathbf{x}_7^{[2]})$ $\mathbf{h}_2^T(\mathbf{x}_{21}^{[1]} + \mathbf{x}_{10}^{[2]})$	$\mathbf{h}_3^T(\mathbf{x}_{17}^{[1]} + \mathbf{x}_7^{[2]})$ $\mathbf{h}_3^T(\mathbf{x}_{22}^{[1]} + \mathbf{x}_9^{[2]})$	$\mathbf{h}_4^T(\mathbf{x}_{18}^{[1]} + \mathbf{x}_6^{[2]})$ $\mathbf{h}_4^T(\mathbf{x}_{23}^{[1]} + \mathbf{x}_9^{[2]})$	$\mathbf{h}_5^T(\mathbf{x}_{19}^{[1]} + \mathbf{x}_6^{[2]})$ $\mathbf{h}_5^T(\mathbf{x}_{24}^{[1]} + \mathbf{x}_8^{[2]})$
repetition 3	round 1	$\mathbf{h}_1^T\mathbf{x}_{10}^{[1]}$ $\mathbf{h}_1^T\mathbf{x}_{11}^{[1]}$ $\mathbf{h}_1^T\mathbf{x}_{12}^{[1]}$ $\mathbf{h}_1^T\mathbf{x}_{11}^{[2]}$ $\mathbf{h}_1^T\mathbf{x}_{12}^{[2]}$ $\mathbf{h}_1^T\mathbf{x}_{14}^{[2]}$	$\mathbf{h}_2^T\mathbf{x}_{13}^{[1]}$ $\mathbf{h}_2^T\mathbf{x}_{14}^{[1]}$ $\mathbf{h}_2^T\mathbf{x}_{15}^{[1]}$ $\mathbf{h}_2^T\mathbf{x}_{11}^{[2]}$ $\mathbf{h}_2^T\mathbf{x}_{13}^{[2]}$ $\mathbf{h}_2^T\mathbf{x}_{14}^{[2]}$	$\mathbf{h}_3^T\mathbf{x}_1^{[1]}$ $\mathbf{h}_3^T\mathbf{x}_2^{[1]}$ $\mathbf{h}_3^T\mathbf{x}_3^{[1]}$ $\mathbf{h}_3^T\mathbf{x}_{11}^{[2]}$ $\mathbf{h}_3^T\mathbf{x}_{13}^{[2]}$ $\mathbf{h}_{13}^T\mathbf{x}_{15}^{[2]}$	$\mathbf{h}_4^T\mathbf{x}_4^{[1]}$ $\mathbf{h}_4^T\mathbf{x}_5^{[1]}$ $\mathbf{h}_4^T\mathbf{x}_6^{[1]}$ $\mathbf{h}_4^T\mathbf{x}_{12}^{[2]}$ $\mathbf{h}_4^T\mathbf{x}_{13}^{[2]}$ $\mathbf{h}_4^T\mathbf{x}_{15}^{[2]}$	$\mathbf{h}_5^T\mathbf{x}_7^{[1]}$ $\mathbf{h}_5^T\mathbf{x}_8^{[1]}$ $\mathbf{h}_5^T\mathbf{x}_9^{[1]}$ $\mathbf{h}_5^T\mathbf{x}_{12}^{[2]}$ $\mathbf{h}_5^T\mathbf{x}_{14}^{[2]}$ $\mathbf{h}_5^T\mathbf{x}_{15}^{[2]}$
	round 2	$\mathbf{h}_1^T(\mathbf{x}_{19}^{[1]} + \mathbf{x}_{13}^{[2]})$ $\mathbf{h}_1^T(\mathbf{x}_{24}^{[1]} + \mathbf{x}_{15}^{[2]})$	$\mathbf{h}_2^T(\mathbf{x}_{20}^{[1]} + \mathbf{x}_{12}^{[2]})$ $\mathbf{h}_2^T(\mathbf{x}_{25}^{[1]} + \mathbf{x}_{15}^{[2]})$	$\mathbf{h}_3^T(\mathbf{x}_{16}^{[1]} + \mathbf{x}_{12}^{[2]})$ $\mathbf{h}_3^T(\mathbf{x}_{21}^{[1]} + \mathbf{x}_{14}^{[2]})$	$\mathbf{h}_4^T(\mathbf{x}_{17}^{[1]} + \mathbf{x}_{11}^{[2]})$ $\mathbf{h}_4^T(\mathbf{x}_{22}^{[1]} + \mathbf{x}_{14}^{[2]})$	$\mathbf{h}_5^T(\mathbf{x}_{18}^{[1]} + \mathbf{x}_{11}^{[2]})$ $\mathbf{h}_5^T(\mathbf{x}_{23}^{[1]} + \mathbf{x}_{13}^{[2]})$

2.5.2 (3,2) Code with $M = 3$

As in the previous example, the messages are randomly and independently interleaved. For this case, the scheme is completed in $M = 3$ rounds and then repeated for $K = 2$ repetitions, see Table 2.3. In the first round, we download $K^{M-1} = 4$ coded bits for W_1 from each database, e.g., $\mathbf{h}_1^T \mathbf{x}_i^{[1]}, i \in \{1, \cdots, 4\}$ from the first database. Similarly, we download one equation from the rows $\mathbf{x}_1^{[1]}$ to $\mathbf{x}_{12}^{[1]}$ by applying the database symmetry. We apply message symmetry to download $N\binom{M-1}{1}K^{M-1} = 24$ undesired coded bits from W_2, W_3. Every 2 coded bits from the undesired bits are grouped together to generate single solved side information vector, e.g., we download as $\mathbf{h}_1^T \mathbf{x}_1^{[2]}, \mathbf{h}_2^T \mathbf{x}_2^{[2]}$ from databases 1,2, $\mathbf{h}_3^T \mathbf{x}_2^{[2]}, \mathbf{h}_1^T \mathbf{x}_1^{[2]}$ from databases 3,1, and similarly for rows $\mathbf{x}_1^{[m]}$ to $\mathbf{x}_6^{[m]}$ where $m = 2, 3$. Hence, we have $N\binom{M-1}{1}K^{M-2} = 12$ side information rows to be used in round two.

In round two, we download sums of every two messages. We exploit the generated side information within the $N - K = 1$ remaining database that does not participate in generating them. For example, we decoded $\mathbf{x}_1^{[2]}$ by downloading equations from databases 1,2, then we use $\mathbf{x}_1^{[2]}$ in database 3 by downloading the sum $\mathbf{h}_3(\mathbf{x}_{15}^{[1]} + \mathbf{x}_1^{[2]})$. Hence, we can download $N\binom{M-1}{1}K^{M-2}(N-K) = 12$ new coded bits of W_1 by using every decoded side information in a sum of W_1 with one of W_2 or W_3. These bits are reliably decoded, since the generated side information can be canceled from the downloaded equation. It remains to add sums of W_2 and W_3 to ensure the privacy. Therefore, we download $N\binom{M-1}{2}K^{M-2}(N-K) = 6$ undesired equations, that will be grouped further to form $N\binom{M-1}{2}K^{M-3}(N-K) = 3$ solved

side information equations in the form of sums of W_2 and W_3. As an example, we download $\mathbf{h}_1^T(\mathbf{x}_7^{[2]} + \mathbf{x}_7^{[3]}), \mathbf{h}_2^T(\mathbf{x}_7^{[2]} + \mathbf{x}_7^{[3]})$ from databases 1,2. In this case the interference from the rows $\mathbf{x}_7^{[2]} + \mathbf{x}_7^{[3]}$ is decoded. Note that we do not solve for the individual $\mathbf{x}_7^{[2]}$ or $\mathbf{x}_7^{[3]}$ but we *align* them in the same subspace, and solve for their sum.

In round three, we use the newly generated side information, e.g., $\mathbf{x}_7^{[2]} + \mathbf{x}_7^{[3]}$, to download extra $N\binom{M-1}{2}K^{M-3}(N-K)^2 = 3$ desired coded bits in the form of sum of three terms, e.g., $\mathbf{h}_3^T(\mathbf{x}_{27}^{[1]} + \mathbf{x}_7^{[2]} + \mathbf{x}_7^{[3]})$. Finally, the previous steps are repeated $K = 2$ times to reliably decode W_1 and the queries are shuffled for privacy. The retrieval rate in this case is $R = \frac{54}{114} = \frac{9}{19} = \frac{1-\frac{2}{3}}{1-(\frac{2}{3})^3}$. The explicit query structure is shown in Table 2.3.

2.6 Converse Proof

In this section, we prove the converse for PIR from MDS-coded databases. The proof extends the techniques in [12] to the case of MDS-coded databases. The proof presented here does not use symmetrization arguments or fixing of an individual query as in the conference version [116], which presents an alternative proof that provides an alternative perspective.

We need the following lemma which states that in the PIR problem from (N, K) MDS-coded databases, the answers from any K databases are statistically independent.

Lemma 2.1 (Independence of answers of any K databases) *In the PIR*

Table 2.3: PIR for code (3,2) and $M = 3$

		DB1	DB2	DB3
repetition 1	round 1	$\mathbf{h}_1^T\mathbf{x}_1^{[1]}$	$\mathbf{h}_2^T\mathbf{x}_5^{[1]}$	$\mathbf{h}_3^T\mathbf{x}_9^{[1]}$
		$\mathbf{h}_1^T\mathbf{x}_2^{[1]}$	$\mathbf{h}_2^T\mathbf{x}_6^{[1]}$	$\mathbf{h}_3^T\mathbf{x}_{10}^{[1]}$
		$\mathbf{h}_1^T\mathbf{x}_3^{[1]}$	$\mathbf{h}_2^T\mathbf{x}_7^{[1]}$	$\mathbf{h}_3^T\mathbf{x}_{11}^{[1]}$
		$\mathbf{h}_1^T\mathbf{x}_4^{[1]}$	$\mathbf{h}_2^T\mathbf{x}_8^{[1]}$	$\mathbf{h}_3^T\mathbf{x}_{12}^{[1]}$
		$\mathbf{h}_1^T\mathbf{x}_1^{[2]}$	$\mathbf{h}_2^T\mathbf{x}_1^{[2]}$	$\mathbf{h}_3^T\mathbf{x}_2^{[2]}$
		$\mathbf{h}_1^T\mathbf{x}_2^{[2]}$	$\mathbf{h}_2^T\mathbf{x}_3^{[2]}$	$\mathbf{h}_3^T\mathbf{x}_3^{[2]}$
		$\mathbf{h}_1^T\mathbf{x}_4^{[2]}$	$\mathbf{h}_2^T\mathbf{x}_4^{[2]}$	$\mathbf{h}_3^T\mathbf{x}_5^{[2]}$
		$\mathbf{h}_1^T\mathbf{x}_5^{[2]}$	$\mathbf{h}_2^T\mathbf{x}_6^{[2]}$	$\mathbf{h}_3^T\mathbf{x}_6^{[2]}$
		$\mathbf{h}_1^T\mathbf{x}_1^{[3]}$	$\mathbf{h}_2^T\mathbf{x}_1^{[3]}$	$\mathbf{h}_3^T\mathbf{x}_2^{[3]}$
		$\mathbf{h}_1^T\mathbf{x}_2^{[3]}$	$\mathbf{h}_2^T\mathbf{x}_3^{[3]}$	$\mathbf{h}_3^T\mathbf{x}_3^{[3]}$
		$\mathbf{h}_1^T\mathbf{x}_4^{[3]}$	$\mathbf{h}_2^T\mathbf{x}_4^{[3]}$	$\mathbf{h}_3^T\mathbf{x}_5^{[3]}$
		$\mathbf{h}_1^T\mathbf{x}_5^{[3]}$	$\mathbf{h}_2^T\mathbf{x}_6^{[3]}$	$\mathbf{h}_3^T\mathbf{x}_6^{[3]}$
	round 2	$\mathbf{h}_1^T(\mathbf{x}_{13}^{[1]} + \mathbf{x}_3^{[2]})$	$\mathbf{h}_2^T(\mathbf{x}_{14}^{[1]} + \mathbf{x}_2^{[2]})$	$\mathbf{h}_3^T(\mathbf{x}_{15}^{[1]} + \mathbf{x}_1^{[2]})$
		$\mathbf{h}_1^T(\mathbf{x}_{16}^{[1]} + \mathbf{x}_3^{[3]})$	$\mathbf{h}_2^T(\mathbf{x}_{17}^{[1]} + \mathbf{x}_2^{[3]})$	$\mathbf{h}_3^T(\mathbf{x}_{18}^{[1]} + \mathbf{x}_1^{[3]})$
		$\mathbf{h}_1^T(\mathbf{x}_7^{[2]} + \mathbf{x}_7^{[3]})$	$\mathbf{h}_2^T(\mathbf{x}_7^{[2]} + \mathbf{x}_7^{[3]})$	$\mathbf{h}_3^T(\mathbf{x}_8^{[2]} + \mathbf{x}_8^{[3]})$
		$\mathbf{h}_1^T(\mathbf{x}_{19}^{[1]} + \mathbf{x}_6^{[2]})$	$\mathbf{h}_2^T(\mathbf{x}_{20}^{[1]} + \mathbf{x}_5^{[2]})$	$\mathbf{h}_3^T(\mathbf{x}_{21}^{[1]} + \mathbf{x}_4^{[2]})$
		$\mathbf{h}_1^T(\mathbf{x}_{22}^{[1]} + \mathbf{x}_6^{[3]})$	$\mathbf{h}_2^T(\mathbf{x}_{23}^{[1]} + \mathbf{x}_5^{[3]})$	$\mathbf{h}_3^T(\mathbf{x}_{24}^{[1]} + \mathbf{x}_4^{[3]})$
		$\mathbf{h}_1^T(\mathbf{x}_8^{[2]} + \mathbf{x}_8^{[3]})$	$\mathbf{h}_2^T(\mathbf{x}_9^{[2]} + \mathbf{x}_9^{[3]})$	$\mathbf{h}_3^T(\mathbf{x}_9^{[2]} + \mathbf{x}_9^{[3]})$
	rd. 3	$\mathbf{h}_1^T(\mathbf{x}_{25}^{[1]} + \mathbf{x}_9^{[2]} + \mathbf{x}_9^{[3]})$	$\mathbf{h}_2^T(\mathbf{x}_{26}^{[1]} + \mathbf{x}_8^{[2]} + \mathbf{x}_8^{[3]})$	$\mathbf{h}_3^T(\mathbf{x}_{27}^{[1]} + \mathbf{x}_7^{[2]} + \mathbf{x}_7^{[3]})$
repetition 2	round 1	$\mathbf{h}_1^T\mathbf{x}_9^{[1]}$	$\mathbf{h}_2^T\mathbf{x}_1^{[1]}$	$\mathbf{h}_3^T\mathbf{x}_5^{[1]}$
		$\mathbf{h}_1^T\mathbf{x}_{10}^{[1]}$	$\mathbf{h}_2^T\mathbf{x}_2^{[1]}$	$\mathbf{h}_3^T\mathbf{x}_6^{[1]}$
		$\mathbf{h}_1^T\mathbf{x}_{11}^{[1]}$	$\mathbf{h}_2^T\mathbf{x}_3^{[1]}$	$\mathbf{h}_3^T\mathbf{x}_7^{[1]}$
		$\mathbf{h}_1^T\mathbf{x}_{12}^{[1]}$	$\mathbf{h}_2^T\mathbf{x}_4^{[1]}$	$\mathbf{h}_3^T\mathbf{x}_8^{[1]}$
		$\mathbf{h}_1^T\mathbf{x}_{10}^{[2]}$	$\mathbf{h}_2^T\mathbf{x}_{10}^{[2]}$	$\mathbf{h}_3^T\mathbf{x}_{11}^{[2]}$
		$\mathbf{h}_1^T\mathbf{x}_{11}^{[2]}$	$\mathbf{h}_2^T\mathbf{x}_{12}^{[2]}$	$\mathbf{h}_3^T\mathbf{x}_{12}^{[2]}$
		$\mathbf{h}_1^T\mathbf{x}_{13}^{[2]}$	$\mathbf{h}_2^T\mathbf{x}_{13}^{[2]}$	$\mathbf{h}_3^T\mathbf{x}_{14}^{[2]}$
		$\mathbf{h}_1^T\mathbf{x}_{14}^{[2]}$	$\mathbf{h}_2^T\mathbf{x}_{15}^{[2]}$	$\mathbf{h}_3^T\mathbf{x}_{15}^{[2]}$
		$\mathbf{h}_1^T\mathbf{x}_{10}^{[3]}$	$\mathbf{h}_2^T\mathbf{x}_{10}^{[3]}$	$\mathbf{h}_3^T\mathbf{x}_{11}^{[3]}$
		$\mathbf{h}_1^T\mathbf{x}_{11}^{[3]}$	$\mathbf{h}_2^T\mathbf{x}_{12}^{[3]}$	$\mathbf{h}_3^T\mathbf{x}_{12}^{[3]}$
		$\mathbf{h}_1^T\mathbf{x}_{13}^{[3]}$	$\mathbf{h}_2^T\mathbf{x}_{13}^{[3]}$	$\mathbf{h}_3^T\mathbf{x}_{14}^{[3]}$
		$\mathbf{h}_1^T\mathbf{x}_{14}^{[3]}$	$\mathbf{h}_2^T\mathbf{x}_{15}^{[3]}$	$\mathbf{h}_3^T\mathbf{x}_{15}^{[3]}$
	round 2	$\mathbf{h}_1^T(\mathbf{x}_{15}^{[1]} + \mathbf{x}_{12}^{[2]})$	$\mathbf{h}_2^T(\mathbf{x}_{13}^{[1]} + \mathbf{x}_{11}^{[2]})$	$\mathbf{h}_3^T(\mathbf{x}_{14}^{[1]} + \mathbf{x}_{10}^{[2]})$
		$\mathbf{h}_1^T(\mathbf{x}_{18}^{[1]} + \mathbf{x}_{12}^{[3]})$	$\mathbf{h}_2^T(\mathbf{x}_{16}^{[1]} + \mathbf{x}_{11}^{[3]})$	$\mathbf{h}_3^T(\mathbf{x}_{17}^{[1]} + \mathbf{x}_{10}^{[3]})$
		$\mathbf{h}_1^T(\mathbf{x}_{16}^{[2]} + \mathbf{x}_{16}^{[3]})$	$\mathbf{h}_2^T(\mathbf{x}_{16}^{[2]} + \mathbf{x}_{16}^{[3]})$	$\mathbf{h}_3^T(\mathbf{x}_{17}^{[2]} + \mathbf{x}_{17}^{[3]})$
		$\mathbf{h}_1^T(\mathbf{x}_{21}^{[1]} + \mathbf{x}_{15}^{[2]})$	$\mathbf{h}_2^T(\mathbf{x}_{19}^{[1]} + \mathbf{x}_{14}^{[2]})$	$\mathbf{h}_3^T(\mathbf{x}_{20}^{[1]} + \mathbf{x}_{13}^{[2]})$
		$\mathbf{h}_1^T(\mathbf{x}_{24}^{[1]} + \mathbf{x}_{15}^{[3]})$	$\mathbf{h}_2^T(\mathbf{x}_{22}^{[1]} + \mathbf{x}_{14}^{[3]})$	$\mathbf{h}_3^T(\mathbf{x}_{23}^{[1]} + \mathbf{x}_{13}^{[3]})$
		$\mathbf{h}_1^T(\mathbf{x}_{17}^{[2]} + \mathbf{x}_{17}^{[3]})$	$\mathbf{h}_2^T(\mathbf{x}_{18}^{[2]} + \mathbf{x}_{18}^{[3]})$	$\mathbf{h}_3^T(\mathbf{x}_{18}^{[2]} + \mathbf{x}_{18}^{[3]})$
	rd. 3	$\mathbf{h}_1^T(\mathbf{x}_{27}^{[1]} + \mathbf{x}_{18}^{[2]} + \mathbf{x}_{18}^{[3]})$	$\mathbf{h}_2^T(\mathbf{x}_{25}^{[1]} + \mathbf{x}_{17}^{[2]} + \mathbf{x}_{17}^{[3]})$	$\mathbf{h}_3^T(\mathbf{x}_{26}^{[1]} + \mathbf{x}_{16}^{[2]} + \mathbf{x}_{16}^{[3]})$

problem from (N, K) MDS-coded databases, for any set $\mathcal{K}$ of databases such that $|\mathcal{K}| = K$,

$$H(A_{\mathcal{K}}^{[m]}|Q_{\mathcal{K}}^{[m]}) = \sum_{n\in\mathcal{K}} H(A_n^{[m]}|Q_n^{[m]}), \qquad m \in \{1, \cdots, M\} \tag{2.27}$$

Furthermore, (2.27) is true if conditioned on any subset of messages $W_{\mathcal{S}}$, i.e.,

$$H(A_{\mathcal{K}}^{[m]}|Q_{\mathcal{K}}^{[m]}, W_{\mathcal{S}}) = \sum_{n\in\mathcal{K}} H(A_n^{[m]}|Q_n^{[m]}, W_{\mathcal{S}}), \qquad m \in \{1, \cdots, M\} \tag{2.28}$$

Proof: Consider a set of databases $\mathcal{K}$ such that $|\mathcal{K}| = K$. We prove first the statistical independence between the vectors $\{\mathbf{y}_n, n \in \mathcal{K}\}$ where $\mathbf{y}_n$ represents the contents of the nth database. The contents of set $\mathcal{K}$ of databases can be written as

$$[\mathbf{y}_n,\ n \in \mathcal{K}] = \begin{bmatrix} W_1 \\ \vdots \\ W_M \end{bmatrix} [\mathbf{h}_n,\ n \in \mathcal{K}] = \begin{bmatrix} W_1 \\ \vdots \\ W_M \end{bmatrix} \mathbf{H}_{\mathcal{K}} \tag{2.29}$$

where $\mathbf{H}_{\mathcal{K}} = [\mathbf{h}_n,\ n \in \mathcal{K}]$ is a $\mathbb{F}_q^{K\times K}$ matrix. By construction of the distributed storage code, the matrix $\mathbf{H}_{\mathcal{K}}$ is an invertible matrix. Using [14, Lemma 1] and the fact that elements of the messages are chosen independently and uniformly over

$\mathbb{F}_q^{\tilde{L}\times K}$, we conclude that

$$[\mathbf{y}_n,\ n\in\mathcal{K}] = \begin{bmatrix} W_1 \\ \vdots \\ W_M \end{bmatrix} \mathbf{H}_{\mathcal{K}} \sim \begin{bmatrix} W_1 \\ \vdots \\ W_M \end{bmatrix} \tag{2.30}$$

where $A \sim B$ denotes that random variables A and B are identically distributed. Therefore, the contents of the databases are statistically equivalent to the messages. Hence, the columns of $[\mathbf{y}_n,\ n\in\mathcal{K}]$ are statistically independent since the elements of the messages are independent.

Since $A_n^{[m]}, n\in\mathcal{K}$ are deterministic functions of $(\mathbf{y}_n, Q_n^{[m]})$, $\{A_n^{[m]} : n\in\mathcal{K}\}$ are statistically independent as they are deterministic functions of independent random variables. Therefore, if $\mathcal{K}=\{n_1, n_2, \cdots, n_K\}$

$$H(A_{\mathcal{K}}^{[m]}|Q_{\mathcal{K}}^{[m]}) = \sum_{i=1}^{K} H(A_{n_i}^{[m]}|A_{1:n_{i-1}}^{[m]}, Q_{\mathcal{K}}^{[m]}) \tag{2.31}$$

$$= \sum_{i=1}^{K} H(A_{n_i}^{[m]}|Q_{\mathcal{K}}^{[m]}) \tag{2.32}$$

$$= \sum_{n\in\mathcal{K}} H(A_n^{[m]}|Q_n^{[m]}) \tag{2.33}$$

where (2.32) follows from the independence of any K answer strings, (2.33) follows from the fact that $Q_{\mathcal{K}}^{[m]} \to Q_n^{[m]} \to A_n^{[m]}$ is a Markov chain. We note that since coding is applied on individual messages, conditioning on any subset of messages $W_{\mathcal{S}}$ with $|W_{\mathcal{S}}| = S$ is equivalent to reducing the problem to storing $M-S$ independent messages instead of M messages. Hence, the statistical independence argument in

(2.28) follows as before. ■

We use Han's inequality [47, Theorem 17.6.1] in a similar way to [14].

Lemma 2.2 (Han's inequality) *Let $\mathcal{K} \subseteq \{1, \cdots, N\}$, such that $|\mathcal{K}| = K$. Then, for any subset of messages $W_{\mathcal{S}}$,*

$$\frac{1}{\binom{N}{K}} \sum_{\mathcal{K}:|\mathcal{K}|=K} H(A_{\mathcal{K}}^{[m]}|W_{\mathcal{S}}, Q_{1:N}^{[m]}) \geq \frac{K}{N} H(A_{1:N}^{[m]}|W_{\mathcal{S}}, Q_{1:N}^{[m]}) \tag{2.34}$$

The following lemma characterizes a lower bound on the interference components in $A_{1:N}^{[1]}$ that result from the interfering messages $W_{2:M}$ which is represented by $\frac{L}{R} - L$. The following lemma is exactly [12, Lemma 5]. The result does not change due to the distributed storage code introduced in our problem. We include the proof of this lemma here for completeness.

Lemma 2.3 (Interference lower bound) *The interference from undesired messages within the answer strings, $\frac{L}{R} - L$, is lower bounded by,*

$$L\left(\frac{1}{R} - 1 + \frac{o(L)}{L}\right) \geq I(W_{2:M}; Q_{1:N}^{[1]}, A_{1:N}^{[1]}|W_1) \tag{2.35}$$

Proof: We start with the right hand side of (2.35),

$$\begin{aligned} I(W_{2:M}; Q_{1:N}^{[1]}, A_{1:N}^{[1]}|W_1) =& I(W_{2:M}; Q_{1:N}^{[1]}, A_{1:N}^{[1]}, W_1) && (2.36)\\ =& I(W_{2:M}; Q_{1:N}^{[1]}, A_{1:N}^{[1]}) + I(W_{2:M}; W_1|Q_{1:N}^{[1]}, A_{1:N}^{[1]}) && (2.37)\\ =& I(W_{2:M}; Q_{1:N}^{[1]}) + I(W_{2:M}; A_{1:N}^{[1]}|Q_{1:N}^{[1]}) + o(L) && (2.38) \end{aligned}$$

$$=I(W_{2:M}; A^{[1]}_{1:N}|Q^{[1]}_{1:N}) + o(L) \tag{2.39}$$

$$=H(A^{[1]}_{1:N}|Q^{[1]}_{1:N}) - H(A^{[1]}_{1:N}|Q^{[1]}_{1:N}, W_{2:M}) + o(L) \tag{2.40}$$

$$\leq \sum_{n=1}^{N} H(A^{[1]}_n) - H(W_1, A^{[1]}_{1:N}|Q^{[1]}_{1:N}, W_{2:M})$$

$$+ H(W_1|Q^{[1]}_{1:N}, A^{[1]}_{1:N}, W_{2:M}) + o(L) \tag{2.41}$$

$$=\frac{L}{R} - H(W_1|Q^{[1]}_{1:N}, W_{2:M}) - H(A^{[1]}_{1:N}|Q^{[1]}_{1:N}, W_{1:M}) + o(L) \tag{2.42}$$

$$=\frac{L}{R} - L + o(L) \tag{2.43}$$

$$=L\left(\frac{1}{R} - 1 + \frac{o(L)}{L}\right) \tag{2.44}$$

where (2.36) follows from the independence of messages, (2.38) and (2.42) follow from the decodability of W_1 from $(Q^{[1]}_{1:N}, A^{[1]}_{1:N})$, (2.39) follows from the independence of the queries $Q^{[1]}_{1:N}$ and the messages $W_{2:M}$, (2.41) follows from the fact that conditioning reduces entropy, and (2.43) follows from the fact that the answers $A^{[1]}_{1:N}$ are deterministic functions of $(Q^{[1]}_{1:N}, W_{1:M})$ and the independence of $(W_1, Q^{[1]}_{1:N}, W_{2:M})$.

■

In the following lemma, we prove an inductive relation for the mutual information term on the right hand side of (2.35).

Lemma 2.4 (Induction lemma) *We have the following inductive relationship,*

$$I(W_{m:M}; Q^{[m-1]}_{1:N}, A^{[m-1]}_{1:N}|W_{1:m-1}) \geq \frac{K}{N} I(W_{m+1:M}; Q^{[m]}_{1:N}, A^{[m]}_{1:N}|W_{1:m}) + \frac{KL\left(1 - \frac{o(L)}{L}\right)}{N} \tag{2.45}$$

Proof: We start with the left hand side of (2.45),

$$
\begin{aligned}
&I(W_{m:M};Q_{1:N}^{[m-1]}, A_{1:N}^{[m-1]}|W_{1:m-1}) \\
&\quad \geq \frac{1}{\binom{N}{K}} \sum_{\mathcal{K}:|\mathcal{K}|=K} I(W_{m:M}; Q_{\mathcal{K}}^{[m-1]}, A_{\mathcal{K}}^{[m-1]}|W_{1:m-1}) && (2.46)\\
&\quad = \frac{1}{\binom{N}{K}} \sum_{\mathcal{K}:|\mathcal{K}|=K} I(W_{m:M}; A_{\mathcal{K}}^{[m-1]}|W_{1:m-1}, Q_{\mathcal{K}}^{[m-1]}) && (2.47)\\
&\quad = \frac{1}{\binom{N}{K}} \sum_{\mathcal{K}:|\mathcal{K}|=K} H(A_{\mathcal{K}}^{[m-1]}|W_{1:m-1}, Q_{\mathcal{K}}^{[m-1]}) && (2.48)\\
&\quad = \frac{1}{\binom{N}{K}} \sum_{\mathcal{K}:|\mathcal{K}|=K} \sum_{n\in\mathcal{K}} H(A_{n}^{[m-1]}|W_{1:m-1}, Q_{n}^{[m-1]}) && (2.49)\\
&\quad = \frac{1}{\binom{N}{K}} \sum_{\mathcal{K}:|\mathcal{K}|=K} \sum_{n\in\mathcal{K}} H(A_{n}^{[m]}|W_{1:m-1}, Q_{n}^{[m]}) && (2.50)\\
&\quad = \frac{1}{\binom{N}{K}} \sum_{\mathcal{K}:|\mathcal{K}|=K} H(A_{\mathcal{K}}^{[m]}|W_{1:m-1}, Q_{\mathcal{K}}^{[m]}) && (2.51)\\
&\quad \geq \frac{1}{\binom{N}{K}} \sum_{\mathcal{K}:|\mathcal{K}|=K} H(A_{\mathcal{K}}^{[m]}|W_{1:m-1}, Q_{1:N}^{[m]}) && (2.52)\\
&\quad \geq \frac{K}{N} H(A_{1:N}^{[m]}|W_{1:m-1}, Q_{1:N}^{[m]}) && (2.53)\\
&\quad = \frac{K}{N} I(W_{m:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]}|W_{1:m-1}) && (2.54)\\
&\quad = \frac{K}{N} \Big[I(W_{m:M}; W_m, Q_{1:N}^{[m]}, A_{1:N}^{[m]}|W_{1:m-1}) - o(L) \Big] && (2.55)\\
&\quad = \frac{K}{N} \Big[I(W_{m:M}; W_m|W_{1:m-1}) + I(W_{m:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]}|W_{1:m}) - o(L) \Big] && (2.56)\\
&\quad = \frac{K}{N} \Big[L + I(W_{m+1:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]}|W_{1:m}) - o(L) \Big] && (2.57)\\
&\quad = \frac{K}{N} I(W_{m+1:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]}|W_{1:m}) + \frac{KL\left(1 - \frac{o(L)}{L}\right)}{N} && (2.58)
\end{aligned}
$$

where (2.46) follows from the fact that for every subset $\mathcal{K}$ such that $|\mathcal{K}| = K$ we have $I(W_{m:M}; Q_{1:N}^{[m-1]}, A_{1:N}^{[m-1]}|W_{1:m-1}) \geq I(W_{m:M}; Q_{\mathcal{K}}^{[m-1]}, A_{\mathcal{K}}^{[m-1]}|W_{1:m-1})$ by the non-

negativity of the mutual information, (2.47) follows from the independence of the messages and the queries, (2.48) follows from the fact that the answers $A_{\mathcal{K}}^{[m-1]}$ are deterministic functions of $(W_{1:M}, Q_{\mathcal{K}}^{[m-1]})$, (2.49) and (2.51) follow from the independence of any K answers as a consequence of Lemma 2.1, (2.50) follows from the privacy constraint, (2.52) follows from conditioning reduces entropy, (2.53) follows from Han's inequality in Lemma 2.2, (2.54) follows from the fact that $A_{1:N}^{[m]}$ is a deterministic function of $(W_{1:M}, Q_{1:N}^{[m]})$ and the independence of the messages and the queries, (2.55) follows from the decodability of W_m from $(Q_{1:N}^{[m]}, A_{1:N}^{[m]})$, and (2.57) follows from $I(W_{m:M}; W_m|W_{1:m-1}) = H(W_m) = L$ from the independence of the messages. ■

Now, we are ready to complete the converse proof by applying Lemma 2.3 and Lemma 2.4 successively. We have

$$
\begin{aligned}
L\left(\frac{1}{R} - 1 + \frac{o(L)}{L}\right) &\geq I(W_{2:M}; Q_{1:N}^{[1]}, A_{1:N}^{[1]}|W_1) && (2.59)\\
&\geq \frac{K}{N} I(W_{3:M}; Q_{1:N}^{[2]}, A_{1:N}^{[2]}|W_{1:2}) + \frac{KL\left(1 - \frac{o(L)}{L}\right)}{N} && (2.60)\\
&\geq \cdots && (2.61)\\
&\geq \frac{K^{M-2}}{N^{M-2}} I(W_{M:M}; Q_{1:N}^{[M-1]}, A_{1:N}^{[M-1]}|W_{1:M-1}) \\
&\quad + \left(\frac{K}{N} + \frac{K^2}{N^2} + \cdots + \frac{K^{M-2}}{N^{M-2}}\right)\left(1 - \frac{o(L)}{L}\right) L && (2.62)\\
&\geq \left(\frac{K}{N} + \frac{K^2}{N^2} + \cdots + \frac{K^{M-1}}{N^{M-1}}\right)\left(1 - \frac{o(L)}{L}\right) L && (2.63)
\end{aligned}
$$

where (2.59) follows from Lemma 2.3, and (2.60)-(2.63) follow from applying

Lemma 2.4 successively for $M-1$ times. Hence, we have

$$\frac{1}{R} \geq \left(1+\frac{K}{N}+\frac{K^2}{N^2}+\cdots+\frac{K^{M-1}}{N^{M-1}}\right)\left(1-\frac{o(L)}{L}\right) \tag{2.64}$$

By taking $L \to \infty$, and noting $\frac{o(L)}{L} \to 0$, we have

$$R \leq \frac{1}{\sum_{i=0}^{M-1}\left(\frac{K}{N}\right)^i} \tag{2.65}$$

$$= \frac{1}{\sum_{i=0}^{M-1} R_c^i} = \frac{1-R_c}{1-R_c^M} \tag{2.66}$$

Remark 2.1 *In the conference version of this work [116], we presented a different converse proof. In this remark, we briefly describe this alternative proof for a more complete and insightful exposition. The converse proof in [116] assumes without loss of generality that the answer strings are symmetric across messages and databases, and an individual answer string (e.g., A_1) can be the same no matter what the desired message is. The converse proof is obtained by induction over M. We start the proof by considering the case of $M=2$ messages as a base induction step. In this case, we derive a lower bound on the interference from W_2 to be [116, Lemma 3],*

$$H(A_{1:N}^{[1]}|W_1, \mathcal{Q}) \geq \frac{KL}{N} \tag{2.67}$$

where $\mathcal{Q} \triangleq \{Q_n^{[m]} : m \in \{1,\cdots,M\},\ n \in \{1,\cdots,N\}\}$. From [116, Lemma 3], we prove that $R \leq \frac{1}{1+\frac{K}{N}}$ for $M=2$, which proves the base induction step. For any M, we prove that the remaining uncertainty on the answer strings after conditioning on

one of the interfering messages is upper bounded by [116, Lemma 4],

$$H(A_{1:N}^{[2]}|W_1, \mathcal{Q}) \leq \frac{N}{K}\left(NH(A_1|\mathcal{Q}) - L\right) \tag{2.68}$$

Consequently, we obtain an inductive relation for any M as,

$$NH(A_1|\mathcal{Q}) \geq \left(1 + \frac{K}{N}\right) L + \frac{K^2}{N} H(A_1|W_1, W_2, \mathcal{Q}) \tag{2.69}$$

Using the induction hypothesis,

$$NH(A_1|\mathcal{Q}) \geq L \sum_{i=0}^{M-1} \left(\frac{K}{N}\right)^i \tag{2.70}$$

and plugging it to the inductive relation concludes the converse proof.

2.7 Conclusions

In this chapter, we considered the private information retrieval (PIR) problem over MDS-coded and non-colluding databases. We employed information-theoretic arguments to derive the optimal retrieval rate for the desired message for any given (N, K) storage code. We showed that the PIR capacity in this case is given by $C = \frac{1-R_c}{1-R_c^M}$. The optimal retrieval rate is strictly higher than the best-known achievable scheme in the literature for any finite number of messages. This result reduces to the capacity of the classical PIR problem, i.e., with repetition-coded databases, by observing that for repetition coding $R_c = \frac{1}{N}$. Our result shows that the optimal

retrieval cost is independent of the explicit structure of the storage code, and the number of databases, but depends only on the code rate R_c and the number of messages M. Interestingly, the result implies that there is no gain of joint design of the MDS storage code and the retrieval procedure. The result also establishes a fundamental tradeoff between the code rate and the PIR capacity for the MDS codes.

CHAPTER 3

Multi-Message Private Information Retrieval

3.1 Introduction

In this chapter, we consider the problem of multi-message private information retrieval (MPIR) from N non-communicating replicated databases. In MPIR, the user is interested in retrieving P messages out of M stored messages without leaking the identity of the retrieved messages. The information-theoretic sum capacity of MPIR C_s^P is the maximum number of desired message symbols that can be retrieved privately per downloaded symbol, where the symbols are defined over the same field. For the case $P \geq \frac{M}{2}$, we determine the exact sum capacity of MPIR as $C_s^P = \frac{1}{1+\frac{M-P}{PN}}$. The achievable scheme in this case is based on downloading MDS-coded mixtures of all messages. For $P \leq \frac{M}{2}$, we develop lower and upper bounds for all M, P, N. These bounds match if the total number of messages M is an integer multiple of the number of desired messages P, i.e., $\frac{M}{P} \in \mathbb{N}$. In this case, $C_s^P = \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{M/P-1}}\right)^{-1}$, i.e., $C_s^P = \frac{1-\frac{1}{N}}{1-(\frac{1}{N})^{M/P}}$ for $N > 1$, and $C_s^P = \frac{P}{M}$ for $N = 1$. The achievable scheme in this case generalizes the single-message capacity achieving scheme to have unbalanced number of stages per round of download. For all the remaining cases, the

difference between the lower and upper bound is at most 0.0082, which occurs for $M = 5$, $P = 2$, $N = 2$. Our results indicate that joint retrieval of desired messages is more efficient than successive use of single-message retrieval schemes even after considering the free savings that result from downloading undesired symbols in each single-message retrieval round.

3.2 Problem Formulation

Consider a classical PIR setting storing M messages (or files). Each message is a vector $W_i \in \mathbb{F}_q^L$, $i \in \{1, \cdots, M\}$, whose elements are picked uniformly and independently from sufficiently large field[1] $\mathbb{F}_q$. Denote the contents of message W_m by the vector $[w_m(1), w_m(2), \cdots, w_m(L)]^T$. The messages are independent and identically distributed, and thus,

$$H(W_i) = L, \quad i \in \{1, \cdots, M\} \tag{3.1}$$

$$H\left(W_{1:M}\right) = ML \tag{3.2}$$

where L is measured in q-ary bits, $W_{1:M} = (W_1, W_2, \cdots, W_M)$. The messages are stored in N non-colluding (non-communicating) databases. Each database stores an identical copy of all M messages, i.e., the databases encode the messages via $(N, 1)$ repetition storage code [117].

In the MPIR problem (see Fig. 3.1), the user aims to retrieve a subset of mes-

[1]We note that using $q = \min\{p^m \geq M : p \text{ is a prime}, m \in \mathbb{N}\}$ is sufficient to ensure the existence of the $P \times M$ MDS generator matrix in Section 4. Furthermore, binary field suffices for the achievable scheme in Section 5.

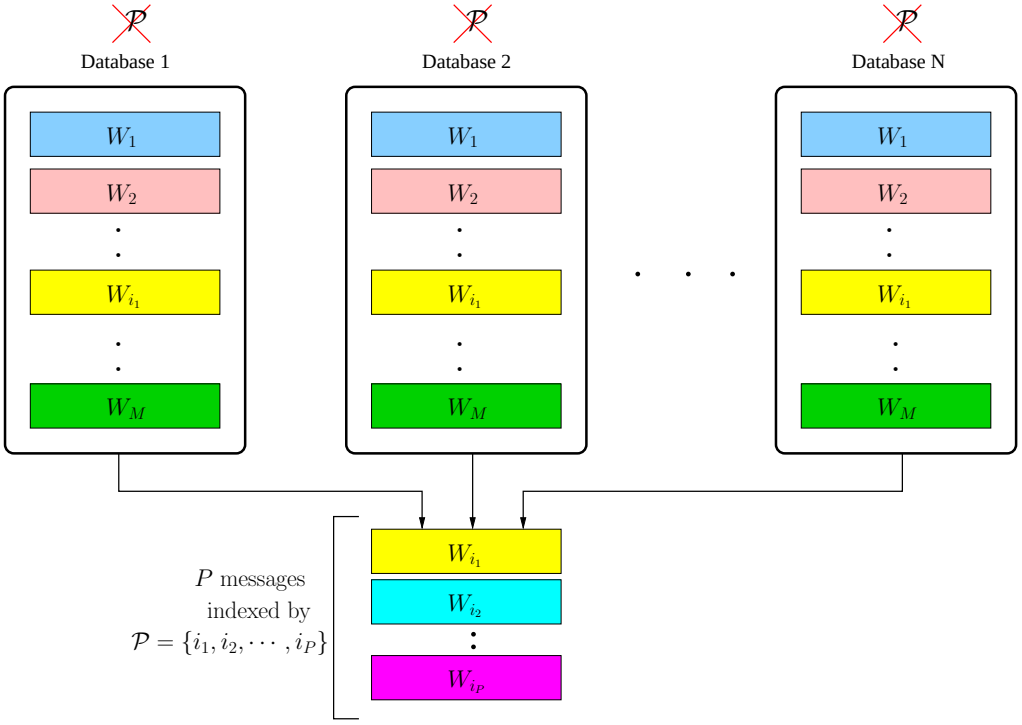


Figure 3.1: The multi-message PIR problem (MPIR).

sages indexed by the index set $\mathcal{P} = \{i_1, \cdots, i_P\} \subseteq \{1, \cdots, M\}$ out of the available messages, where $|\mathcal{P}| = P$, without leaking the identity of the subset $\mathcal{P}$. We assume that the cardinality of the potential message set, P, is known to all databases. To retrieve $W_{\mathcal{P}} = (W_{i_1}, W_{i_2}, \cdots, W_{i_P})$, the user generates a query $Q_n^{[\mathcal{P}]}$ and sends it to the nth database. The user does not have any knowledge about the messages in advance, hence the messages and the queries are statistically independent,

$$I\left(W_1,\cdots,W_M;Q_1^{[\mathcal{P}]},\cdots,Q_N^{[\mathcal{P}]}\right) = I\left(W_{1:M};Q_{1:N}^{[\mathcal{P}]}\right)=0 \tag{3.3}$$

The privacy is satisfied by ensuring statistical independence between the queries

and the message index set $\mathcal{P} = \{i_1, \cdots, i_P\}$, i.e., the privacy constraint is given by,

$$I\left(Q_n^{[i_1,\cdots,i_P]}; i_1, \cdots, i_P\right) = I\left(Q_n^{[\mathcal{P}]}; \mathcal{P}\right) = 0 \tag{3.4}$$

for all $n \in \{1, \cdots, N\}$.

The nth database responds with an answer string $A_n^{[\mathcal{P}]}$, which is a deterministic function of the queries and the messages, hence

$$H(A_n^{[\mathcal{P}]}|Q_n^{[\mathcal{P}]}, W_{1:M}) = 0 \tag{3.5}$$

We further note that by the data processing inequality and (3.4),

$$I\left(A_n^{[\mathcal{P}]}; \mathcal{P}\right) = 0, \quad n \in \{1, \cdots, N\} \tag{3.6}$$

In addition, the user should be able to reconstruct the messages $W_{\mathcal{P}}$ reliably from the collected answers from all databases given the knowledge of the queries. Thus, we write the reliability constraint as,

$$H(W_{i_1}, \cdots, W_{i_P}|A_1^{[\mathcal{P}]}, \cdots, A_N^{[\mathcal{P}]}, Q_1^{[\mathcal{P}]}, \cdots, Q_N^{[\mathcal{P}]}) = H\left(W_{\mathcal{P}}|A_{1:N}^{[\mathcal{P}]}, Q_{1:N}^{[\mathcal{P}]}\right) = 0 \tag{3.7}$$

We denote the retrieval rate of the ith message by R_i, where $i \in \mathcal{P}$. The retrieval rate of the ith message is the ratio between the length of message i and

the total download cost of the message set $\mathcal{P}$ that includes W_i. Hence,

$$R_i = \frac{H(W_i)}{\sum_{n=1}^{N} H\left(A_n^{[\mathcal{P}]}\right)} \tag{3.8}$$

The sum retrieval rate of $W_{\mathcal{P}}$ is given by,

$$\sum_{i=1}^{P} R_i = \frac{H(W_{\mathcal{P}})}{\sum_{n=1}^{N} H\left(A_n^{[\mathcal{P}]}\right)} = \frac{PL}{\sum_{n=1}^{N} H\left(A_n^{[\mathcal{P}]}\right)} \tag{3.9}$$

The sum capacity of the MPIR problem is given by

$$C_s^P = \sup \sum_{i=1}^{P} R_i \tag{3.10}$$

where the sup is over all private retrieval schemes.

In this chapter, we follow the information-theoretic assumptions of large enough message size, large enough field size, and ignore the upload cost as in [8, 12, 14, 117]. A formal treatment of the capacity under message and field size constraints for $P = 1$ can be found in [18]. We note that the MPIR problem described here reduces to the classical PIR problem when $P = 1$, whose capacity is characterized in [12].

3.3 Main Results and Discussions

Our first result is the exact characterization of the sum capacity for the case $P \geq \frac{M}{2}$, i.e., when the user wishes to privately retrieve at least half of the messages stored

in the databases.

Theorem 3.1 *For the MPIR problem with non-colluding and replicated databases, if the number of desired messages P is at least half of the number of overall stored messages M, i.e., if $P \geq \frac{M}{2}$, then the sum capacity is given by,*

$$C_s^P = \frac{1}{1 + \frac{M-P}{PN}} \tag{3.11}$$

The achievability proof for Theorem 3.1 is given in Section 3.4, and the converse proof is given in Section 3.6.1. We note that when $P = 1$, the constraint of Theorem 3.1 is equivalent to $M = 2$, and the result in (3.11) reduces to the known result of [12] for $P = 1$, $M = 2$, which is $\frac{1}{1+\frac{1}{N}}$. We observe that the sum capacity in (3.11) is a strictly increasing function of N, and $C_s^P \to 1$ as $N \to \infty$. We also observe that the sum capacity in this regime is a strictly increasing function of P, and approaches 1 as[2] $P \to M$.

The following corollary compares our result and the rate corresponding to the repeated use of single-message retrieval scheme [12].

Corollary 3.1 *For the MPIR problem with $P \geq \frac{M}{2}$, the repetition of the single-message retrieval scheme of [12] P times in a row, which achieves a sum rate of,*

$$R_s^{rep} = \frac{(N-1)(N^{M-1}+P-1)}{N^M - 1} \tag{3.12}$$

[2]Note that in the degenerate case, when $P = M$, the privacy constraint is trivially satisfied as $H(\mathcal{P}) = H(\mathcal{P}|Q_n^{[\mathcal{P}]}) = 0$ as there is no uncertainty about the identity of the desired messages if $P = M$ either with or without the knowledge of the queries. Thus, the optimal sum retrieval rate is 1 which is achieved by downloading all the messages.

is strictly sub-optimal with respect to the exact capacity in (3.11).

Corollary 3.1 implies that applying Sun-Jafar scheme [12] P times is sub-optimal, even if the user uses the undesired symbols, which are downloaded as a byproduct of Sun-Jafar scheme, as a *head start* in downloading the remaining messages because in this case the user would achieve $R_s^{rep} < C_s^P$.

Proof: In order to use the single-message capacity achieving PIR scheme as an MPIR scheme, the user repeats the single-message achievable scheme for each individual message that belongs to $\mathcal{P}$. We note that at each repetition, the scheme downloads extra decodable symbols from other messages. By this argument, the following rate R_s^{rep} is achievable using a repetition of the single-message scheme,

$$R_s^{rep} = C + \Delta(M, P, N) \tag{3.13}$$

where C is the single-message capacity which is given by $C = \frac{1-\frac{1}{N}}{1-(\frac{1}{N})^M}$ [12], and $\Delta(M, P, N)$ is the rate of the extra decodable symbols that belong to $\mathcal{P}$. To calculate $\Delta(M, P, N)$, we note that the total download cost D is given by $D = \frac{L}{C}$ by definition. Since $L = N^M$ in the single-message scheme, $D = \frac{N^M(1-(\frac{1}{N})^M)}{1-\frac{1}{N}} = \frac{N^{M+1}-N}{N-1}$. The single-message scheme downloads one symbol from every message from every database, i.e., the scheme downloads extra $(P-1)N$ symbols from the remaining desired messages that belong to $\mathcal{P}$, thus,

$$\Delta(M, P, N) = \frac{(P-1)N(N-1)}{N^{M+1}-N} = \frac{(P-1)(N-1)}{N^M-1} \tag{3.14}$$

Using this in (3.13) gives the R_s^{rep} expression in (3.12).

Now, the difference between the capacity in (3.11) and achievable rate in (3.12) is,

$$C_s^P - R_s^{rep} = \frac{PN}{P(N-1)+M} - \frac{(N-1)(N^{M-1}+P-1)}{N^M-1} \tag{3.15}$$

$$= \frac{\eta(P,M,N)}{(N^M-1)(P(N-1)+M)} \tag{3.16}$$

It suffices to prove that $\eta(P,M,N) \geq 0$ for all P, M, N when $P \geq \frac{M}{2}$ and $N \geq 2$. Note,

$$\begin{aligned}\eta(P,M,N) =&(2P-M)N^M + (M-P)N^{M-1} - P(P-1)N^2 \\ &+ ((P-1)(2P-M)-P)N + (M-P)(P-1)\end{aligned} \tag{3.17}$$

In the regime $P \geq \frac{M}{2}$, coefficients of N^M, N^{M-1}, N^0 are non-negative. Denote the negative terms in $\eta(\cdot)$ by $\nu(P,N)$ which is $\nu(P,N) = P(P-1)N^2 + PN$. We note $\nu(P,N) < P^2N^2$ when $N > 1$, which is the case here. Thus,

$$\begin{aligned}\eta(P,M,N) \geq&(2P-M)N^M + (M-P)N^{M-1} + (P-1)(2P-M)N \\ &+ (M-P)(P-1) - P^2N^2 \qquad (3.18)\\ >&(2P-M)N^M + (M-P)N^{M-1} - P^2N^2 \qquad (3.19)\\ =&N^2\left((2P-M)N^{M-2} + (M-P)N^{M-3} - P^2\right) \qquad (3.20)\\ \geq&N^2\left((2P-M)2^{M-2} + (M-P)2^{M-3} - P^2\right) \qquad (3.21)\end{aligned}$$

$$=N^2\left(2^{M-3}(3P-M)-P^2\right) \tag{3.22}$$

$$\geq N^2\left(2^{M-3}\cdot\frac{M}{2}-M^2\right) \tag{3.23}$$

$$=MN^2\left(2^{M-4}-M\right) \tag{3.24}$$

where (3.21) follows from the fact that $(2P-M)N^{M-2}+(M-P)N^{M-3}-P^2$ is monotone increasing in $N\geq 2$ for $M\geq 3$, and (3.23) follows from $\frac{M}{2}\leq P\leq M$. From (3.24), we conclude that $\eta(M,P,N)>0$ for all $M\geq 7$, $P\geq\frac{M}{2}$ and $N\geq 2$. Examining the expression in (3.17) for the remaining cases manually, i.e., when $M\leq 6$, we note that $\eta(M,P,N)>0$ in these cases as well. Therefore, $\eta(M,P,N)>0$ for all possible cases, and the MPIR capacity is strictly larger than the rate achieved by repeating the optimum single-message PIR scheme. ■

For the example in the introduction, where $M=3$, $P=2$, $N=2$, our MPIR scheme achieves a sum capacity of $\frac{4}{5}$ in (3.11), which is strictly larger than the repeating-based achievable sum rate of $\frac{5}{7}$ in (3.12).

The following corollary gives an achievable rate region for the MPIR problem.

Corollary 3.2 *For the MPIR problem, for the case $P\geq\frac{M}{2}$, the following rate region is achievable,*

$$\begin{aligned}\mathcal{C}=\text{conv}\{&(C,\delta,\cdots,\delta),(\delta,C,\cdots,\delta),\cdots,(\delta,\cdots,\delta,C),\\&(C,0,0,\cdots,0),(0,C,0,\cdots,0),\cdots,(0,0,\cdots,C),\\&(0,0,\cdots,0),\left(C^P,C^P,\cdots,C^P\right)\}\end{aligned} \tag{3.25}$$

where

$$
\begin{aligned}
C &= \frac{1-\frac{1}{N}}{1-(\frac{1}{N})^M}, \\
C^P &= \frac{C_s^P}{P} = \frac{N}{PN+(M-P)}, \\
\delta &= \frac{\Delta(M,P,N)}{P-1} = \frac{N-1}{N^M-1}
\end{aligned}
\tag{3.26}
$$

and where conv *denotes the convex hull, and all corner points lie in the P-dimensional space.*

Proof: This is a direct consequence of Theorem 3.1 and Corollary 3.1. The corner point $\left(C, \frac{\Delta(M,P,N)}{P-1}, \frac{\Delta(M,P,N)}{P-1}, \cdots, \frac{\Delta(M,P,N)}{P-1}\right) = \left(\frac{1-\frac{1}{N}}{1-(\frac{1}{N})^M}, \frac{N-1}{N^M-1}, \frac{N-1}{N^M-1}, \cdots, \frac{N-1}{N^M-1}\right)$ is achievable from the single-message achievable scheme. Due to the symmetry of the problem any other permutation for the coordinates of this corner point is also achievable by changing the roles of the desired messages. Theorem 3.1 gives the symmetric sum capacity corner point for the case of $P \geq \frac{M}{2}$, namely $\left(\frac{C_s^P}{P}, \frac{C_s^P}{P}, \cdots, \frac{C_s^P}{P}\right) = \left(\frac{N}{PN+(M-P)}, \frac{N}{PN+(M-P)}, \cdots, \frac{N}{PN+(M-P)}\right)$. By time sharing of these corner points along with the origin, the region in (3.25) is achievable. ■

As an example for this achievable region, consider again the example in the introduction, where $M = 3$, $P = 2$, $N = 2$. In this case, we have a two-dimensional rate region with three corner points: $(\frac{4}{7}, \frac{1}{7})$, which corresponds to the single-message capacity achieving point that aims at retrieving W_1; $(\frac{1}{7}, \frac{4}{7})$, which corresponds to single-message capacity achieving point that aims at retrieving W_2; and $(\frac{2}{5}, \frac{2}{5})$, which corresponds to the symmetric sum capacity point. The convex hull of these corner

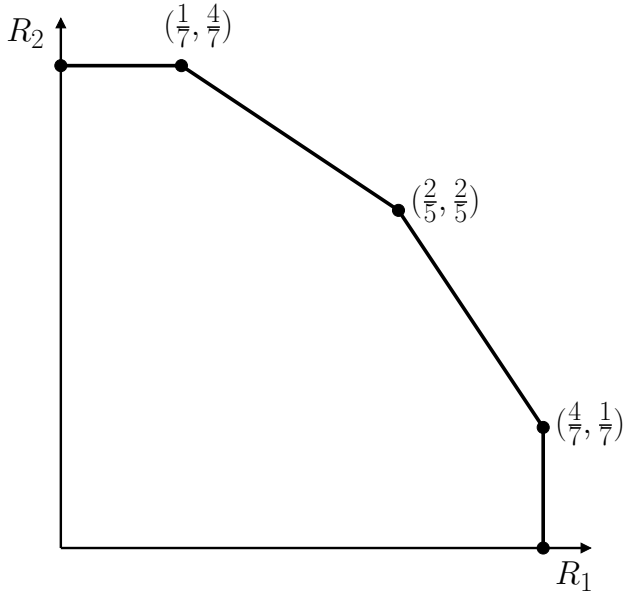


Figure 3.2: The achievable rate region of $M = 3$, $P = 2$, $N = 2$.

points together with the points on the axes gives the achievable region in Fig. 3.2. We note that in general, the rate region in Corollary 3.2 is merely *an achievable region*. The capacity region that characterizes the exact tradeoff between the retrieval rates for the P messages remains an open problem despite the optimality of the corner points. A converse argument is needed to show the optimality of time-sharing (if the rate region is indeed the capacity region).

For the case $P \leq \frac{M}{2}$, we have the following result, where the lower and upper bound match if $\frac{M}{P} \in \mathbb{N}$.

Theorem 3.2 *For the MPIR problem with non-colluding and replicated databases, when $P \leq \frac{M}{2}$, the sum capacity is lower and upper bounded as,*

$$\underline{R}_s \leq C_s^P \leq \bar{R}_s \tag{3.27}$$

where the upper bound $\bar{R}_s$ *is given by,*

$$\bar{R}_s = \frac{1}{1+\frac{1}{N}+\cdots+\frac{1}{N^{\lfloor \frac{M}{P} \rfloor -1}}+\left(\frac{M}{P}-\lfloor \frac{M}{P} \rfloor\right)\frac{1}{N^{\lfloor \frac{M}{P} \rfloor}}} \tag{3.28}$$

$$= \frac{1}{\frac{1-(\frac{1}{N})^{\lfloor \frac{M}{P} \rfloor}}{1-\frac{1}{N}}+\left(\frac{M}{P}-\lfloor \frac{M}{P} \rfloor\right)\frac{1}{N^{\lfloor \frac{M}{P} \rfloor}}} \tag{3.29}$$

For the lower bound, define r_i *as,*

$$r_i = \frac{e^{j2\pi(i-1)/P}}{N^{1/P}-e^{j2\pi(i-1)/P}}, \quad i = 1, \cdots, P \tag{3.30}$$

where $j = \sqrt{-1}$*, and denote* γ_i*,* $i = 1, \cdots, P$*, to be the solutions of the linear equations* $\sum_{i=1}^{P} \gamma_i r_i^{-P} = (N-1)^{M-P}$*, and* $\sum_{i=1}^{P} \gamma_i r_i^{-k} = 0$*,* $k = 1, \cdots, P-1$*, then* $\underline{R}_s$ *is given by,*

$$\underline{R}_s = \frac{\sum_{i=1}^{P} \gamma_i r_i^{M-P}\left[\left(1+\frac{1}{r_i}\right)^{M}-\left(1+\frac{1}{r_i}\right)^{M-P}\right]}{\sum_{i=1}^{P} \gamma_i r_i^{M-P}\left[\left(1+\frac{1}{r_i}\right)^{M}-1\right]} \tag{3.31}$$

The achievability lower bound in Theorem 3.2 is shown in Section 3.5 and the upper bound is derived in Section 3.6.2. The following corollary states that the bounds in Theorem 3.2 match if the total number of messages is an integer multiple of the number of desired messages.

Corollary 3.3 *For the MPIR problem with non-colluding and replicated databases,*

if $\frac{M}{P}$ *is an integer, then the bounds in (3.27) match, and hence, for*[3] $N > 1$

$$C_s^P = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^{\frac{M}{P}}}, \quad \frac{M}{P} \in \mathbb{N}, \tag{3.32}$$

Proof: For the upper bound, observe that if $\frac{M}{P} \in \mathbb{N}$, then $\frac{M}{P} = \left\lfloor \frac{M}{P} \right\rfloor$. Hence, (3.28) becomes

$$\bar{R}_s = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^{\frac{M}{P}}} \tag{3.33}$$

For the lower bound, consider the case $\frac{M}{P} \in \mathbb{N}$. From (3.30),

$$\left(1 + \frac{1}{r_i}\right)^M = \left(\frac{N^{1/P}}{e^{j2\pi(i-1)/P}}\right)^M = N^{\frac{M}{P}} \tag{3.34}$$

since $e^{j2\pi(i-1)M/P} = 1$ for $\frac{M}{P} \in \mathbb{N}$. Similarly, $\left(1 + \frac{1}{r_i}\right)^{M-P} = N^{\frac{M}{P}-1}$. Hence, if $\frac{M}{P} \in \mathbb{N}$,

$$\underline{R}_s = \frac{\sum_{i=1}^P \gamma_i r_i^{M-P} \left[N^{\frac{M}{P}} - N^{\frac{M}{P}-1}\right]}{\sum_{i=1}^P \gamma_i r_i^{M-P} \left[N^{\frac{M}{P}} - 1\right]} \tag{3.35}$$

$$= \frac{N^{\frac{M}{P}} - N^{\frac{M}{P}-1}}{N^{\frac{M}{P}} - 1} \tag{3.36}$$

$$= \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^{\frac{M}{P}}} \tag{3.37}$$

Thus, $\underline{R}_s = C_s^P = \bar{R}_s$ if $\frac{M}{P} \in \mathbb{N}$, and we have an exact capacity result in this case.

■

[3] If $N = 1$, the optimal retrieval scheme is to download the contents of the database, hence $C_s^P = \frac{P}{M}$.

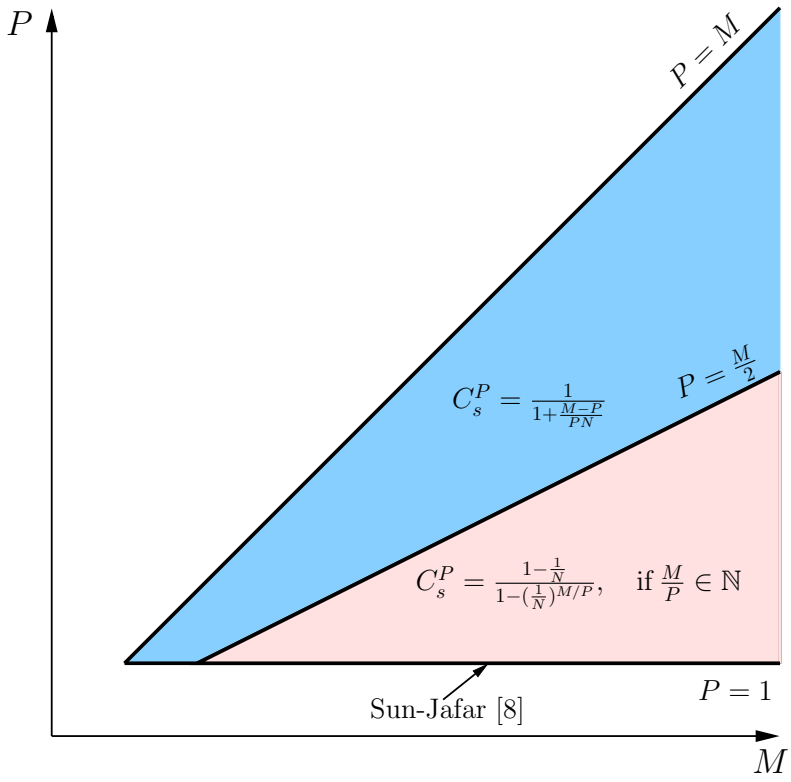
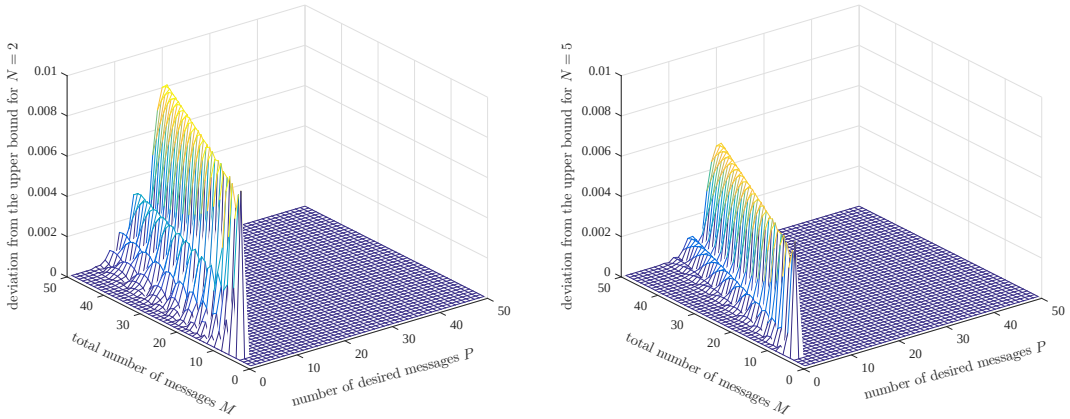


Figure 3.3: Summary of the state of the results.

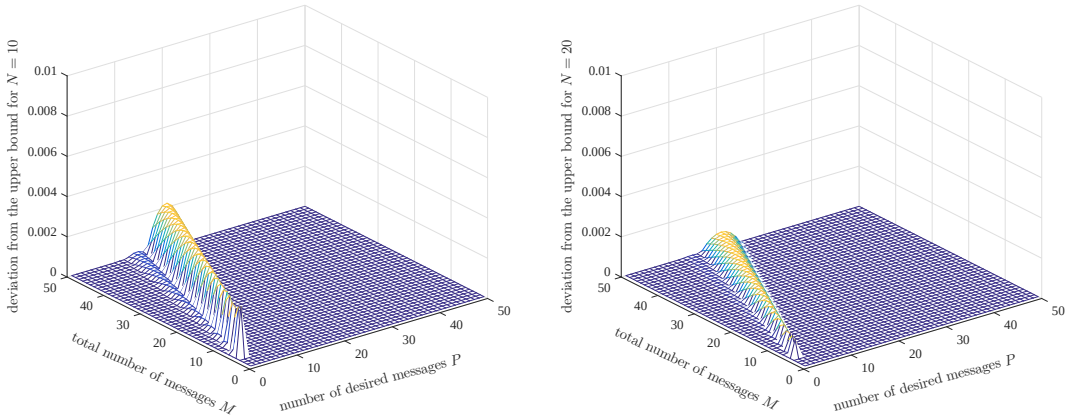
Examining the result, we observe that when the total number of messages is an integer multiple of the number of desired messages, the sum capacity of the MPIR is the same as the capacity of the single-message PIR with the number of messages equal to $\frac{M}{P}$. Note that, although at first the result may seem as if every P messages can be lumped together as a single message, and the achievable scheme in [12] can be used, this is not the case. The reason for this is that, we need to ensure the privacy constraint for *every subset* of messages of size[4] P. That is why, in this chapter, we develop a new achievable scheme.

[4]We note that this is similar to the TPIR problem when $\frac{N}{T} \in \mathbb{N}$, in which case one cannot simply lump every T databases together and apply the capacity-achieving scheme of PIR with non-colluding databases for the new system that consists of $\frac{N}{T}$ databases. In both problems, the use of MDS codes is important to induce symmetry across the group of messages/databases.



(a) $N = 2$

(b) $N = 5$



(c) $N = 10$

(d) $N = 20$

Figure 3.4: Deviation of the achievable sum rate from the upper bound.

The state of the results is summarized in Fig. 3.3: Consider the (M, P) plane, where naturally $M \geq P$. The valid part of the plane is divided into two regions. The first region is confined between the lines $P = \frac{M}{2}$ and $P = M$; the sum capacity in this region is exactly characterized (Theorem 3.1). The second region is confined between the lines $P = 1$ and $P = \frac{M}{2}$; the sum capacity in this region is characterized only for the cases when $\frac{M}{P} \in \mathbb{N}$ (Corollary 3.3). The line $P = 1$ corresponds to the previously known result for the single-message PIR [12]. The exact capacity for the rest of the cases is still an open problem; however, the achievable scheme in

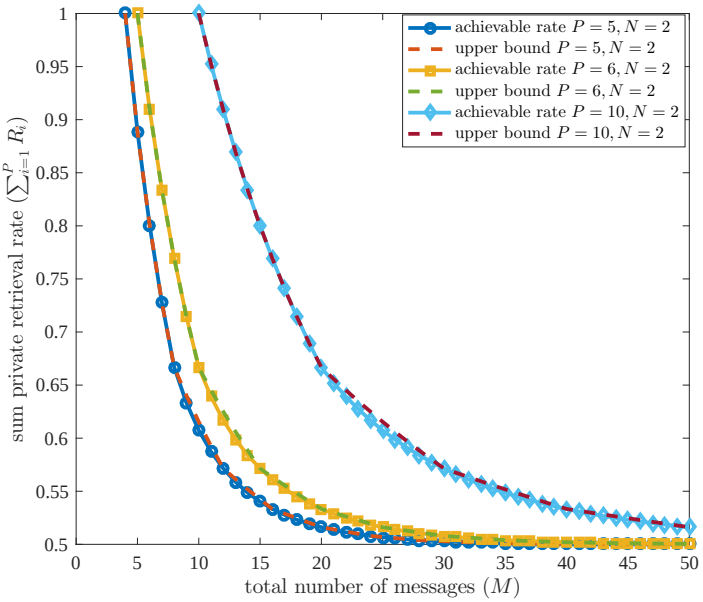


Figure 3.5: Effect of changing M for fixed $P = 5, 6, 10$ and fixed $N = 2$.

Theorem 3.2 yields near-optimal sum rates for all the remaining cases with the largest difference of 0.0082 from the upper bound, as discussed next.

Fig. 3.4 shows the difference of the achievable rate $\underline{R}_s$ and the upper bound $\bar{R}_s$ in Theorem 3.2, i.e., $\bar{R}_s - \underline{R}_s$. The figure shows that the difference decreases as N increases. This difference in all cases is small and is upper bounded by 0.0082, which occurs when $M = 5$, $P = 2$, $N = 2$. In addition, the difference is zero for the cases $P \geq \frac{M}{2}$ (Theorem 3.1) or $\frac{M}{P} \in \mathbb{N}$ (Corollary 3.3).

Fig. 3.5 shows the effect of changing M for fixed (P, N). We observe that as M increases, the sum rate monotonically decreases and has a limit of[5] $1 - \frac{1}{N}$. In addition, Fig. 3.6 shows the effect of changing N for fixed (P, M). We observe that as N increases, the sum rate increases and approaches 1, as expected.

[5]Although it seems at first that $C_s^P \to 1 - \frac{1}{N}$ if $M \to \infty$, we emphasize that this is true if only $P = o(M)$, i.e., P does not scale with M. If $P = \gamma M$, then as $M \to \infty$, we have $C_s^P = \frac{1}{1+\frac{1-\gamma}{\gamma N}} > 1 - \frac{1}{N}$, when $\gamma > \frac{1}{2}$, and $C_s^P = \frac{1-\frac{1}{N}}{1-(\frac{1}{N})^{1/\gamma}} > 1 - \frac{1}{N}$, when $\frac{1}{\gamma} \geq 2 \in \mathbb{N}$.

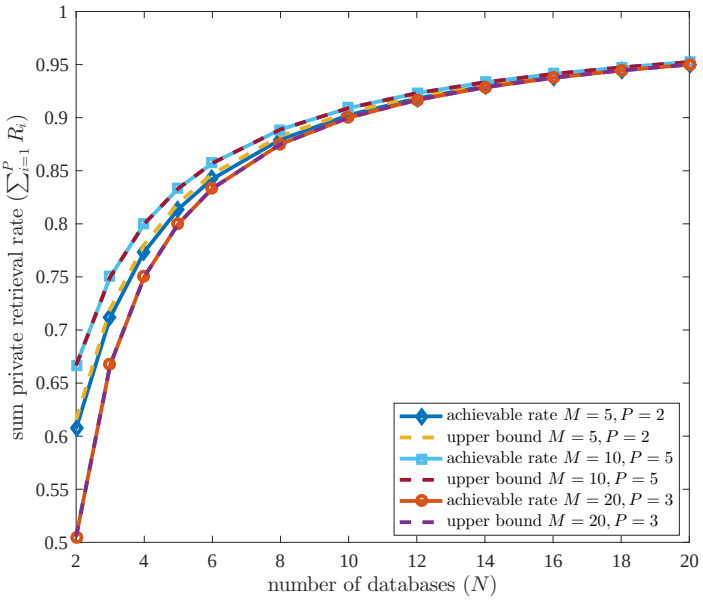


Figure 3.6: Effect of changing N for fixed $(M, P) = (5, 2), (10, 5), (20, 3)$.

3.4 Achievability Proof for the Case $P \geq \frac{M}{2}$

In this section, we present the general achievable scheme that attains the upper bound for the case $P \geq \frac{M}{2}$. The scheme applies the concepts of message symmetry, database symmetry, and exploiting side information as in [12]. However, our scheme requires the extra ingredient of MDS coding of the desired symbols and the side information in its second stage. We note also that, here, by message symmetry, we mean symmetry across group of messages of size P, which is realized by MDS coding.

3.4.1 Motivating Example: $M = 3$, $P = 2$ Messages, $N = 2$ Databases

We start with a simple motivating example in this sub-section. The scheme operates over message size $N^2 = 4$. For sake of clarity, we assume that the three messages after interleaving their indices are $W_1 = (a_1, \cdots, a_4)^T$, $W_2 = (b_1, \cdots, b_4)^T$, and $W_3 = (c_1, \cdots, c_4)^T$. We use $\mathbf{G}_{2\times 3}$ Reed-Solomon generator matrix over $\mathbb{F}_3$ as

$$\mathbf{G}_{2\times 3} = \begin{bmatrix} 1 & 1 & 1 \\ \\ 1 & 2 & 0 \end{bmatrix} \tag{3.38}$$

The user picks a random permutation for the columns of $\mathbf{G}_{2\times 3}$ from the 6 possible permutations, e.g., in this example we use the permutation $2, 1, 3$. In the first round, the user starts by downloading one symbol from each database and each message, i.e., the user downloads (a_1, b_1, c_1) from the first database, and (a_2, b_2, c_2) from the second database. In the second round, the user encodes the side information from database 2 which is c_2 with two new symbols from W_1, W_2 which are (a_3, b_3) using the permuted generator matrix, i.e., the user downloads two equations from database 1 in the second round,

$$\mathbf{GS}_1 \begin{bmatrix} a_3 \\ \\ b_3 \\ \\ c_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ \\ 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ \\ 1 & 0 & 0 \\ \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_3 \\ \\ b_3 \\ \\ c_2 \end{bmatrix} = \begin{bmatrix} a_3 + b_3 + c_2 \\ \\ 2a_3 + b_3 \end{bmatrix} \tag{3.39}$$

The user repeats this operation for the second database with (a_4, b_4) as desired symbols and c_1 as the side information from the first database.

For the decodability: The user subtracts out c_2 from round two in the first database, then the user can decode (a_3, b_3) from $a_3 + b_3$ and $2a_3 + b_3$. Similarly, by subtracting out c_1 from round two in the second database, the user can decode (a_4, b_4) from $a_4 + b_4$ and $2a_4 + b_4$.

For the privacy: Single bit retrievals of (a_1, b_1, c_1) and (a_2, b_2, c_2) from the two databases in the first round satisfy message symmetry and database symmetry, and do not leak any information. In addition, due to the private shuffling of bit indices, the different coefficients of 1, 2 and 0 in front of the bits in the MDS-coded summations in the second round do not leak any information either; see a formal proof in Section 3.4.3. To see the privacy constraint intuitively from another angle, we note that the user can alter the queries for the second database when the queries for the first database are fixed, when the user wishes to retrieve another set of two messages. For instance, if the user wishes to retrieve (W_1, W_3) instead of (W_1, W_2), it can alter the queries for the second database by changing every c_2 in the queries of the second database with c_3, c_1 with c_4, b_2 with b_3, and b_4 with b_1.

The query table for this case is shown in Table 3.1 below. The scheme retrieves $a_1, \cdots, a_4$ and $b_1, \cdots, b_4$, i.e., 8 bits in 10 downloads (5 from each database). Thus, the achievable sum rate for this scheme is $\frac{8}{10} = \frac{4}{5} = \frac{1}{1+\frac{M-P}{PN}}$. If we use the single-message optimal scheme in [12], which is given in [12, Example 4.3] for this specific case, twice in a row to retrieve two messages, we achieve a sum rate of $\frac{20}{28} = \frac{5}{7} < \frac{4}{5}$ as discussed in the introduction.

Table 3.1: The query table for the case $M = 3, P = 2, N = 2$.

Database 1	Database 2
a_1, b_1, c_1	a_2, b_2, c_2
$a_3 + b_3 + c_2$	$a_4 + b_4 + c_1$
$2a_3 + b_3$	$2a_4 + b_4$

3.4.2 General Achievable Scheme

The scheme requires $L = N^2$, and is completed in two rounds. The main ingredient of the scheme is the MDS coding of the desired symbols and side information in the second round. The details of the scheme are as follows.

1. *Index preparation:* The user interleaves the contents of each message randomly and independently from the remaining messages using a random interleaver $\pi_m(.)$ which is known privately to the user only, i.e.,

$$x_m(i) = w_m(\pi_m(i)), \quad i \in \{1, \cdots, L\} \tag{3.40}$$

   where $X_m = [x_m(1), \cdots, x_m(L)]^T$ is the interleaved message. Thus, the downloaded symbol $x_m(i)$ at any database appears to be chosen at random and independent from the desired message subset $\mathcal{P}$.

2. *Round one:* As in [12], the user downloads one symbol from every message from every database, i.e., the user downloads $(x_1(n), x_2(n), \cdots, x_M(n))$ from the nth database. This implements *message symmetry*, *symmetry across databases*, and satisfies the privacy constraint.

3. *Round two:* The user downloads a coded mixture of new symbols from

the desired messages and the undesired symbols downloaded from the other databases. Specifically,

(a) The user picks an MDS generator matrix $\mathbf{G} \in \mathbb{F}_q^{P\times M}$, which has the property that every $P \times P$ submatrix is full-rank. This implies that if the user can cancel out any $M - P$ symbols from the mixture, the remaining symbols can be decoded. One explicit MDS generator matrix is the Reed-Solomon generator matrix over $\mathbb{F}_q$, where $q > M$, [118, 119]. The matrix is constructed by choosing M distinct elements of $\mathbb{F}_q$. Let us denote these elements by $\{\theta_1, \theta_2, \cdots, \theta_M\}$. Then,

$$\mathbf{G} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ \theta_1 & \theta_2 & \theta_3 & \cdots & \theta_M \\ \theta_1^2 & \theta_2^2 & 3^2 & \cdots & \theta_M^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \theta_1^{P-1} & \theta_2^{P-1} & \theta_3^{P-1} & \cdots & \theta_M^{P-1} \end{bmatrix}_{P\times M} \tag{3.41}$$

(b) The user picks uniformly and independently at random the permutation matrices $\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_{N-1}$ of size $M \times M$. These matrices shuffle the order of the columns of $\mathbf{G}$ to be independent of $\mathcal{P}$.

(c) At the first database, the user downloads an MDS-coded version of P new symbols from the desired set $\mathcal{P}$ and $M - P$ undesired symbols that are already decoded from the second database in the first round, i.e., the

user downloads P equations of the form

$$\mathbf{GS}_1[x_{i_1}(N+1) \quad x_{i_2}(N+1) \cdots x_{i_P}(N+1) \quad x_{j_1}(2) \quad x_{j_2}(2) \cdots x_{j_{M-P}}(2)]^T \tag{3.42}$$

where $\mathcal{P} = \{i_1, i_2, \cdots, i_P\}$ are the indices of the desired messages and $\bar{\mathcal{P}} = \{j_1, j_2, \cdots, j_{M-P}\}$ are the indices of the undesired messages. In this case, the user can cancel out the undesired messages and be left with a $P \times P$ invertible system of equations that it can solve to get $[x_{i_1}(N+1), x_{i_2}(N+1), \cdots, x_{i_P}(N+1)]$. This implements *exploiting side information* as in [12].

(d) The user repeats the last step for each set of side information from database 3 to database N, each with different permutation matrix.

(e) By *database symmetry*, the user repeats all steps of round two at all other databases.

3.4.3 Decodability, Privacy, and Calculation of the Achievable Rate

Now, we verify that this achievable scheme satisfies the reliability and privacy constraints.

For the reliability: The user gets individual symbols from all databases in the first round, and hence they are all decodable by definition. In the second round, the user can subtract out all the undesired message symbols using the undesired symbols downloaded from all other databases during the first round. Consequently,

the user is left with a $P \times P$ system of equations which is guaranteed to be invertible by the MDS property, hence all symbols that belong to $W_{\mathcal{P}}$ are decodable.

For the privacy: At each database, for every message subset $\mathcal{P}$ of size P, the achievable scheme retrieves randomly interleaved symbols which are encoded by the following matrix:

$$\mathbf{H}_{\mathcal{P}} = \begin{bmatrix} \mathbf{I}_P & \mathbf{0}_P & \mathbf{0}_P & \cdots & \mathbf{0}_P \\ \mathbf{0}_P & \mathbf{G}_{\mathcal{P}}^1 & \mathbf{0}_P & \cdots & \mathbf{0}_P \\ \mathbf{0}_P & \mathbf{0}_P & \mathbf{G}_{\mathcal{P}}^2 & \cdots & \mathbf{0}_P \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_P & \mathbf{0}_P & \mathbf{0}_P & \cdots & \mathbf{G}_{\mathcal{P}}^{N-1} \end{bmatrix} \tag{3.43}$$

where $\mathbf{G}_{\mathcal{P}}^n = \mathbf{G}\mathbf{S}_n(:, \mathcal{P})$ are the columns of the encoding matrix that correspond to the message subset $\mathcal{P}$ after applying the random permutation $\mathbf{S}_n$. Since the permutation matrices are chosen uniformly and independently from each other, the probability distribution of $\mathbf{H}_{\mathcal{P}}$ is uniform irrespective to $\mathcal{P}$ (the probability of realizing such a matrix is $\left(\frac{(M-P)!}{M!}\right)^{N-1}$). Furthermore, the symbols are chosen randomly and uniformly by applying the random interleaver. Hence, the retrieval scheme is private.

To calculate the achievable rate: We note that at each database, the user downloads M individual symbols in the first round that includes P desired symbols. The user exploits the side information from the remaining $(N-1)$ databases to generate P equations for each side information set. Each set of P equations in turn

generates P desired symbols. Hence, the achievable rate is calculated as,

$$\sum_{i=1}^{P} R_i = \frac{\text{total number of desired symbols}}{\text{total downloaded equations}} \tag{3.44}$$

$$= \frac{N(P + P(N-1))}{N(M + P(N-1))} \tag{3.45}$$

$$= \frac{PN}{(M-P) + PN} \tag{3.46}$$

$$= \frac{1}{1 + \frac{M-P}{PN}} \tag{3.47}$$

3.4.4 Further Examples for the Case $P \geq \frac{M}{2}$

In this section, we illustrate our achievable scheme with two more basic examples. In Section 3.4.1, we considered the case $M = 3$, $P = 2$, $N = 2$. In the next two subsections, we will consider examples with larger M, P (Section 3.4.4.1), and larger N (Section 3.4.4.2).

3.4.4.1 $M = 5$ Messages, $P = 3$ Messages, $N = 2$ Databases

Let $\mathcal{P} = \{1, 2, 3\}$, and a to e denote the contents of W_1 to W_5, respectively. The achievable scheme is similar to the example in Section 3.4.1. The main difference is that in this case, we use 5×5 permutation matrix for $\mathbf{S}_1$ and $\mathbf{G}_{3\times5}$ Reed-Solomon generator matrix over $\mathbb{F}_5$ as:

$$\mathbf{G}_{3\times5} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 0 \\ 1 & 4 & 4 & 1 & 0 \end{bmatrix} \tag{3.48}$$

The query table is shown in Table 3.2 below with the following random permutation for the columns: $2, 5, 1, 3, 4$. The reliability and privacy constraints are satisfied due to the MDS property that implies that any subset of 3 messages corresponds to a 3×3 invertible submatrix if the remaining symbols are decodable from the other database. This scheme retrieves $a_1, \cdots, a_4$, $b_1, \cdots, b_4$ and $c_1, \cdots, c_4$, hence 12 bits in 16 downloads (8 from each database). Thus, the achievable sum rate is $\frac{12}{16} = \frac{3}{4}$ which equals the sum capacity $\frac{1}{1+\frac{M-P}{PN}}$ in (3.11). This strictly outperforms the repetition-based achievable sum rate $\frac{18}{31}$ in (3.12).

Table 3.2: The query table for $M = 5, P = 3, N = 2$.

Database 1	Database 2
a_1, b_1, c_1, d_1, e_1	a_2, b_2, c_2, d_2, e_2
$a_3 + b_3 + c_3 + d_2 + e_2$	$a_4 + b_4 + c_4 + d_1 + e_1$
$2a_3 + c_3 + 3d_2 + 4e_2$	$2a_4 + c_4 + 3d_1 + 4e_1$
$4a_3 + c_3 + 4d_2 + e_2$	$4a_4 + c_4 + 4d_1 + e_1$

3.4.4.2 $M = 4$ Messages, $P = 2$ Messages, $N = 3$ Databases

Next, we give an example with a larger N. Here, the message size is $N^2 = 9$. With a generator matrix $\mathbf{G}_{2\times 4} = \mathbf{G}_{3\times 5}([1:2],[1:4])$ to be the upper left submatrix of the previous example and two set of random permutations (corresponding to $\mathbf{S}_1, \mathbf{S}_2$) as $1, 3, 2, 4$, and $4, 1, 3, 2$. The query table is shown in Table 3.3 below. This scheme retrieves $a_1, \cdots, a_9$ and $b_1, \cdots, b_9$, hence 18 bits in 24 downloads (8 from each database). Thus, the achievable rate is $\frac{18}{24} = \frac{3}{4} = \frac{1}{1+\frac{M-P}{PN}}$. This strictly outperforms the repetition-based achievable scheme sum rate $\frac{7}{10}$ in (3.12).

Table 3.3: The query table for the case $M = 4, P = 2, N = 3$.

Database 1	Database 2	Database 3
a_1, b_1, c_1, d_1	a_2, b_2, c_2, d_2	a_3, b_3, c_3, d_3
$a_4 + b_4 + c_2 + d_2$	$a_6 + b_6 + c_1 + d_1$	$a_8 + b_8 + c_1 + d_1$
$a_4 + 3b_4 + 2c_2 + 4d_2$	$a_6 + 3b_6 + 2c_1 + 4d_1$	$a_8 + 3b_8 + 2c_1 + 4d_1$
$a_5 + b_5 + c_3 + d_3$	$a_7 + b_7 + c_3 + d_3$	$a_9 + b_9 + c_2 + d_2$
$4a_5 + b_5 + 3c_3 + 2d_3$	$4a_7 + b_7 + 3c_3 + 2d_3$	$4a_9 + b_9 + 3c_2 + 2d_2$

3.5 Achievability Proof for the Case $P \leq \frac{M}{2}$

In this section, we describe an achievable scheme for the case $P \leq \frac{M}{2}$. We show that this scheme is optimal when the total number of messages M is an integer multiple of the number of desired messages P. The scheme incurs a small loss from the upper bound for all other cases. The scheme generalizes the ideas in [12]. Different than [12], our scheme uses unequal number of stages for each round of download. Interestingly, the number of stages at each round can be thought of as the output of an all-poles IIR filter. Our scheme reduces to [12] if we let $P = 1$. In the sequel, we define the ith round as the download queries that retrieve sum of i different symbols. We define the stage as a block of queries that exhausts all $\binom{M}{i}$ combinations of the sum of i symbols in the ith round.

3.5.1 Motivating Example: $M = 5$, $P = 2$ Messages, $N = 2$ Databases

To motivate our achievable scheme, consider the case of retrieving two messages denoted by letters (a, b) from five stored messages denoted by letters (a, b, c, d, e). Instead of designing the queries beginning from the top as usual, i.e., beginning by

downloading individual symbols, we design the scheme backwards starting from the last round that corresponds to downloading sums of all five messages and trace back to identify the side information needed at each round from the other database. Our steps described below can be followed through in the query table in Table 3.4.

Now, let us fix the number of stages in the 5th round to be 1 as in [12] since $N = 2$. Round 5 corresponds to downloading the sum of all five messages and contains one combination of symbols $a + b + c + d + e$; please see the last line in Table 3.4. Since we wish to retrieve (a, b), we need one side information equation in the form of $c + d + e$ from earlier rounds. The combination $c + d + e$ can be created directly from round 3 without using round 4. Hence, we suppress round 4, as it does not create any useful side information in our case, and download one stage from round 3 to generate one side information equation $c + d + e$.

In round 3, we download sums of 3 messages. Each stage of round 3 consists of $\binom{5}{3} = 10$ equations. One of those 10 equations is in the desired $c + d + e$ form, and the remaining 9 of them have either a or b or both a, b in them. In tabulating all these 9 combinations, we recognize two categories of side information equations needed from earlier rounds. The first category corresponds to equations of the form $a+b+(c, d, e)$, where (c, d, e) means possible choices for the rest of the equation, i.e., these equations have both a and b in them and plus one more symbol in the form of c or d or e. This category requires downloading one stage of individual symbols (i.e., an individual c or d or e), that is, one stage of round 1. We note also that one of the symbols (a, b) should be known as a side information from the second database in order to solve for the remaining new symbol. The second category corresponds

to equations of the form $a+(c+d, c+e, d+e)$ and $b+(c+d, c+e, d+e)$, i.e., these equations have only one of a or b but not both. This category requires two stages of round 2, as we need different side information equations that contain sum of twos, e.g., $c+d$, $c+e$, $d+e$.

In round 2, we download sums of 2 messages. Each stage of the second round contains $\binom{5}{2} = 10$ equations. In each stage, we need one category of side information equations, which is $a+(c, d, e)$ and $b+(c, d, e)$. This necessitates two different stages of individual symbols, i.e., two stages of round 1 for each stage of round 2.

Denoting α_i to be the number of stages needed for the ith round, we sum all the required stages for round 1 to be $\alpha_1 = 2 \cdot 2 + 1 = 5$ stages. Hence, the user identifies the number stages as $\alpha_1 = 5, \alpha_2 = 2, \alpha_3 = 1, \alpha_4 = 0, \alpha_5 = 1$. These can be observed in the query table in Table 3.4. Note that, we have $\alpha_1 = 5$ stages in round 1 where we download individual bits; then we have $\alpha_2 = 2$ stages in round 2 where we download sums of two symbols; then we have $\alpha_3 = 1$ stage in round 3 where we download sums of three symbols; we skip round 4 as $\alpha_4 = 0$; and we have $\alpha_5 = 1$ stage of round 5 where we download sum of all five symbols.

Now, after designing the structure of the queries and the number of stages needed for each round, we apply the rest of the scheme described in [12]. The user randomly interleaves the messages as usual. In the first round, the user downloads one symbol from each message at each database. This is repeated $\alpha_1 = 5$ times for each database. Hence, the user downloads $a_{1:10}, b_{1:10}, c_{1:10}, d_{1:10}, e_{1:10}$ from the two databases. In the second round, the user downloads sums of two messages. Each stage contains $\binom{5}{2} = 10$ equations. This is repeated $\alpha_2 = 2$ times. For

example, in database 1, user exploits c_6, d_6, e_6 to get a_{12}, a_{13}, a_{14} and c_7, d_7, e_7 to obtain b_{11}, b_{12}, b_{13}. These are from round 1. Round 2 generates $c_{11} + d_{11}$, $c_{12} + e_{11}$, $d_{12} + e_{12}$ from stage 1, and $c_{13} + d_{13}$, $c_{14} + e_{13}$, $d_{14} + e_{14}$ from stage 2 as side information for round 3. In round 3, the user downloads sum of three symbols. There are $\binom{5}{3} = 10$ of them. Symbols c_{10}, d_{10}, e_{10} downloaded from round 1 in database 2 are used to be summed with mixtures of $a + b$. The two sets of side information generated in the second round are exploited in the equations that have one a or b. Note that for each such equation, one of a or b is new and the other one is decoded from database 2. Round 3 generates one side information as $c_{19} + d_{19} + e_{19}$ that is used in round 5. This last round includes the sum of all five messages.

Therefore, as seen in Table 3.4, we have retrieved $a_1, \cdots, a_{34}$ and $b_1 \cdots, b_{34}$, i.e., 68 bits in a total of 112 downloads (56 from each database). Thus, the achievable sum rate is $\frac{68}{112} = \frac{17}{28}$. This is $\underline{R}_s$ in Theorem 3.2, whereas the upper bound $\bar{R}_s$ in Theorem 3.2 is $\frac{1}{1+\frac{1}{N}+\frac{1}{2N^2}} = \frac{8}{13}$. The gap between $\underline{R}_s$ and $\bar{R}_s$ is equal to $\frac{3}{364} \simeq 0.0082$, which also is the largest possible gap between $\underline{R}_s$ and $\bar{R}_s$ over all possible values of M, P and N.

3.5.2 Calculation of the Number of Stages

The main new ingredient of our scheme in comparison to the scheme in [12] is the unequal number of stages in each round. In [12], the scheme is completed in M rounds, and each round contains only 1 stage only when $N = 2$. To generalize the ideas in Section 3.5.1 and calculate the number of stages needed per round, we use

Table 3.4: The query table for the case $M = 5, P = 2, N = 2$.

		Database 1	Database 2
round 1	stg 1	a_1, b_1, c_1, d_1, e_1	a_6, b_6, c_6, d_6, e_6
	stg 2	a_2, b_2, c_2, d_2, e_2	a_7, b_7, c_7, d_7, e_7
	stg 3	a_3, b_3, c_3, d_3, e_3	a_8, b_8, c_8, d_8, e_8
	stg 4	a_4, b_4, c_4, d_4, e_4	a_9, b_9, c_9, d_9, e_9
	stg 5	a_5, b_5, c_5, d_5, e_5	$a_{10}, b_{10}, c_{10}, d_{10}, e_{10}$
round 2	stage 1	$a_{11} + b_6$ $a_{12} + c_6$ $a_{13} + d_6$ $a_{14} + e_6$ $b_{11} + c_7$ $b_{12} + d_7$ $b_{13} + e_7$ $c_{11} + d_{11}$ $c_{12} + e_{11}$ $d_{12} + e_{12}$	$a_{18} + b_1$ $a_{19} + c_1$ $a_{20} + d_1$ $a_{21} + e_1$ $b_{18} + c_2$ $b_{19} + d_2$ $b_{20} + e_2$ $c_{15} + d_{15}$ $c_{16} + e_{15}$ $d_{16} + e_{16}$
	stage 2	$a_6 + b_{14}$ $a_{15} + c_8$ $a_{16} + d_8$ $a_{17} + e_8$ $b_{15} + c_9$ $b_{16} + d_9$ $b_{17} + e_9$ $c_{13} + d_{13}$ $c_{14} + e_{13}$ $d_{14} + e_{14}$	$a_1 + b_{21}$ $a_{22} + c_3$ $a_{23} + d_3$ $a_{24} + e_3$ $b_{22} + c_4$ $b_{23} + d_4$ $b_{24} + e_4$ $c_{17} + d_{17}$ $c_{18} + e_{17}$ $d_{18} + e_{18}$
round 3	stage 1	$a_{25} + b_7 + c_{10}$ $a_7 + b_{25} + d_{10}$ $a_{26} + b_8 + e_{10}$ $a_{27} + c_{15} + d_{15}$ $a_{28} + c_{16} + e_{15}$ $a_{29} + d_{16} + e_{16}$ $b_{26} + c_{17} + d_{17}$ $b_{27} + c_{18} + e_{17}$ $b_{28} + d_{18} + e_{18}$ $c_{19} + d_{19} + e_{19}$	$a_2 + b_{29} + c_5$ $a_{30} + b_2 + d_5$ $a_3 + b_{30} + e_5$ $a_{31} + c_{11} + d_{11}$ $a_{32} + c_{12} + e_{11}$ $a_{33} + d_{12} + e_{12}$ $b_{31} + c_{13} + d_{13}$ $b_{32} + c_{14} + e_{13}$ $b_{33} + d_{14} + e_{14}$ $c_{20} + d_{20} + e_{20}$
rd. 5	stg 1	$a_8 + b_{34} + c_{20} + d_{20} + e_{20}$	$a_{34} + b_3 + c_{19} + d_{19} + e_{19}$

Vandermonde's identity

$$\binom{M}{i} = \sum_{k=0}^{P} \binom{P}{k} \binom{M-P}{i-k} \tag{3.49}$$

The relation in (3.49) states that any combination of i objects from a group of M objects must have k objects from a group of size P and $i-k$ objects from a group of size $M-P$. In our context, the first group is the subset of the desired messages and the second group is the subset of the undesired messages. Then, the relation can be interpreted in our setting as follows: In the ith round, the $\binom{M}{i}$ combinations of all possible sums of i terms can be sorted into $P+1$ categories: The first category (i.e., $k=0$), contains no terms from the desired messages, the second category contains 1 term from the desired messages and $i-1$ terms from the undesired messages, etc. The relation gives also the number of query subgroups of each category $\binom{P}{k}$ and the number of queries in each subgroup $\binom{M-P}{i-k}$.

Let us consider the following concrete example for clarification: Consider that we have 6 messages denoted by (a, b, c, d, e, f), and the desired group to be retrieved is (a, b). Consider round 4 that consists of all combinations of sums of 4 symbols. From Vandermonde's identity, we know that $\binom{6}{4} = \binom{2}{0}\binom{4}{4} + \binom{2}{1}\binom{4}{3} + \binom{2}{2}\binom{4}{2}$. Which means that there are three categories of sums: First category is with only undesired messages; we have $\binom{2}{0} = 1$ query subgroup of the form $c+d+e+f$. The second category is to have 1 term from the desired group and the remaining are undesired; we have $\binom{2}{1} = 2$ query subgroups, one corresponds to a with combinations of 3 terms from c, d, e, f, and the other to b with combinations of 3 terms from c, d, e, f.

Each query subgroup contains $\binom{4}{3}$ queries, i.e., the first query subgroup is of the form $a+(c+d+e, c+d+f, c+e+f, d+e+f)$ and the second query subgroup is of the form $b+(c+d+e, c+d+f, c+e+f, d+e+f)$. Third category is to have 2 terms from the desired group and 2 terms from the undesired group; we have $\binom{2}{2}=1$ query subgroup of this category that takes the form $a+b+(c+d, c+e, \cdots)$. The number of queries of this group is $\binom{4}{2}$ corresponding to all combinations of 2 undesired symbols.

Back to the calculation of the number of stages: To be able to cancel the undesired symbols from an i-term sum, the user needs to download these undesired symbols as side information in the previous rounds. Hence, round i requires downloading $\binom{P}{1}$ stages in round $(i-1)$, $\binom{P}{2}$ stages in round $(i-2)$, etc. Note that these stages need to be downloaded from the remaining $(N-1)$ databases. Then, each database needs to download $\frac{1}{N-1}\binom{P}{1}$ stages in round $(i-1)$, $\frac{1}{N-1}\binom{P}{2}$ stages in round $(i-2)$, etc.

From this observation, we can trace back the number of stages needed at each round. Denote α_i to be the number of stages in round i. Fix the number of stages in the last round (round M) to be $\alpha_M=(N-1)^{M-P}$ stages. This choice ensures that the number of stages in any round is an integer. Note that in round M, the user downloads a sum of all M messages, this requires side information in the form of the sum of the undesired $M-P$ messages. Hence, we suppress the rounds $M-P+1$ through $M-1$ since they do not generate any useful side information. Note that the side information equations in round M at each database are collected from the remaining $(N-1)$ databases. Then, the number of stages in round $(M-P)$ should

be $(N-1)^{M-P-1}$. Therefore, we write

$$\alpha_M = (N-1)^{M-P} \tag{3.50}$$

$$\alpha_{M-1} = \cdots = \alpha_{M-P+1} = 0 \tag{3.51}$$

$$\alpha_{M-P} = \frac{1}{N-1}\alpha_M = \frac{1}{N-1}\sum_{i=1}^{P}\binom{P}{i}\alpha_{M-P+i} \tag{3.52}$$

Now, in round $(M-P)$, each stage requires $\binom{P}{1}$ stages from round $(M-P-1)$, $\binom{P}{2}$ stages from round $(M-P-2)$, and so on so forth, and these stages are divided across $(N-1)$ databases. Continuing with the same argument, for each round, we write

$$\alpha_{M-P-1} = \frac{1}{N-1}\binom{P}{1}\alpha_{M-P} = \frac{1}{N-1}\sum_{i=1}^{P}\binom{P}{i}\alpha_{M-P-1+i} \tag{3.53}$$

$$\begin{aligned}\alpha_{M-P-2} &= \frac{1}{N-1}\binom{P}{1}\alpha_{M-P-1} + \frac{1}{N-1}\binom{P}{2}\alpha_{M-P} \\ &= \frac{1}{N-1}\sum_{i=1}^{P}\binom{P}{i}\alpha_{M-P-2+i}\end{aligned} \tag{3.54}$$

$$\vdots$$

$$\alpha_k = \frac{1}{N-1}\sum_{i=1}^{P}\binom{P}{i}\alpha_{k+i} \tag{3.55}$$

Interestingly, this pattern closely resembles the output of an IIR filter $y[n]$ [54], with the difference equation,

$$y[n] = \frac{1}{N-1}\sum_{i=1}^{P}\binom{P}{i}y[n-i] \tag{3.56}$$

and with the initial conditions $y[-P] = (N-1)^{M-P}$, $y[-P+1] = \cdots = y[-1] = 0$. Note that the only difference between the two seemingly different settings is the orientation of the time axis. The calculation of the number of stages is obtained backwards in contrast to the output of this IIR filter. Hence, we can systematically obtain the number of stages at each round by observing the output of the IIR filter characterized by (3.56), and mapping it to the number of stages via $\alpha_k = y[(M-P)-k]$.

We note that for the special case $P = 1$, the number of stages can be obtained from the first order filter $y[n] = \frac{1}{N-1}y[n-1]$. The output of this filter is $y[n] = (N-1)^{M-2-n}$. Then, the number of stages in round k is $\alpha_k = y[M-1-k] = (N-1)^{k-1}$, which is exactly the number of stages used in [12]; in particular if $N = 2$, then $\alpha_k = 1$ for all k.

3.5.3 General Achievable Scheme

1. *Index preparation:* The user interleaves the contents of each message randomly and independently from the remaining messages using a random interleaver $\pi_m(.)$ which is known privately to the user only, i.e.,

$$x_m(i) = w_m(\pi_m(i)), \quad i \in \{1, \cdots, L\} \tag{3.57}$$

2. *Number of stages:* We calculate the number of stages needed in each round. This can be done systematically by finding the output of the IIR filter char-

acterized by,

$$y[n] = \frac{1}{N-1} \sum_{i=1}^{P} \binom{P}{i} y[n-i] \tag{3.58}$$

with the initial conditions $y[-P] = (N-1)^{M-P}$, $y[-P+1] = \cdots = y[-1] = 0$. The number of stages in round i is $\alpha_i = y[(M-P)-i]$ as discussed in Section 3.5.2.

3. *Initialization:* From the first database, the user downloads one symbol from each message that belongs to the desired message set $\mathcal{P}$. The user sets the round index to $i = 1$.

4. *Message symmetry:* In round i, the user downloads sums of i terms from different symbols from the first database. To satisfy the privacy constraint, the user should download an equal amount of symbols from all messages. Therefore, the user downloads the remaining $\binom{M-P}{i}$ combinations in round i from the undesired symbol set $\bar{\mathcal{P}}$. For example: In round 1, the user downloads one symbol from every undesired message with a total of $\binom{M-P}{1} = M-P$ such symbols.

5. *Repetition of stages:* In the first database, the user repeats the operation in round i according to the number of calculated stages α_i. This in total results in downloading $\alpha_i \binom{M-P}{i}$ undesired equations, and $\alpha_i \left(\binom{M}{i} - \binom{M-P}{i} \right)$ desired equations.

6. *Symmetry across databases:* The user implements symmetry across databases

by downloading $\alpha_i\binom{M-P}{i}$ new undesired equations, and $\alpha_i\left(\binom{M}{i}-\binom{M-P}{i}\right)$ new desired equations from each database. These undesired equations will be used as side information in subsequent rounds. For example: In round 1, each database generates $\alpha_1(M-P)$ undesired equations in the form of individual symbols. Hence, each database can exploit up to $\alpha_1(N-1)(M-P)$ side information equations from other $(N-1)$ databases.

7. *Exploiting side information:* Until now, we did not specify how the desired equations are constructed. Since each stage in round i can be categorized using Vandermonde's identity as in the previous section, we form the desired equations as a sum of the desired symbols and the undesired symbols that can be decoded from other databases in the former $(i-1)$ rounds. If the user sums two or more symbols from $\mathcal{P}$, the user downloads one new symbol from one message only and the remaining symbols from $\mathcal{P}$ should be derived from other databases. Thus, in round $(i+1)$, the user mixes one symbol of $\mathcal{P}$ with the sum of i undesired symbols from round i. This should be repeated for all $\binom{P}{1}$ desired symbols. Then, the user mixes each sum of 2 desired symbols with the sum of $(i-1)$ undesired symbols generated in the $(i-1)$th round. This should be repeated for all the $\binom{P}{2}$ combinations of the desired symbols, and so on.

8. *Repeating steps:* Repeat steps 4, 5, 6, 7 by setting $i=i+1$ until $i=M-P-1$.

9. *Last round:* We note that rounds $M-P+1$ to $M-1$ do not generate useful side information. Hence, $\alpha_{M-P+1}=\cdots=\alpha_{M-1}=0$. In round M, which corresponds to summing all M messages, the user mixes P symbols from $\mathcal{P}$

(only one of them is new and the remaining are previously decoded from the other $(N-1)$ databases) and $M-P$ undesired symbol mixture that was generated in round $(M-P)$.

10. *Shuffling the order of queries:* After preparing the query table, the order of the queries are shuffled uniformly, so that all possible orders of queries are equally likely regardless of $\mathcal{P}$.

3.5.4 Decodability, Privacy, and Calculation of the Achievable Rate

Now, we verify that the proposed scheme satisfies the reliability and privacy constraints.

For the reliability: The scheme is designed to download the *exact* number of undesired equations that will be used as side information equation at subsequent rounds in other databases.[6] Hence, each desired symbol at any round is mixed with a known mixture of symbols that can be decoded from other databases. Note that if the scheme encounters the case of having a mixture of desired symbols, one of them only is chosen to be new and the remaining symbols are downloaded previously from other databases. Thus, the reliability constraint is satisfied by canceling out the side information.

For the privacy: The randomized mapping of message bits and the randomization of the order of queries guarantees privacy as in [12]. It can be checked that

[6]Check for instance in Table 3.4 that all of the downloads (equations) involving undesired symbols from database 2 are used in database 1: singles $c_6, d_6, e_6,\ c_7, d_7, e_7,\ c_8, d_8, e_8,\ c_9, d_9, e_9,\ c_{10}, d_{10}, e_{10}$; sums of twos $c_{15}+d_{15}$, $c_{16}+e_{15}$, $d_{16}+e_{16}$, $c_{17}+d_{17}$, $c_{18}+e_{17}$, $d_{18}+e_{18}$; sum of threes $c_{20}+d_{20}+e_{20}$, all downloaded from database 2 are all used as side information in database 1.

when we fix the queries for one database, we can adjust the queries for the remaining databases such that the user can decode any $\mathcal{P}$ subset of messages. This is true since all combinations of messages are generated by our scheme.

To calculate the achievable rate: From Vandermonde's identity $\binom{M}{i} = \sum_{p=0}^{P} \binom{P}{p}\binom{M-P}{i-p}$, round i requires downloading $\binom{P}{p}$ stages in round $(i-p)$. These stages should be downloaded from the remaining $(N-1)$ databases. Hence, as shown in the previous section, the number of stages at each round is calculated as the output of an IIR filter whose input-output relation is given in (3.56) with the initial conditions $y[-P] = (N-1)^{M-P}$, $y[-P+1] = \cdots = y[-1] = 0$, with the conversion of time index of the filter to the round index of the schemes as $\alpha_i = y[(M-P)-i]$. These initial conditions imply that the user downloads $(N-1)^{M-P}$ stages in the last round that corresponds to downloading the sum of all messages. The $(P-1)$ rounds before the last round are suppressed because we only need to form sums of $(M-P)$ messages to be used in the last round.

Now, to calculate the number of stages for round i, we first solve for the roots of the characteristic equation of (3.56) [54],

$$r^P - \frac{1}{N-1}\sum_{i=1}^{P}\binom{P}{i} r^{P-i} = 0 \tag{3.59}$$

which is equivalent to

$$r^P - \frac{r^P}{N-1}\sum_{i=1}^{P}\binom{P}{i} r^{-i} = 0 \tag{3.60}$$

which further reduces to

$$r^P - \frac{r^P}{N-1}\left[\left(1+\frac{1}{r}\right)^P - 1\right] = 0 \tag{3.61}$$

using the binomial theorem. Simplifying (3.61), we have

$$Nr^P - (r+1)^P = 0 \tag{3.62}$$

By applying the bijective mapping $t = N^{1/P} \cdot \frac{r}{r+1}$, (3.62) is equivalent to $t^P = 1$. The roots for this equation are the normal roots of unity, i.e., $t_k = e^{j2\pi(k-1)/P}$, $k = 1, \cdots, P$, where $j = \sqrt{-1}$. Hence, the roots of the characteristic equation are given by,

$$r_k = \frac{t_k}{N^{1/P} - t_k} = \frac{e^{j2\pi(k-1)/P}}{N^{1/P} - e^{j2\pi(k-1)/P}}, \quad k = 1, \cdots, P \tag{3.63}$$

Thus, the complete response of the IIR filter is given by $y[n] = \sum_{i=1}^{P} \gamma_i r_i^n$, where γ_i are constants that result from solving the initial conditions, i.e., $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_P)^T$ is the solution of the system of equations,

$$\begin{bmatrix} r_1^{-P} & r_2^{-P} & \cdots & r_P^{-P} \\ r_1^{-P+1} & r_2^{-P+1} & \cdots & r_P^{-P+1} \\ \vdots & \vdots & \cdots & \vdots \\ r_1^{-1} & r_2^{-1} & \cdots & r_P^{-1} \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_P \end{bmatrix} = \begin{bmatrix} (N-1)^{M-P} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.64}$$

Now, we are ready to calculate the number of stages α_k in round k. Since $\alpha_k = y[(M-P)-k]$ by construction, then

$$\alpha_k = \sum_{i=1}^{P} \gamma_i r_i^{M-P-k} \tag{3.65}$$

In round k, the user downloads sums of k symbols. The user repeats this round for α_k stages. Each stage contains all the combinations of any k symbols which there are $\binom{M}{k}$ of them. Hence, the total download cost D is,

$$D = \sum_{k=1}^{M} \binom{M}{k} \alpha_k \tag{3.66}$$

$$= \sum_{k=1}^{M} \sum_{i=1}^{P} \binom{M}{k} \gamma_i r_i^{M-P-k} \tag{3.67}$$

$$= \sum_{i=1}^{P} \gamma_i r_i^{M-P} \sum_{k=1}^{M} \binom{M}{k} r_i^{-k} \tag{3.68}$$

$$= \sum_{i=1}^{P} \gamma_i r_i^{M-P} \left[\left(1 + \frac{1}{r_i} \right)^{M} - 1 \right] \tag{3.69}$$

Considering the undesired equations: in round k, the user downloads all combinations of the $(M-P)$ undesired messages which there are $\binom{M-P}{k}$ of them. Therefore, similar to the above calculation, the total number of undesired equations U is,

$$U = \sum_{i=1}^{P} \gamma_i r_i^{M-P} \left[\left(1 + \frac{1}{r_i} \right)^{M-P} - 1 \right] \tag{3.70}$$

Hence, the achievable rate $\underline{R}_s$ is

$$\underline{R}_s = \frac{D-U}{D} \tag{3.71}$$

$$= \frac{\sum_{i=1}^{P} \gamma_i r_i^{M-P} \left[\left(1+\frac{1}{r_i}\right)^M - \left(1+\frac{1}{r_i}\right)^{M-P} \right]}{\sum_{i=1}^{P} \gamma_i r_i^{M-P} \left[\left(1+\frac{1}{r_i}\right)^M - 1 \right]} \tag{3.72}$$

which is (3.31) in Theorem 3.2.

3.5.5 Further Examples for the Case $P \leq \frac{M}{2}$

In this section, we illustrate our proposed scheme with a few additional basic examples. In Section 3.5.1, we considered the case $M = 5$, $P = 2$, $N = 2$. In the next three sub-sections, we consider three more examples. In the example in Section 3.5.5.1, the ratio $\frac{M}{P}$ is exactly equal to 2, thus, both the achievable scheme here and the achievable scheme in Section 3.4 can be used; we comment about the differences and advantages of both schemes. In the example in Section 3.5.5.2, we present the case of a larger N for the example in Section 3.5.1. In the example in Section 3.5.5.3, we present a case with larger M, P and N.

3.5.5.1 $M = 4$ Messages, $P = 2$ Messages, $N = 2$ Databases

The first step of the achievable scheme is to identify the number of stages needed for each round of download. The IIR filter in (3.56) that determines the number of

stages reduces in this case to

$$y[n] = 2y[n-1] + y[n-2] \tag{3.73}$$

with the initial conditions $y[-2] = 1, y[-1] = 0$. The number of stages in round k is $\alpha_k = y[2-k]$. Since M is small, we can calculate the output iteratively without using the canonical filter output as,

$$\alpha_4 = y[-2] = 1 \tag{3.74}$$

$$\alpha_3 = y[-1] = 0 \tag{3.75}$$

$$\alpha_2 = y[0] = 2y[-1] + y[-2] = 1 \tag{3.76}$$

$$\alpha_1 = y[1] = 2y[0] + y[-1] = 2 \tag{3.77}$$

Hence, we should download 2 stages of individual symbols (round 1), and 1 stage of sums of two symbols (round 2). We should suppress the round that retrieves sums of three symbols (round 3), and have 1 stage of sums of all four symbols (round 4).

The user initializes the scheme by randomly and independently interleaving the symbols of each message. The query table for this example is shown in Table 3.5. In round 1, the user downloads individual symbols from all messages at each database. The user downloads a_1, b_1, c_1, d_1 and a_2, b_2, c_2, d_2 from database 1, as $\alpha_1 = 2$. This is repeated for database 2. In round 2, the user downloads sums of two symbols. There are $\binom{4}{2} = 6$ such equations. At database 1, the undesired symbols from database 2 in the first round are exploited in some of these sums. These equations are either

in the form $a+(c,d)$ or in the form $b+(c,d)$. This necessitates two sets of different individual symbols to be downloaded from database 2 in the first round, or otherwise the symbols are repeated and privacy is compromised. Moreover, we note that the user downloads a_5+b_3 which uses b_3 as side information even though W_2 is desired; this is reversed in database 2 to download a_1+b_7 with a_1 as a side information to have a symmetric scheme. Round 2 concludes with downloading c_5+d_5 and c_6+d_6 at the two databases, which will be used as side information in the last round. Round 3 is skipped and the user proceeds to round 4 (last round) directly. In round 4, the user downloads sum of four symbols, and uses the side information downloaded in round 2 and any decoded symbols for the other desired message. For example, in database 1, the user downloads $a_3+b_{10}+c_6+d_6$, hence, the side information c_6+d_6 is exploited in this round as well as a_3. The user finishes the scheme by shuffling the order of all queries randomly. The user retrieves $a_1,\cdots,a_{10}$ and $b_1,\cdots,b_{10}$ privately in 30 downloads (15 from each database) and achieves a sum rate of $\frac{20}{30}=\frac{2}{3}=\frac{1}{1+\frac{1}{N}}$, which matches the upper bound in Theorem 3.2. This sum rate outperforms the repetition-based achievable rate which is $\frac{3}{5}$ in (3.12).

We note that this case can be solved using the achievable scheme presented in Section 3.4 as well since $\frac{M}{P}=2$ in this case. In fact, this is equivalent to the case considered in Section 3.4.4.2, if the number of databases is reduced from $N=3$ to $N=2$. Starting from Table 3.3 in Section 3.4.4.2 and removing the downloads from database 3, we obtain the query table which uses MDS-coded queries shown in Table 3.6 below. Via the scheme in Table 3.6 below, the user retrieves $a_1,\cdots,a_4$ and $b_1,\cdots,b_4$ privately in 12 downloads (6 from each database), therefore achieving

Table 3.5: The query table for the case $M = 4, P = 2, N = 2$.

		Database 1	Database 2
rd. 1	stg 1	a_1, b_1, c_1, d_1	a_3, b_3, c_3, d_3
	stg 2	a_2, b_2, c_2, d_2	a_4, b_4, c_4, d_4
round 2	stage 1	$a_5 + b_3$ $a_6 + c_3$ $a_7 + d_3$ $b_5 + c_4$ $b_6 + d_4$ $c_5 + d_5$	$a_1 + b_7$ $a_8 + c_1$ $a_9 + d_1$ $b_8 + c_2$ $b_9 + d_2$ $c_6 + d_6$
rd. 4	stg 1	$a_3 + b_{10} + c_6 + d_6$	$a_{10} + b_1 + c_5 + d_5$

the same optimal sum rate of $\frac{8}{12} = \frac{2}{3} = \frac{1}{1+\frac{1}{N}}$.

Table 3.6: Alternative query table for the case $M = 4, P = 2, N = 2$.

Database 1	Database 2
a_1, b_1, c_1, d_1	a_2, b_2, c_2, d_2
$a_3 + b_3 + c_2 + d_2$	$a_4 + b_4 + c_1 + d_1$
$a_3 + 3b_3 + 2c_2 + 4d_2$	$a_4 + 3b_4 + 2c_1 + 4d_1$

We presented this case here even though it could be solved using the scheme in Section 3.4, in order to give an example where the second achievable scheme achieves the upper bound in Theorem 3.2 and yields a capacity result since $\frac{M}{P}$ is an integer. Interestingly, we observe that for all cases where $P = \frac{M}{2}$, the two achievable schemes are both optimal. The two schemes present an interesting trade-off between the field size and the upload cost: The first achievable scheme in Section 3.4 requires using an MDS code with field size $q \geq M$ but the number of queries for each database is limited to $M + P$. On the other hand, the second achievable scheme here in Section 3.5 does not use any coding and can work with the storage field size, however, the number of queries increase exponentially since the number of stages for each round is related to an unstable IIR filter.

3.5.5.2 $M = 5$ Messages, $P = 2$ Messages, $N = 3$ Databases

In this example, we show an explicit query structure for $N > 2$. In this case the corresponding difference equation for the IIR filter is

$$y[n] = y[n-1] + \frac{1}{2}y[n-2] \tag{3.78}$$

with the initial conditions $y[-1] = 0$, $y[-2] = (N-1)^{M-P} = 8$. Thus, the number of stages in each round are: $\alpha_1 = 6$, $\alpha_2 = 4$, $\alpha_3 = 4$, $\alpha_4 = 0, \alpha_5 = 8$. The query table is shown in Tables 3.7, 3.8 and 3.9. This scheme retrieves $a_1, \cdots, a_{126}$ and $b_1, \cdots, b_{126}$ privately in 354 downloads (177 from each database), therefore, achieving a sum rate of $\frac{252}{354} = \frac{42}{59} < \frac{1}{1+\frac{1}{N}+\frac{1}{2N^2}} = \frac{18}{25}$. The gap is $\frac{12}{1475} \simeq 0.0081$.

3.5.5.3 $M = 7$ Messages, $P = 3$ Messages, $N = 3$ Databases

Finally, in this section, we consider an example with $N = 3$ databases and larger M and P than in previous examples, where we describe the structure and the calculation of the number of queries without specifying the explicit query table as it grows quite large. We first calculate the number of stages at each round. The corresponding IIR filter is

$$y[n] = \frac{1}{2}(3y[n-1] + 3y[n-2] + y[n-3]) \tag{3.79}$$

Table 3.7: The query table for the case $M = 5, P = 2, N = 3$.

		Database 1	Database 2	Database 3
round 1	stg 1	a_1, b_1, c_1, d_1, e_1	a_7, b_7, c_7, d_7, e_7	$a_{13}, b_{13}, c_{13}, d_{13}, e_{13}$
	stg 2	a_2, b_2, c_2, d_2, e_2	a_8, b_8, c_8, d_8, e_8	$a_{14}, b_{14}, c_{14}, d_{14}, e_{14}$
	stg 3	a_3, b_3, c_3, d_3, e_3	a_9, b_9, c_9, d_9, e_9	$a_{15}, b_{15}, c_{15}, d_{15}, e_{15}$
	stg 4	a_4, b_4, c_4, d_4, e_4	$a_{10}, b_{10}, c_{10}, d_{10}, e_{10}$	$a_{16}, b_{16}, c_{16}, d_{16}, e_{16}$
	stg 5	a_5, b_5, c_5, d_5, e_5	$a_{11}, b_{11}, c_{11}, d_{11}, e_{11}$	$a_{17}, b_{17}, c_{17}, d_{17}, e_{17}$
	stg 6	a_6, b_6, c_6, d_6, e_6	$a_{12}, b_{12}, c_{12}, d_{12}, e_{12}$	$a_{18}, b_{18}, c_{18}, d_{18}, e_{18}$
round 2	stage 1	$a_{19} + b_7$	$a_{33} + b_1$	$a_{47} + b_1$
		$a_{20} + c_7$	$a_{34} + c_1$	$a_{48} + c_1$
		$a_{21} + d_7$	$a_{35} + d_1$	$a_{49} + d_1$
		$a_{22} + e_7$	$a_{36} + e_1$	$a_{50} + e_1$
		$b_{19} + c_8$	$b_{33} + c_2$	$b_{47} + c_2$
		$b_{20} + d_8$	$b_{34} + d_2$	$b_{48} + d_2$
		$b_{21} + e_8$	$b_{35} + e_2$	$b_{49} + e_2$
		$c_{19} + d_{19}$	$c_{27} + d_{27}$	$c_{35} + d_{35}$
		$c_{20} + e_{19}$	$c_{28} + e_{27}$	$c_{36} + e_{35}$
		$d_{20} + e_{20}$	$d_{28} + e_{28}$	$d_{36} + e_{36}$
	stage 2	$a_7 + b_{22}$	$a_1 + b_{36}$	$a_1 + b_{50}$
		$a_{23} + c_9$	$a_{37} + c_3$	$a_{51} + c_3$
		$a_{24} + d_9$	$a_{38} + d_3$	$a_{52} + d_3$
		$a_{25} + e_9$	$a_{39} + e_3$	$a_{53} + e_3$
		$b_{23} + c_{10}$	$b_{37} + c_4$	$b_{51} + c_4$
		$b_{24} + d_{10}$	$b_{38} + d_4$	$b_{52} + d_4$
		$b_{25} + e_{10}$	$b_{39} + e_4$	$b_{53} + e_4$
		$c_{21} + d_{21}$	$c_{29} + d_{29}$	$c_{37} + d_{37}$
		$c_{22} + e_{21}$	$c_{30} + e_{29}$	$c_{38} + e_{37}$
		$d_{22} + e_{22}$	$d_{30} + e_{30}$	$d_{38} + e_{38}$
	stage 3	$a_{26} + b_{13}$	$a_{40} + b_{13}$	$a_{54} + b_7$
		$a_{27} + c_{13}$	$a_{41} + c_{13}$	$a_{55} + c_7$
		$a_{28} + d_{13}$	$a_{42} + d_{13}$	$a_{56} + d_7$
		$a_{29} + e_{13}$	$a_{43} + e_{13}$	$a_{57} + e_7$
		$b_{26} + c_{14}$	$b_{40} + c_{14}$	$b_{54} + c_8$
		$b_{27} + d_{14}$	$b_{41} + d_{14}$	$b_{55} + d_8$
		$b_{28} + e_{14}$	$b_{42} + e_{14}$	$b_{56} + e_8$
		$c_{23} + d_{23}$	$c_{31} + d_{31}$	$c_{39} + d_{39}$
		$c_{24} + e_{23}$	$c_{32} + e_{31}$	$c_{40} + e_{39}$
		$d_{24} + e_{24}$	$d_{32} + e_{32}$	$d_{40} + e_{40}$

Table 3.8: The query table for the case $M = 5, P = 2, N = 3$ (cont.).

		Database 1	Database 2	Database 3
round 2	stage 4	$a_{13} + b_{29}$	$a_{13} + b_{43}$	$a_7 + b_{57}$
		$a_{30} + c_{15}$	$a_{44} + c_{15}$	$a_{58} + c_9$
		$a_{31} + d_{15}$	$a_{45} + d_{15}$	$a_{59} + d_9$
		$a_{32} + e_{15}$	$a_{46} + e_{15}$	$a_{60} + e_9$
		$b_{30} + c_{16}$	$b_{44} + c_{16}$	$b_{58} + c_{10}$
		$b_{31} + d_{16}$	$b_{45} + d_{16}$	$b_{59} + d_{10}$
		$b_{32} + e_{16}$	$b_{46} + e_{16}$	$b_{60} + e_{10}$
		$c_{25} + d_{25}$	$c_{33} + d_{33}$	$c_{41} + d_{41}$
		$c_{26} + e_{25}$	$c_{34} + e_{33}$	$c_{42} + e_{41}$
		$d_{26} + e_{26}$	$d_{34} + e_{34}$	$d_{42} + e_{42}$
round 3	stage 1	$a_{61} + b_8 + c_{11}$	$a_{79} + b_2 + c_5$	$a_{97} + b_2 + c_5$
		$a_8 + b_{61} + d_{11}$	$a_2 + b_{79} + d_5$	$a_2 + b_{97} + d_5$
		$a_{62} + b_9 + e_{11}$	$a_{80} + b_3 + e_5$	$a_{98} + b_3 + e_5$
		$a_{63} + c_{27} + d_{27}$	$a_{81} + c_{19} + d_{19}$	$a_{99} + c_{19} + d_{19}$
		$a_{64} + c_{28} + e_{27}$	$a_{82} + c_{20} + e_{19}$	$a_{100} + c_{20} + e_{19}$
		$a_{65} + d_{28} + e_{28}$	$a_{83} + d_{20} + e_{20}$	$a_{101} + d_{20} + e_{20}$
		$b_{62} + c_{29} + d_{29}$	$b_{80} + c_{21} + d_{21}$	$b_{98} + c_{21} + d_{21}$
		$b_{63} + c_{30} + e_{29}$	$b_{81} + c_{22} + e_{21}$	$b_{99} + c_{22} + e_{21}$
		$b_{64} + d_{30} + e_{30}$	$b_{82} + d_{22} + e_{22}$	$b_{100} + d_{22} + e_{22}$
		$c_{43} + d_{43} + e_{43}$	$c_{47} + d_{47} + e_{47}$	$c_{51} + d_{51} + e_{51}$
	stage 2	$a_9 + b_{65} + c_{12}$	$a_3 + b_{83} + c_6$	$a_3 + b_{101} + c_6$
		$a_{66} + b_{10} + d_{12}$	$a_{84} + b_4 + d_6$	$a_{102} + b_4 + d_6$
		$a_{10} + b_{66} + e_{12}$	$a_4 + b_{84} + e_6$	$a_4 + b_{102} + e_6$
		$a_{67} + c_{31} + d_{31}$	$a_{85} + c_{23} + d_{23}$	$a_{103} + c_{23} + d_{23}$
		$a_{68} + c_{32} + e_{31}$	$a_{86} + c_{24} + e_{23}$	$a_{104} + c_{24} + e_{23}$
		$a_{69} + d_{32} + e_{32}$	$a_{87} + d_{24} + e_{24}$	$a_{105} + d_{24} + e_{24}$
		$b_{67} + c_{33} + d_{33}$	$b_{85} + c_{25} + d_{25}$	$b_{103} + c_{25} + d_{25}$
		$b_{68} + c_{34} + e_{33}$	$b_{86} + c_{26} + e_{25}$	$b_{104} + c_{26} + e_{25}$
		$b_{69} + d_{34} + e_{34}$	$b_{87} + d_{26} + e_{26}$	$b_{105} + d_{26} + e_{26}$
		$c_{44} + d_{44} + e_{44}$	$c_{48} + d_{48} + e_{48}$	$c_{52} + d_{52} + e_{52}$
	stage 3	$a_{70} + b_{14} + c_{17}$	$a_{88} + b_{14} + c_{17}$	$a_{106} + b_8 + c_{11}$
		$a_{14} + b_{70} + d_{17}$	$a_{14} + b_{88} + d_{17}$	$a_8 + b_{106} + d_{11}$
		$a_{71} + b_{15} + e_{17}$	$a_{89} + b_{15} + e_{17}$	$a_{107} + b_9 + e_{11}$
		$a_{72} + c_{35} + d_{35}$	$a_{90} + c_{35} + d_{35}$	$a_{108} + c_{27} + d_{27}$
		$a_{73} + c_{36} + e_{35}$	$a_{91} + c_{36} + e_{35}$	$a_{109} + c_{28} + e_{27}$
		$a_{74} + d_{36} + e_{36}$	$a_{92} + d_{36} + e_{36}$	$a_{110} + d_{28} + e_{28}$
		$b_{71} + c_{37} + d_{37}$	$b_{89} + c_{37} + d_{37}$	$b_{107} + c_{29} + d_{29}$
		$b_{72} + c_{38} + e_{37}$	$b_{90} + c_{38} + e_{37}$	$b_{108} + c_{30} + e_{29}$
		$b_{73} + d_{38} + e_{38}$	$b_{91} + d_{38} + e_{38}$	$b_{109} + d_{30} + e_{30}$
		$c_{45} + d_{45} + e_{45}$	$c_{49} + d_{49} + e_{49}$	$c_{53} + d_{53} + e_{53}$

Table 3.9: The query table for the case $M = 5, P = 2, N = 3$ (cont.).

		Database 1	Database 2	Database 3
round 3	stage 4	$a_{15} + b_{74} + c_{18}$ $a_{75} + b_{16} + d_{18}$ $a_{16} + b_{75} + e_{18}$ $a_{76} + c_{39} + d_{39}$ $a_{77} + c_{40} + e_{39}$ $a_{78} + d_{40} + e_{40}$ $b_{76} + c_{41} + d_{41}$ $b_{77} + c_{42} + e_{41}$ $b_{78} + d_{42} + e_{42}$ $c_{46} + d_{46} + e_{46}$	$a_{15} + b_{92} + c_{18}$ $a_{93} + b_{16} + d_{18}$ $a_{16} + b_{93} + e_{18}$ $a_{94} + c_{39} + d_{39}$ $a_{95} + c_{40} + e_{39}$ $a_{96} + d_{40} + e_{40}$ $b_{94} + c_{41} + d_{41}$ $b_{95} + c_{42} + e_{41}$ $b_{96} + d_{42} + e_{42}$ $c_{50} + d_{50} + e_{50}$	$a_9 + b_{110} + c_{12}$ $a_{111} + b_{10} + d_{12}$ $a_{10} + b_{111} + e_{12}$ $a_{112} + c_{31} + d_{31}$ $a_{113} + c_{32} + e_{31}$ $a_{114} + d_{32} + e_{32}$ $b_{112} + c_{33} + d_{33}$ $b_{113} + c_{34} + e_{33}$ $b_{114} + d_{34} + e_{34}$ $c_{54} + d_{54} + e_{54}$
round 5	stg 1	$a_{115} + b_{11} + c_{47} + d_{47} + e_{47}$	$a_{119} + b_5 + c_{43} + d_{43} + e_{43}$	$a_{123} + b_5 + c_{43} + d_{43} + e_{43}$
	stg 2	$a_{11} + b_{115} + c_{48} + d_{48} + e_{48}$	$a_5 + b_{119} + c_{44} + d_{44} + e_{44}$	$a_5 + b_{123} + c_{44} + d_{44} + e_{44}$
	stg 3	$a_{116} + b_{12} + c_{49} + d_{49} + e_{49}$	$a_{120} + b_6 + c_{45} + d_{45} + e_{45}$	$a_{124} + b_6 + c_{45} + d_{45} + e_{45}$
	stg 4	$a_{12} + b_{116} + c_{50} + d_{50} + e_{50}$	$a_6 + b_{120} + c_{46} + d_{46} + e_{46}$	$a_6 + b_{124} + c_{46} + d_{46} + e_{46}$
	stg 5	$a_{117} + b_{17} + c_{51} + d_{51} + e_{51}$	$a_{121} + b_{17} + c_{51} + d_{51} + e_{51}$	$a_{125} + b_{11} + c_{47} + d_{47} + e_{47}$
	stg 6	$a_{17} + b_{117} + c_{52} + d_{52} + e_{52}$	$a_{17} + b_{121} + c_{52} + d_{52} + e_{52}$	$a_{11} + b_{125} + c_{48} + d_{48} + e_{48}$
	stg 7	$a_{118} + b_{18} + c_{53} + d_{53} + e_{53}$	$a_{122} + b_{18} + c_{53} + d_{53} + e_{53}$	$a_{126} + b_{12} + c_{49} + d_{49} + e_{49}$
	stg 8	$a_{18} + b_{118} + c_{54} + d_{54} + e_{54}$	$a_{18} + b_{122} + c_{54} + d_{54} + e_{54}$	$a_{12} + b_{126} + c_{50} + d_{50} + e_{50}$

with the initial conditions $y[-3] = (N-1)^{M-P} = 16$, $y[-2] = 0$, $y[-1] = 0$. Hence, the number of stages for each round $\alpha_k = y[4-k]$, $k = 1, \cdots, 7$, are calculated iteratively as $\alpha_1 = 67$, $\alpha_2 = 30$, $\alpha_3 = 12$, $\alpha_4 = 8$, $\alpha_5 = 0$, $\alpha_6 = 0$, $\alpha_7 = 16$.

In round 1, the user downloads 67 individual symbols from each message and from each database. Each database can use the side information generated by the other two databases. Hence, each database has $67 \cdot 2 = 134$ side information equations in the form of single symbols from round 1 to exploit. In round 2, the user downloads sums of two symbols. Each stage in round 2 requires 3 stages from round 1, since the user faces with $a+(d, e, f, g)$, $b+(d, e, f, g)$ or $c+(d, e, f, g)$ cases. Then, round 2 requires $30 \cdot 3 = 90$ stages from the generated side information in round 1, and we are left with $134 - 90 = 44$ more stages of round 1. Each database can use the side information stages from the other two databases, i.e., each can use up to $2 \cdot 30 = 60$ stages of side information in the form of sums of two.

In round 3, the user downloads sums of three symbols, which can be either of $a+b+(d,e,f,g)$, $a+c+(d,e,f,g)$, $b+c+(d,e,f,g)$, $a+(d+e,d+f,\cdots)$, and similarly for b,c. Therefore, each stage in round 3 requires 3 stages from round 2, and 3 stages from round 1. This in total requires $12\cdot 3=36$ stages from round 1 and 36 stages from round 2, and we will be left with 8 stages from round 1 and 24 stages from round 2. Round 3 generates $2\cdot 12=24$ stages of side information in the form of sums of threes. In round 4, the user downloads sums of 4 symbols, which can be either $a+b+(d+e,d+f,\cdots)$, and similarly for $b+c$ and $a+c$, $a+(d+e+f,d+e+g,\cdots)$ and similarly for b, c, or $a+b+c+(d,e,f,g)$. This means that for each stage of round 3, the user needs 1 stage of round 1, 3 stages of round 2, and 3 stages of round 3. This in total requires $8\cdot 3=24$ stages from round 2 and 3 and $8\cdot 1$ stages from round 1 and hence, we exhaust all the generated side information by round 4. Round 4 generates 8 stages of side information in the form of sums of fours. This will be used in the last round to get $8\cdot 2$ new symbols from the desired messages.

The achievable sum rate in this case is $\frac{3933}{5445}=\frac{437}{605}<\frac{1}{1+\frac{1}{N}+\frac{1}{3N^2}}=\frac{27}{37}$. The gap is $\frac{166}{22385}\simeq 0.0074$.

3.6 Converse Proof

In this section, we derive an upper bound for the MPIR problem[7]. The derived upper bound is tight when $P\geq\frac{M}{2}$ and when $\frac{M}{P}\in\mathbb{N}$. We follow the notations and

[7]We note that the assumption that $W_i\in\mathbb{F}_q^L$ is indeed unnecessary in terms of converse arguments. Consequently, our converse proof is valid for any storage alphabet.

simplifications in [12, 117], and we define

$$\mathcal{Q} \triangleq \left\{ Q_n^{[\mathcal{P}]} : \mathcal{P} \subseteq \{1,\cdots,M\},\ |\mathcal{P}| = P,\ n \in \{1,\cdots,N\} \right\} \tag{3.80}$$

and

$$A_{n_1:n_2}^{[\mathcal{P}]} \triangleq \left\{ A_{n_1}^{[\mathcal{P}]}, A_{n_1+1}^{[\mathcal{P}]}, \cdots, A_{n_2}^{[\mathcal{P}]} \right\} \tag{3.81}$$

for $n_1 \leq n_2,\ n_1, n_2 \in \{1, \cdots, N\}$.

Without loss of generality, the following simplifications hold for the MPIR problem:

1. We can assume that the MPIR scheme is symmetric. Since for every asymmetric scheme, there exists an equal rate symmetric scheme that can be constructed by replicating all permutations of databases and messages.

2. To invoke the privacy constraint, we fix the response of one database to be the same irrespective of the desired set of messages $\mathcal{P}$, i.e., $A_n^{[\mathcal{P}_i]} = A_n$, where $|\mathcal{P}_i| = P$ for every $i \in \{1, 2, \cdots, \beta\}$ for some $n \in \{1, \cdots, N\}$, and $\beta = \binom{M}{P}$. No loss of generality is incurred due to the fact that the queries and the answers are statistically independent from $\mathcal{P}$. In the sequel, we fix the answer string of the first database, i.e.,

$$A_1^{[\mathcal{P}]} = A_1,\ \forall \mathcal{P} \tag{3.82}$$

The following lemma is a consequence of the symmetry assumption; its proof can be found in [12].

Lemma 3.1 (Symmetry [12]) *For any* $W_{\mathcal{S}} = \{W_i : i \in \mathcal{S}\}$

$$H(A_n^{[\mathcal{P}]}|W_{\mathcal{S}}, \mathcal{Q}) = H(A_1^{[\mathcal{P}]}|W_{\mathcal{S}}, \mathcal{Q}), \quad n \in \{1,\cdots,N\} \tag{3.83}$$

$$H(A_1|\mathcal{Q}) = H(A_n^{[\mathcal{P}]}|\mathcal{Q}), \quad n \in \{1,\cdots,N\},\ \forall \mathcal{P} \tag{3.84}$$

We construct the converse proof by induction over $\lfloor \frac{M}{P} \rfloor$ in a similar way to [12,117]. The base induction step is obtained for $1 \leq \frac{M}{P} \leq 2$ (this is the case $P \geq \frac{M}{2}$ as it was referred to so far, where the user wants to retrieve at least half of the messages). We obtain an inductive relation for the case $\frac{M}{P} > 2$. The converse proof extends the proof in [12] for $P > 1$.

3.6.1 Converse Proof for the Case $1 \leq \frac{M}{P} \leq 2$

To prove the converse for the case $1 \leq \frac{M}{P} \leq 2$, we need the following lemma which gives a lower bound on the interference within an answer string.

Lemma 3.2 (Interference Lower Bound) *For the MPIR problem with* $P \geq \frac{M}{2}$, *the uncertainty of the interfering messages* $W_{P+1:M}$ *within the answer string* $A_1^{[1:P]}$ *is lower bounded as,*

$$H(A_1^{[1:P]}|W_{1:P}, \mathcal{Q}) \geq \frac{(M-P)L}{N} \tag{3.85}$$

Furthermore, (3.85) is true for any set of desired messages $\mathcal{P}$ with $|\mathcal{P}| = P$, i.e.,

$$H(A_1^{[\mathcal{P}]}|W_{\mathcal{P}}, \mathcal{Q}) \geq \frac{(M-P)L}{N} \tag{3.86}$$

Proof: For clarity of presentation, we assume that $\mathcal{P} = \{1, \cdots, P\}$ without loss of generality. Hence,

$$\begin{aligned}
(M-P)L &= H(W_{P+1:M}) && (3.87)\\
&= H(W_{P+1:M}|W_{1:P}, \mathcal{Q}) && (3.88)\\
&= H(W_{P+1:M}|W_{1:P}, \mathcal{Q}) - H(W_{P+1:M}|A_{1:N}^{[M-P+1:M]}, W_{1:P}, \mathcal{Q}) && (3.89)\\
&= I(W_{P+1:M}; A_{1:N}^{[M-P+1:M]}|W_{1:P}, \mathcal{Q}) && (3.90)\\
&= H(A_{1:N}^{[M-P+1:M]}|W_{1:P}, \mathcal{Q}) && (3.91)\\
&\leq \sum_{n=1}^{N} H(A_n^{[M-P+1:M]}|W_{1:P}, \mathcal{Q}) && (3.92)\\
&= NH(A_1|W_{1:P}, \mathcal{Q}) && (3.93)
\end{aligned}$$

where (3.88) follows from the independence of the messages $W_{P+1:M}$ from the messages $W_{1:P}$ and the queries as in (3.2) and (3.3); (3.89) follows from the reliability constraint (3.7), since messages $W_{P+1:M}$ can be decoded correctly from the answer strings $A_{1:N}^{[M-P+1:M]}$ if $P \geq \frac{M}{2}$ as $\{P+1, \cdots, M\} \subseteq \{M-P+1, \cdots, M\}$ in this regime; (3.91) follows from the fact that the answer strings are deterministic functions of all messages and queries $(\mathcal{Q}, W_{1:M})$; and (3.93) follows from the independence bound and Lemma 3.1.

Consequently, $H(A_1|W_{1:P}, \mathcal{Q}) \geq \frac{(M-P)L}{N}$. The proof of the general statement can be done replacing $W_{1:P}$ by $W_{\mathcal{P}}$, $W_{P+1:M}$ by $W_{\bar{\mathcal{P}}}$ which corresponds to the complement set of messages of $W_{\mathcal{P}}$, and the answer strings $A_{1:N}^{[M-P+1:M]}$ by $A_{1:N}^{[\mathcal{P}^*]}$, where $\bar{\mathcal{P}} \subseteq \mathcal{P}^*$, $|\mathcal{P}^*| = P$. ■

Now, we are ready to prove the converse of the case $P \geq \frac{M}{2}$. We use a similar converse technique to the case of $M = 2, P = 1$ in [12],

$$
\begin{aligned}
ML &= H(W_{1:M}) && (3.94)\\
&= H(W_{1:M}|\mathcal{Q}) && (3.95)\\
&= H(W_{1:M}|\mathcal{Q}) - H(W_{1:M}|A_{1:N}^{[\mathcal{P}_1]}, A_{1:N}^{[\mathcal{P}_2]}, \cdots, A_{1:N}^{[\mathcal{P}_\beta]}, \mathcal{Q}) && (3.96)\\
&= I(W_{1:M}; A_{1:N}^{[\mathcal{P}_1]}, A_{1:N}^{[\mathcal{P}_2]}, \cdots, A_{1:N}^{[\mathcal{P}_\beta]}|\mathcal{Q}) && (3.97)\\
&= H(A_{1:N}^{[\mathcal{P}_1]}, A_{1:N}^{[\mathcal{P}_2]}, \cdots, A_{1:N}^{[\mathcal{P}_\beta]}|\mathcal{Q}) && (3.98)\\
&= H(A_1, A_{2:N}^{[\mathcal{P}_1]}, A_{2:N}^{[\mathcal{P}_2]}, \cdots, A_{2:N}^{[\mathcal{P}_\beta]}|\mathcal{Q}) && (3.99)\\
&= H(A_1, A_{2:N}^{[\mathcal{P}_1]}|\mathcal{Q}) + H(A_{2:N}^{[\mathcal{P}_2]}, \cdots, A_{2:N}^{[\mathcal{P}_\beta]}|A_1, A_{2:N}^{[\mathcal{P}_1]}, \mathcal{Q}) && (3.100)\\
&= H(A_1, A_{2:N}^{[\mathcal{P}_1]}|\mathcal{Q}) + H(A_{2:N}^{[\mathcal{P}_2]}, \cdots, A_{2:N}^{[\mathcal{P}_\beta]}|A_1, A_{2:N}^{[\mathcal{P}_1]}, W_{\mathcal{P}_1}, \mathcal{Q}) && (3.101)\\
&\leq \sum_{n=1}^{N} H(A_n^{[\mathcal{P}_1]}|\mathcal{Q}) + H(A_{2:N}^{[\mathcal{P}_2]}, \cdots, A_{2:N}^{[\mathcal{P}_\beta]}|A_1, W_{\mathcal{P}_1}, \mathcal{Q}) && (3.102)\\
&= \sum_{n=1}^{N} H(A_n^{[\mathcal{P}_1]}|\mathcal{Q}) + H(A_{1:N}^{[\mathcal{P}_2]}, \cdots, A_{1:N}^{[\mathcal{P}_\beta]}|W_{\mathcal{P}_1}, \mathcal{Q}) - H(A_1|W_{\mathcal{P}_1}, \mathcal{Q}) && (3.103)
\end{aligned}
$$

where $\beta = \binom{M}{P}$ represents the total number of message subsets of size P that can be constructed from M messages; (3.95) follows from the independence between the messages and the queries; (3.96) follows from the reliability constraint in (3.7) with noting that $A_{1:N}^{[\mathcal{P}_1]}, A_{1:N}^{[\mathcal{P}_2]}, \cdots, A_{1:N}^{[\mathcal{P}_\beta]}$ represent all answer strings from all databases

to every possible subset of messages $\mathcal{P}_i \subseteq \{1, \cdots, M\}$, $i = 1, 2, \cdots, \beta$, hence all messages can be correctly decoded as all possible answer strings are known; (3.98) follows from the fact that answer strings are deterministic functions of the messages and the queries; (3.99) follows from simplification (3.82) without loss of generality; (3.101) follows from the fact that the messages $W_{\mathcal{P}} = (W_{i_1}, W_{i_2}, \cdots, W_{i_P})$ can be reconstructed from $A_{1:N}^{[\mathcal{P}]}$; and (3.102) is a consequence of the fact that conditioning does not increase entropy and Lemma 3.1.

Now, every message appears in $\binom{M-1}{P-1}$ different message subsets of size P, therefore the answer strings $(A_{1:N}^{[\mathcal{P}_2]}, \cdots, A_{1:N}^{[\mathcal{P}_\beta]})$ are sufficient to construct all messages $W_{1:M}$ irrespective of $\mathcal{P}_1$. Therefore,

$$H(A_{1:N}^{[\mathcal{P}_2]}, \cdots, A_{1:N}^{[\mathcal{P}_\beta]} | W_{\mathcal{P}_1}, \mathcal{Q}) = (M - P)L \tag{3.104}$$

Using this and Lemma 3.2 in (3.103) yields

$$ML \leq \sum_{n=1}^{N} H(A_n^{[\mathcal{P}_1]} | \mathcal{Q}) + (M - P)L - \frac{(M - P)L}{N} \tag{3.105}$$

which can be written as,

$$PL + \frac{(M - P)L}{N} \leq \sum_{n=1}^{N} H(A_n^{[\mathcal{P}_1]} | \mathcal{Q}) \tag{3.106}$$

which further can be written as,

$$\left(1+\frac{M-P}{PN}\right)PL \leq \sum_{n=1}^{N} H(A_n^{[\mathcal{P}_1]}|\mathcal{Q}) \tag{3.107}$$

which leads to the desired converse result,

$$\begin{aligned}\sum_{i=1}^{P} R_i &= \frac{PL}{\sum_{n=1}^{N} H\left(A_n^{[\mathcal{P}]}\right)} \\ &\leq \frac{PL}{\sum_{n=1}^{N} H\left(A_n^{[\mathcal{P}]}|\mathcal{Q}\right)} \\ &\leq \frac{1}{1+\frac{M-P}{PN}}\end{aligned} \tag{3.108}$$

3.6.2 Converse Proof for the Case $\frac{M}{P} > 2$

In the sequel, we derive an inductive relation that can be used in addition to the base induction step of $1 \leq \frac{M}{P} \leq 2$ derived in the previous sub-section to obtain an upper bound for the MPIR problem. The idea we pursue here is similar in spirit to the one in [12], where the authors developed a base converse step for $M = 2$ messages, and developed an induction over the number of messages M for the case $M > 2$. Here, we have developed a base converse step for $1 \leq \frac{M}{P} \leq 2$, and now develop an induction over $\left\lfloor \frac{M}{P} \right\rfloor$ for the case $\frac{M}{P} > 2$.

The following lemma upper bounds the remaining uncertainty of the answer strings after knowing a subset of size P of the interference messages.

Lemma 3.3 (Interference Conditioning Lemma) *The remaining uncertainty*

in the answer strings $A_{2:N}^{[\mathcal{P}_2]}$ *after conditioning on the messages indexed by* $\mathcal{P}_1$*, such that* $\mathcal{P}_1 \cap \mathcal{P}_2 = \phi$*,* $|\mathcal{P}_1| = |\mathcal{P}_2| = P$ *is upper bounded by,*

$$H(A_{2:N}^{[\mathcal{P}_2]}|W_{\mathcal{P}_1}, \mathcal{Q}) \leq (N-1)[NH(A_1|\mathcal{Q}) - PL] \tag{3.109}$$

Proof: We begin with

$$\begin{aligned}
&H(A_{2:N}^{[\mathcal{P}_2]}|W_{\mathcal{P}_1}, \mathcal{Q}) \\
&\quad \leq \sum_{n=2}^{N} H(A_n^{[\mathcal{P}_2]}|W_{\mathcal{P}_1}, \mathcal{Q}) && (3.110)\\
&\quad \leq \sum_{n=2}^{N} H(A_{1:n-1}^{[\mathcal{P}_1]}, A_n^{[\mathcal{P}_2]}, A_{n+1:N}^{[\mathcal{P}_1]}|W_{\mathcal{P}_1}, \mathcal{Q}) && (3.111)\\
&\quad = \sum_{n=2}^{N} H(A_{1:n-1}^{[\mathcal{P}_1]}, A_n^{[\mathcal{P}_2]}, A_{n+1:N}^{[\mathcal{P}_1]}, W_{\mathcal{P}_1}|\mathcal{Q}) - H(W_{\mathcal{P}_1}|\mathcal{Q}) && (3.112)\\
&\quad = \sum_{n=2}^{N} H(A_{1:n-1}^{[\mathcal{P}_1]}, A_n^{[\mathcal{P}_2]}, A_{n+1:N}^{[\mathcal{P}_1]}|\mathcal{Q}) + H(W_{\mathcal{P}_1}|A_{1:n-1}^{[\mathcal{P}_1]}, A_n^{[\mathcal{P}_2]}, A_{n+1:N}^{[\mathcal{P}_1]}) - H(W_{\mathcal{P}_1}) \\
& && (3.113)\\
&\quad \leq \sum_{n=2}^{N} NH(A_1|\mathcal{Q}) - H(W_{\mathcal{P}_1}) && (3.114)\\
&\quad = (N-1)[NH(A_1|\mathcal{Q}) - PL] && (3.115)
\end{aligned}$$

where (3.110) follows from the independence bound; (3.111) follows from the non-negativity of entropy; (3.113) follows from the statistical independence between the messages and the queries; and (3.114) follows from the decodability of $W_{\mathcal{P}_1}$ given the answer strings $(A_{1:n-1}^{[\mathcal{P}_1]}, A_n^{[\mathcal{P}_2]}, A_{n+1:N}^{[\mathcal{P}_1]})$, which is tantamount to the privacy constraint as in the second simpli-

fication. ■

Now, we derive the inductive relation for $\frac{M}{P} > 2$. Without loss of generality, let $\mathcal{P}_1 = \{1, \cdots, P\}$ and $\mathcal{P}_2 = \{P+1, \cdots, 2P\}$. Then, starting from (3.99), we write

$$
\begin{aligned}
ML =& H(A_1, A_{2:N}^{[\mathcal{P}_1]}, A_{2:N}^{[\mathcal{P}_2]}, \cdots, A_{2:N}^{[\mathcal{P}_\beta]} | \mathcal{Q}) && (3.116)\\
=& H(A_1, A_{2:N}^{[\mathcal{P}_1]} | \mathcal{Q}) + H(A_{2:N}^{[\mathcal{P}_2]} | A_1, A_{2:N}^{[\mathcal{P}_1]}, \mathcal{Q}) \\
&+ H(A_{2:N}^{[\mathcal{P}_3]}, \cdots, A_{2:N}^{[\mathcal{P}_\beta]} | A_1, A_{2:N}^{[\mathcal{P}_1]}, A_{2:N}^{[\mathcal{P}_2]}, \mathcal{Q}) && (3.117)\\
\leq& NH(A_1|\mathcal{Q}) + H(A_{2:N}^{[\mathcal{P}_2]} | A_1, A_{2:N}^{[\mathcal{P}_1]}, W_{1:P}, \mathcal{Q}) \\
&+ H(A_{2:N}^{[\mathcal{P}_3]}, \cdots, A_{2:N}^{[\mathcal{P}_\beta]} | A_1, A_{2:N}^{[\mathcal{P}_1]}, A_{2:N}^{[\mathcal{P}_2]}, W_{1:2P}, \mathcal{Q}) && (3.118)\\
\leq& NH(A_1|\mathcal{Q}) + H(A_{2:N}^{[\mathcal{P}_2]} | W_{1:P}, \mathcal{Q}) + H(A_{2:N}^{[\mathcal{P}_3]}, \cdots, A_{2:N}^{[\mathcal{P}_\beta]} | A_1, W_{1:2P}, \mathcal{Q}) && (3.119)\\
=& NH(A_1|\mathcal{Q}) + H(A_{2:N}^{[\mathcal{P}_2]} | W_{1:P}, \mathcal{Q}) - H(A_1 | W_{1:2P}, \mathcal{Q}) \\
&+ H(A_{1:N}^{[\mathcal{P}_3]}, \cdots, A_{1:N}^{[\mathcal{P}_\beta]} | W_{1:2P}, \mathcal{Q}) && (3.120)\\
=& NH(A_1|\mathcal{Q}) + H(A_{2:N}^{[\mathcal{P}_2]} | W_{1:P}, \mathcal{Q}) - H(A_1 | W_{1:2P}, \mathcal{Q}) + (M-2P)L && (3.121)\\
\leq& NH(A_1|\mathcal{Q}) + (N-1)[NH(A_1|\mathcal{Q}) - PL] \\
&- H(A_1 | W_{1:2P}, \mathcal{Q}) + (M-2P)L && (3.122)
\end{aligned}
$$

where (3.118) follows from the decodability of $W_{1:2P}$ given $(A_1, A_{2:N}^{[\mathcal{P}_1]}, A_{2:N}^{[\mathcal{P}_2]})$, the symmetry lemma and the independence bound; (3.119) follows from the fact that conditioning does not increase entropy. In (3.121), we note that subsets $(\mathcal{P}_3, \cdots, \mathcal{P}_\beta)$ include all messages $(W_1, \cdots, W_M)$ because every message appears in $\binom{M-1}{P-1}$ subsets. Hence, $H(A_{1:N}^{[\mathcal{P}_3]}, \cdots, A_{1:N}^{[\mathcal{P}_\beta]}|$

$W_{1:2P}, \mathcal{Q}) = (M-2P)L$ since $W_{2P+1:M}$ is decodable from $(A_{1:N}^{[\mathcal{P}_3]}, \cdots, A_{1:N}^{[\mathcal{P}_\beta]})$ after knowing $W_{1:2P}$. Finally, (3.122) follows from the interference conditioning lemma.

Consequently, (3.122) can be written as

$$N^2 H(A_1|\mathcal{Q}) \geq (N+1)PL + H(A_1|W_{1:2P}, \mathcal{Q}) \tag{3.123}$$

which is equivalent to

$$NH(A_1|\mathcal{Q}) \geq \left(1+\frac{1}{N}\right) PL + \frac{1}{N} H(A_1|W_{1:2P}, \mathcal{Q}) \tag{3.124}$$

Now, (3.124) constructs an inductive relation, since evaluating $NH(A_1|W_{1:2P}, \mathcal{Q})$ is the same as $NH(A_1|\mathcal{Q})$ with $(M-2P)$ messages, i.e., the problem of MPIR with M messages for fixed P is reduced to an MPIR problem with $(M-2P)$ messages for the same fixed P. We note that (3.124) generalizes the inductive relation in [12] for $P=1$.

We can write the induction hypothesis for MPIR with M messages as

$$NH(A_1|\mathcal{Q}) \geq PL\left[\sum_{i=0}^{\lfloor \frac{M}{P} \rfloor - 1} \frac{1}{N^i} + \left(\frac{M}{P} - \left\lfloor \frac{M}{P} \right\rfloor\right) \frac{1}{N^{\lfloor \frac{M}{P} \rfloor}}\right] \tag{3.125}$$

Next, we proceed with proving this relation for $M+1$ messages. From the induction hypothesis, we have

$$NH(A_1|W_{1:2P}, \mathcal{Q})$$

$$\geq PL\left[\sum_{i=0}^{\lfloor\frac{M-2P+1}{P}\rfloor-1}\frac{1}{N^i}+\left(\frac{M-2P+1}{P}-\left\lfloor\frac{M-2P+1}{P}\right\rfloor\right)\frac{1}{N^{\lfloor\frac{M-2P+1}{P}\rfloor}}\right] \tag{3.126}$$

$$=PL\left[\sum_{i=0}^{\lfloor\frac{M+1}{P}\rfloor-3}\frac{1}{N^i}+\left(\frac{M+1}{P}-\left\lfloor\frac{M+1}{P}\right\rfloor\right)\frac{1}{N^{\lfloor\frac{M+1}{P}\rfloor-2}}\right] \tag{3.127}$$

substituting this in (3.124),

$$NH(A_1|\mathcal{Q})$$

$$\geq\left(1+\frac{1}{N}\right)PL+\frac{PL}{N^2}\left[\sum_{i=0}^{\lfloor\frac{M+1}{P}\rfloor-3}\frac{1}{N^i}+\left(\frac{M+1}{P}-\left\lfloor\frac{M+1}{P}\right\rfloor\right)\frac{1}{N^{\lfloor\frac{M+1}{P}\rfloor-2}}\right] \tag{3.128}$$

$$=PL\left[\sum_{i=0}^{\lfloor\frac{M+1}{P}\rfloor-1}\frac{1}{N^i}+\left(\frac{M+1}{P}-\left\lfloor\frac{M+1}{P}\right\rfloor\right)\frac{1}{N^{\lfloor\frac{M+1}{P}\rfloor}}\right] \tag{3.129}$$

which concludes the induction argument.

Consequently, the upper bound for the MPIR problem can be obtained as,

$$\sum_{i=1}^{P}R_i=\frac{PL}{\sum_{n=1}^{N}H\left(A_n^{[\mathcal{P}]}\right)} \tag{3.130}$$

$$\leq\frac{PL}{NH(A_1|\mathcal{Q})} \tag{3.131}$$

$$=\frac{1}{\sum_{i=0}^{\lfloor\frac{M}{P}\rfloor-1}\frac{1}{N^i}+\left(\frac{M}{P}-\left\lfloor\frac{M}{P}\right\rfloor\right)\frac{1}{N^{\lfloor\frac{M}{P}\rfloor}}} \tag{3.132}$$

$$=\left(\frac{1-(\frac{1}{N})^{\lfloor\frac{M}{P}\rfloor}}{1-\frac{1}{N}}+\left(\frac{M}{P}-\left\lfloor\frac{M}{P}\right\rfloor\right)\frac{1}{N^{\lfloor\frac{M}{P}\rfloor}}\right)^{-1} \tag{3.133}$$

where (3.132) follows from (3.129); and (3.133) follows from evaluating the sum in (3.132).

3.7 Conclusions

In this chapter, we introduced the multi-message private information retrieval (MPIR) problem from an information-theoretic perspective. The problem generalizes the PIR problem in [12] which retrieves a single message privately. We determined the exact sum capacity for this problem when the number of desired messages is at least half of the number of total stored messages to be $C_s^P = \frac{1}{1+\frac{M-P}{PN}}$. We showed that joint retrieval of the desired messages strictly outperforms repeating the single-message capacity achieving scheme for each message. Furthermore, we showed that if the total number of messages is an integer multiple of the number of desired messages, then the sum capacity is $C_s^P = \frac{1-\frac{1}{N}}{1-(\frac{1}{N})^{M/P}}$, which resembles the single-message PIR capacity expression when the number of messages is $\frac{M}{P}$. For the remaining cases, we derived lower and upper bounds. We observed numerically that the gap between the lower and upper bounds decreases monotonically in N, and the worst case gap is 0.0082 which occurs for the case $N = 2$ when $M = 5$, $P = 2$.

The MPIR problem can be extended in several interesting directions. First, we recall from earlier remarks in the chapter that the sum capacity for $M/P \notin \mathbb{N}$ is still an open problem, in addition to characterizing the optimal capacity region. Second, the MDS-coded MPIR as an extension of [117] is an interesting open problem, as the contents of the databases are themselves coded via an MDS code in [117]. This is

a challenging problem, in particular if $P \geq \frac{M}{2}$, because our achievable scheme here uses a $P \times M$ MDS code; it would be interesting to see how the storage MDS code and the retrieval MDS code would interact. Similar difficulties would exist in the MPIR problem with colluding databases (extending [14]), robust MPIR problem (extending [14]), and MPIR problem with Byzantine databases (extending [120]), as all these problems adopt some version of MDS coding for retrieval purposes. Furthermore, one can examine whether multiround MPIR enhances the MPIR retrieval rate or not (extending the case of single-message retrieval in [19]), and study the effects of limited message size on MPIR (extending [18]). Our converse techniques may be generalized to be applicable to these scenarios. Some progress in these MPIR problems has been made recently in [25].

CHAPTER 4

Private Information Retrieval from Byzantine and Colluding Databases

4.1 Introduction

In this chapter, we consider the problem of single-round PIR from N replicated databases, where B databases are outdated (unsynchronized), or even worse, adversarial (Byzantine), and therefore, can return incorrect answers. In the PIR problem with Byzantine databases (BPIR), a user wishes to retrieve a specific message from a set of M messages with zero-error, irrespective of the actions performed by the Byzantine databases. We consider the T-privacy constraint in this chapter, where any T databases can collude, and exchange the queries submitted by the user. We derive the information-theoretic capacity of this problem, which is the maximum number of *correct symbols* that can be retrieved privately (under the T-privacy constraint) for every symbol of the downloaded data. We determine the exact BPIR capacity to be $C = \frac{N-2B}{N} \cdot \frac{1-\frac{T}{N-2B}}{1-(\frac{T}{N-2B})^M}$, if $2B + T < N$. This capacity expression shows that the effect of Byzantine databases on the retrieval rate is equivalent to removing $2B$ databases from the system, with a penalty factor of $\frac{N-2B}{N}$, which signi-

fies that even though the number of databases needed for PIR is effectively $N-2B$, the user still needs to access the entire N databases. The result shows that for the unsynchronized PIR problem, if the user does not have any knowledge about the fraction of the messages that are mis-synchronized, the single-round capacity is the same as the BPIR capacity. Our achievable scheme extends the optimal achievable scheme for the robust PIR (RPIR) problem to correct the *errors* introduced by the Byzantine databases as opposed to *erasures* in the RPIR problem. Our converse proof uses the idea of the cut-set bound in the network coding problem against adversarial nodes.

4.2 Problem Formulation

Consider a single-round PIR setting with N replicated databases storing M messages (or files). The messages $\mathcal{W} = \{W_1, \cdots, W_M\}$ are independent and uniformly distributed over a large enough finite field $\mathbb{F}_q$. Each message $W_i \in \mathbb{F}_q^L$ is a vector of length L (q-ary symbols),

$$H(W_i) = L, \quad i = 1, \cdots, M \tag{4.1}$$

$$H(\mathcal{W}) = H(W_1, \cdots, W_M) = ML \tag{4.2}$$

Each database stores a copy from the complete set of messages $\mathcal{W}$, i.e., this distributed storage system applies an $(N, 1)$ repetition code [117]. Denote the contents of the nth database by Ω_n. Ideally, $\Omega_n = \mathcal{W}$ for all $n \in \{1, \cdots, N\}$.

In the PIR problem, a user wishes to retrieve a message $W_i \in \mathcal{W}$ without

revealing any information about the message index i. The user submits a single-round query $Q_n^{[i]}$ to the nth database. The user does not know the stored messages in advance, therefore, the message set $\mathcal{W}$ and the queries are statistically independent,

$$I\left(\mathcal{W}; Q_{1:N}^{[i]}\right) = I\left(W_1, \cdots, W_M; Q_{1:N}^{[i]}\right) = 0 \tag{4.3}$$

where $Q_{1:N}^{[i]} = \{Q_1^{[i]}, Q_2^{[i]}, \cdots, Q_N^{[i]}\}$ is the set of all queries to the N databases for message i.

Ideally, the classical PIR formulation assumes that all databases store the correct database contents (i.e., up-to-date contents), and respond truthfully with the correct answering strings $A_{1:N}^{[i]} = \{A_1^{[i]}, \cdots, A_N^{[i]}\}$. In the BPIR setting, on the other hand, there exists a set $\mathcal{B}$ of databases, that is unknown to the user, such that $|\mathcal{B}| = B$, which are called Byzantine databases. These databases can respond arbitrarily to the user by introducing errors to the answer strings $A_{\mathcal{B}}^{[i]} = \{A_j^{[i]} : j \in \mathcal{B}\}$, i.e.,

$$H\left(A_n^{[i]} | Q_n^{[i]}, \mathcal{W}\right) > 0, \quad n \in \mathcal{B},\ |\mathcal{B}| = B \tag{4.4}$$

We assume that these Byzantine databases can coordinate upon submitting the answers. In this chapter, we do not assume a specific pattern to the errors. The remaining set of databases $\bar{\mathcal{B}} = \{1, \cdots, N\} \setminus \mathcal{B}$ respond truthfully to the user, i.e., the answer strings of $\bar{\mathcal{B}}$ are a deterministic function of the queries and the correct

contents of the databases $\mathcal{W}$,

$$H\left(A_n^{[i]}|Q_n^{[i]},\mathcal{W}\right)=0,\quad n\in\bar{\mathcal{B}},\ |\bar{\mathcal{B}}|=N-B \tag{4.5}$$

We consider a T-privacy constraint as in the TPIR problem in [14], where any T databases can communicate and exchange the queries submitted by the user. To ensure the T-privacy constraint, the queries to any set $\mathcal{T}\subset\{1,\cdots,N\}$ of databases, such that $|\mathcal{T}|=T$, need to be statistically independent of the desired message index i, i.e.,

$$I\left(i;Q_{\mathcal{T}}^{[i]}\right)=0,\quad \text{for all}\ \mathcal{T}\subset\{1,\cdots,N\},\ |\mathcal{T}|=T \tag{4.6}$$

where $Q_{\mathcal{T}}^{[i]}$ are the queries submitted to the set $\mathcal{T}$ of databases.

We remark here to differentiate the actions of *colluding* between the databases which is done to figure out the desired message, and *coordination* between the Byzantine databases which is done to introduce errors in the answer strings. In addition to the difference in their purposes, these two actions differ in the manner they are performed: colluding between any T databases occurs upon receiving the queries from the user, while coordination between the B Byzantine databases occurs upon submitting the answers to the user. We do not assume any specific relation between the T colluding databases and the B Byzantine databases. This is a more general formulation of the problem; the user in this case has the knowledge that there are B Byzantine databases and T colluding databases, but does not know anything fur-

ther. In general, these two sets may be identical, one may be a subset of the other, they may be disjoint, or they may have a non-trivial intersection[1].

The user should be able to reconstruct the desired message W_i, no matter what the Byzantine databases do, i.e., if there exists a set of databases $\bar{\mathcal{B}}$, that is unknown to the user, such that (4.5) holds, then the reliability constraint is given by,

$$H(W_i|A_{1:N}^{[i]}, Q_{1:N}^{[i]}) = 0 \tag{4.7}$$

We define the *resilient* PIR rate R for the BPIR problem as the ratio between the message size L and the total download cost under the reliability constraint in (4.7) for any possible action of the Byzantine databases, and the T-privacy constraint in (4.6), i.e.,

$$R = \frac{L}{\sum_{n=1}^{N} H(A_n^{[i]})} \tag{4.8}$$

The capacity of BPIR is $C = \sup R$ over all possible single-round retrieval schemes.

In this chapter, we follow the information-theoretic assumptions of large enough message size, large enough field size, and ignore the upload cost as in

[1]For instance, they may be disjoint if the intentions of these databases are different, e.g., if the T colluding databases are only curious to learn the interests of the user without disrupting the retrieval process, while the B Byzantine databases do not care about the identity of the desired message but just want to block the retrieval process itself. An example where Byzantine behavior may not require collusion, or even communication, is when B databases become outdated (unsynchronized) with the same outdated content. This happens without a communication between the databases, but results in errors at the user's side as if these B databases are coordinating, as they have the same wrong content. This discussion clarifies that collusion (which requires communication between databases) and Byzantine behavior (which may or may not require communication or coordination between databases) can be completely different.

[8, 14, 117]. A formal treatment of the capacity under message size constraints can be found in [18]. The BPIR with colluding databases reduces to the TPIR problem in [14] if $B = 0$.

Some scenarios that fit our formulation include:

- *Unsynchronized setting [7]:* In this case, there exists a set $\mathcal{B}$ of databases, such that $|\mathcal{B}| = B$, in which they store different versions of the database contents (see Fig. 4.1), i.e.,

$$\Omega_n \neq \mathcal{W}, \quad n \in \mathcal{B},\ |\mathcal{B}| = B \tag{4.9}$$

  Note that unlike [7], we assume that the user has no knowledge about the fraction of the messages that are mis-synchronized. Hence, our achievable schemes must be resilient against the worst-case that the entirety of the database is mis-synchronized. Furthermore, the scheme in [7] is a two-round scheme, hence we cannot compare our rates with the rates in [7]; we consider only single-round schemes here.

- *Adversarial attacks [55–57]:* In this case, the databases in $\mathcal{B}$ intend to preclude the retrieval process at the user by introducing a carefully-designed error sequence (see Fig. 4.2). This can be done by altering the contents of the databases to an erroneous version as in the unsynchronized setting; or by altering the answering strings themselves, i.e., the nth database returns the

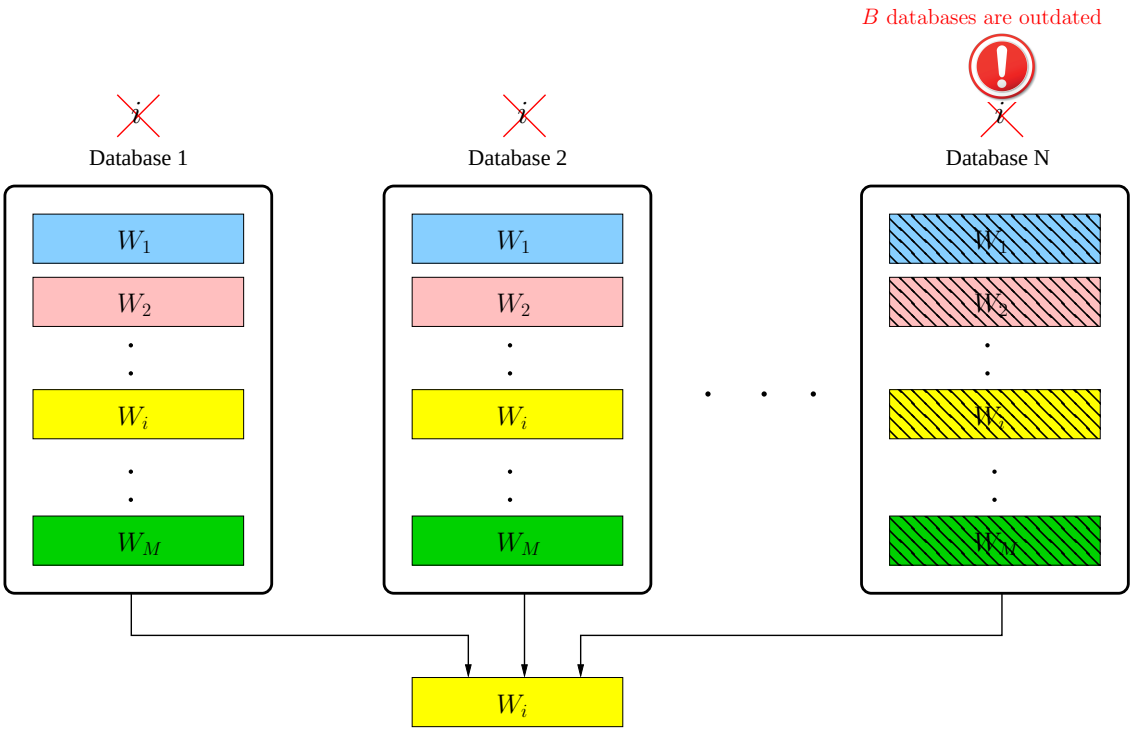


Figure 4.1: PIR from unsynchronized databases.

answer string $\tilde{A}_n^{[i]}$ such that,

$$\tilde{A}_n^{[i]} \neq A_n^{[i]}, \quad n \in \mathcal{B}, \ |\mathcal{B}| = B \tag{4.10}$$

or by doing both.

4.3 Main Result and Discussions

The main result of this chapter is to characterize the capacity of the BPIR problem under T-privacy constraint, where B databases are adversarial (Byzantine) and can return malicious answers, and at the same time the privacy should be kept against any T colluding databases.

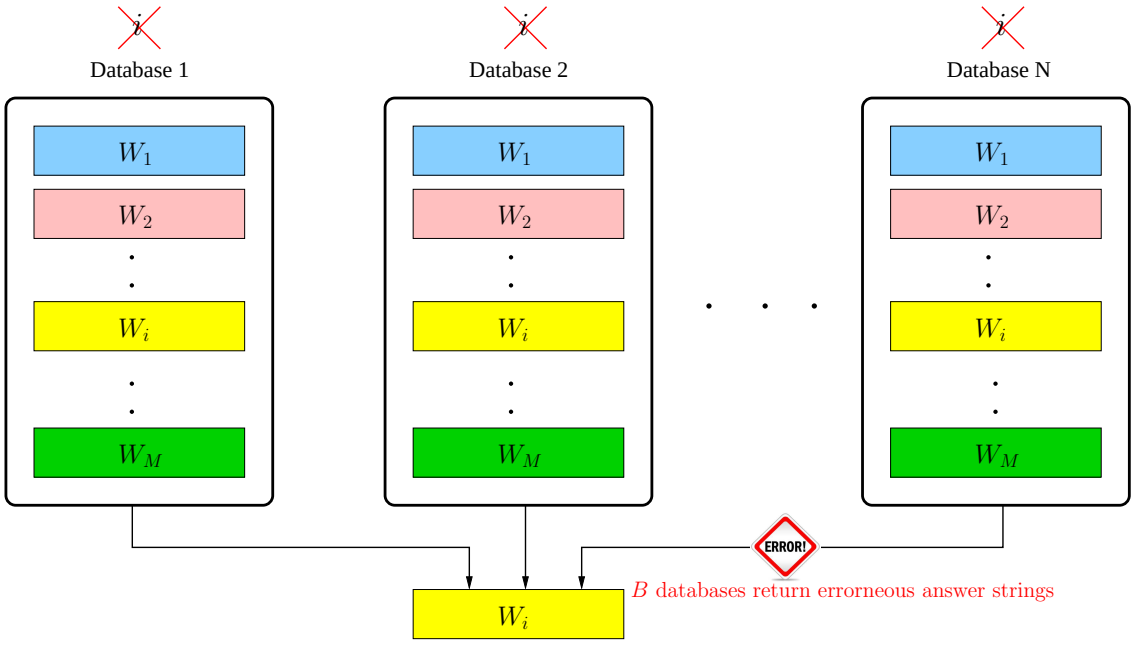


Figure 4.2: PIR under adversarial attacks.

Theorem 4.1 *For the single-round BPIR problem with B Byzantine databases, and T colluding databases, such that $2B + T < N$, the capacity is given by,*

$$C = \frac{N-2B}{N} \cdot \frac{1-\frac{T}{N-2B}}{1-\left(\frac{T}{N-2B}\right)^M} \tag{4.11}$$

$$= \frac{N-2B}{N} \cdot \left(1 + \frac{T}{N-2B} + \frac{T^2}{(N-2B)^2} + \cdots + \frac{T^{M-1}}{(N-2B)^{M-1}}\right)^{-1} \tag{4.12}$$

On the other hand, if $2B+1 \leq N \leq 2B+T$, then the user is forced to download the entire database from at least from $(2B+1)$ different databases, hence $C = \frac{1}{(2B+1)M}$, which is the trivial rate in the BPIR problem. Otherwise, the problem is infeasible and $C = 0$.

The achievability proof for Theorem 4.1 is given in Section 4.4, and the converse proof is given in Section 4.5. We have a few remarks.

Remark 4.1 *The BPIR capacity in (4.11) is the same as the capacity of PIR with T colluding databases if the number of databases is $N-2B$ with a penalty factor of $\frac{N-2B}{N}$. This means that the harm introduced by the B Byzantine databases is equivalent to removing a part from the storage system of size $2B$, but the user still needs to download from all N databases, as it does not know which $N-2B$ databases are honest. This results in the penalty term $\frac{N-2B}{N}$. If $B=0$, the expression in (4.11) reduces to*

$$C_{colluded} = \frac{1-\frac{T}{N}}{1-\left(\frac{T}{N}\right)^M} \tag{4.13}$$

which is the capacity expression in [14] as expected. Fig. 4.3 shows the severe effect of the Byzantine databases on the retrieval rate for fixed $T=2$ and $M=3$ as a function of N.

Remark 4.2 *Comparing the BPIR capacity in Theorem 4.1 with the robust capacity C_{robust} in [14], where U databases are merely unresponsive,*

$$C_{robust} = \frac{1-\frac{T}{N-U}}{1-\left(\frac{T}{N-U}\right)^M} \tag{4.14}$$

we note that the number of redundant databases, which are needed to correct the errors *introduced by the Byzantine databases, is* twice *the number of redundant*

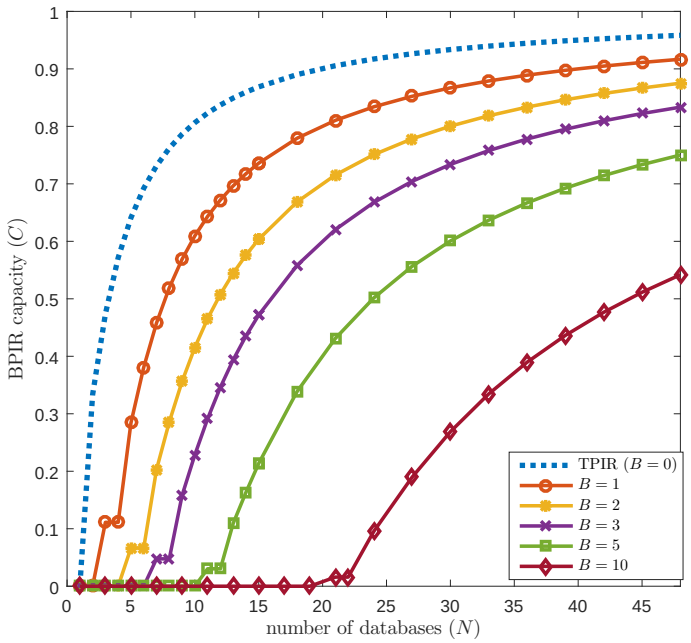


Figure 4.3: The effect of Byzantine databases on the BPIR capacity as a function of N for fixed $T = 2$, $M = 3$.

databases needed to correct the erasures *introduced in the case of unresponsive databases. We also note that the penalty factor is missing in the RPIR problem, since in the RPIR problem, the user does not get the chance to download from the unresponsive databases, in contrast to the BPIR problem, in which the user downloads answer strings from all databases. This is due to the fact that the user cannot identify the Byzantine databases before decoding the entire answer strings in the BPIR setting, while in the RPIR setting, the user identifies the unresponsive databases as they simply do not return answer strings.*

Remark 4.3 *The trivial rate for the BPIR problem is* $\frac{1}{(2B+1)M}$*, which is much less than the trivial rate without the Byzantine databases,* $\frac{1}{M}$*. The reason for this is that the user cannot download the entire database only once in BPIR, but it must down-*

load $(2B+1)$ *different copies of the database in order to decode the desired message via majority decoding. If* $N < 2B+1$*, the capacity is* $C = 0$*, as the Byzantine databases can always confuse the user to decode the desired message incorrectly.*

Remark 4.4 *When the number of messages is large, i.e., as* $M \to \infty$*, the BPIR capacity* $C \to (\frac{N-2B}{N})(1 - \frac{T}{N-2B}) = 1 - \frac{2B+T}{N}$*, i.e., for large enough number of messages, the capacity expression acts as if there are no Byzantine databases and* $2B+T$ *databases are colluding.*

Remark 4.5 *If* T *and* B *are fixed and do not scale with* N*, i.e.,* $T = B = o(N)$*, then the capacity is a strictly increasing function in* N *and* $C \to 1$ *as* $N \to \infty$*. If the number of the Byzantine databases scales with* N*, i.e.,* $B = \gamma N$*, where* $\gamma \in \left[0, \frac{1}{2}(1 - \frac{T}{N})\right)$*, then* $C \to 1 - 2\gamma$ *as* $N \to \infty$*. If* $2\gamma + \frac{1}{N} \leq 1 \leq 2\gamma + \frac{T}{N}$*, then the only possible rate is the trivial rate* $\frac{1}{(2B+1)M}$*. As* $N \to \infty$*, then* $\gamma \to \frac{1}{2}$*, and* $C \to 0$*. This entails that the asymptotic behaviour of the BPIR capacity is a linear function with a slope of* -2 *as in Fig. 4.4, i.e., the asymptotic rate as* $N \to \infty$ *is decreased by* twice *the ratio of the Byzantine databases. A similar behaviour is observed for secure distributed storage systems against Byzantine attacks in [61]. The problem is infeasible if* $\gamma > \frac{1}{2}$*, i.e.,* $C = 0$*. This feasibility result conforms with the best result of a* uniquely decodable *BPIR scheme in [57] which needs* $B < \frac{N}{2}$*.*

Remark 4.6 *Surprisingly, our retrieval scheme in Section 4.4 is a linear scheme in contrast to the network coding problem in [60] that states that linear coding schemes are not sufficient. We note that although the retrieval process is itself linear, the*

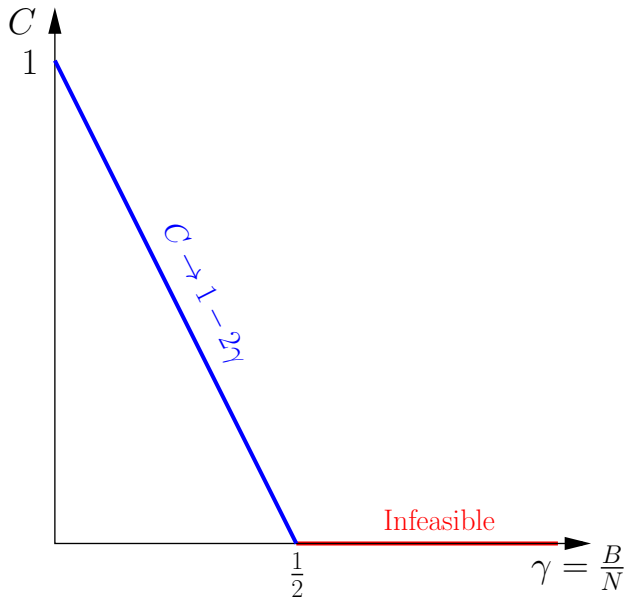


Figure 4.4: The asymptotic BPIR capacity C as $N \to \infty$ as a function of $\gamma = \frac{B}{N}$.

decoding process employs a successive interference cancellation decoder, which is non-linear.

Remark 4.7 *The capacity expression in Theorem 4.1 is also the capacity result for the unsychronized PIR problem [7]. This occurs under the restriction to single-round schemes and the assumption that the user only knows that there exist B databases that are unsynchronized, but does not know the fraction of messages that are mis-synchronized. The achievability scheme in Section 4.4 is a valid achievable scheme for the unsynchronized PIR problem, since the adversary in the Byzantine setting is stronger. For the converse proof, we restricted the actions of the adversarial databases to changing the contents of the stored messages, i.e., altering Ω_n from $\mathcal{W}$ to $\tilde{\mathcal{W}}$, which is the same setting as the unsynchronized PIR with no restriction on the fraction of messages that can be mis-synchronized.*

4.4 Achievability Proof

In this section, we present an achievable scheme that is resilient to the errors introduced by the Byzantine databases. The achievable scheme does not assume any specific error pattern. Hence, our achievable scheme enables *correct decoding* of any desired message if any B databases become outdated, or even worse, intentionally commit an adversarial attack to confuse the user. The achievable scheme generalizes the RPIR scheme presented in [14]. Our scheme has two new ingredients, namely, correcting errors in the side information using punctured MDS codes, and correcting errors in the desired message by an outer layer of MDS code. Error correction in both cases is performed via a nearest-codeword decoder.

4.4.1 Preliminaries

We start by presenting some preliminary results that will be needed. The following lemma states that if an MDS code is punctured by a puncture pattern whose length is smaller than the minimum distance of the original MDS code, then it remains an MDS code [121].

Lemma 4.1 (MDS code puncturing [121]) *If $\mathcal{C}$ is an (n, k) MDS code, then by puncturing the code by a sequence of length z, i.e., deleting a sequence of size z from output codewords of $\mathcal{C}$, such that $z < n - k$, the resulting punctured code $\mathcal{C}_z$ is an $(n - z, k)$ MDS code.*

The second lemma is regarding the statistical effect of operating on a random matrix by a deterministic full-rank matrix. The proof of this lemma can be found in [14].

Lemma 4.2 (Statistical effect of full-rank matrices [14]) *Let* $\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_M \in \mathbb{F}_q^{\alpha \times \alpha}$ *be* M *random matrices, drawn independently and uniformly from all* $\alpha \times \alpha$ *full-rank matrices over* $\mathbb{F}_q$*. Let* $\mathbf{G}_1, \mathbf{G}_2, \cdots, \mathbf{G}_M \in \mathbb{F}_q^{\beta \times \beta}$ *be* M *invertible square matrices of dimension* $\beta \times \beta$ *over* $\mathbb{F}_q$*. Let* $\mathcal{I}_1, \cdots, \mathcal{I}_M \in \mathbb{N}^\beta$ *be* M *index vectors, each containing* β *distinct indices from* $\{1, \cdots, \alpha\}$*, then*

$$\{\mathbf{G}_1\mathbf{S}_1(\mathcal{I}_1, :), \cdots, \mathbf{G}_M\mathbf{S}_M(\mathcal{I}_M, :)\} \sim \{(\mathbf{S}_1([1:\beta], :), \cdots, \mathbf{S}_M([1:\beta], :)\} \tag{4.15}$$

where $\sim$ *denotes statistical equivalence,* $\mathbf{S}_i(\mathcal{I}_i, :)$, $\mathbf{S}_i([1:\beta], :)$ *denote* $\beta \times \alpha$ *matrices with rows indexed by* $\mathcal{I}_i$ *and* $\{1, 2, \cdots, \beta\}$*, respectively.*

The next lemma summarizes the code capabilities of handling errors and erasures for linear block codes [122, Theorem 1.7].

Lemma 4.3 (Code capabilities [122]) *Let* $\mathcal{C}$ *be an* $[n, k, d]$ *linear block code over* $\mathbb{F}_q$*. Let* ρ *be the number of erasures introduced by the channel. Let* $\tau \in \mathbb{N}$*, such that* $2\tau + \rho \leq d - 1$*, then there exists a nearest-codeword decoder that recovers all errors and erasures if the number or errors (excluding erasures) is* τ *or less.*

Lemma 4.3 implies that in the case of no erasures, the maximum number of errors $\tau \leq \left\lfloor \frac{d-1}{2} \right\rfloor$.

4.4.2 Motivating Example: $M = 2$ Messages, $N = 5$, $T = 2$, $B = 1$ Databases

Assume without loss of generality that W_1 is the desired message. Let a_i and b_i be the ith symbol mixture of messages W_1 and W_2, respectively. The specific construction of these mixtures will be presented shortly. We begin the retrieval process by downloading $T^{M-1} = 2$ symbols from W_1, which are a_1, a_2 as in [14]. By *message symmetry*, we download 2 symbols from W_2, which are b_1, b_2. By *database symmetry*, we download 2 symbols from W_1 and 2 symbols from W_2 from all other databases.

Now, we want to generate the maximum number of side information equations in order to maximize the retrieval rate. From Lemma 4.3, we see that the number of errors that can be corrected increases with d. We know that MDS codes meet the Singleton bound [62] with equality, hence encoding both desired and undesired messages by MDS codes is desirable. In addition, Lemma 4.3 implies a *doubling effect*, which suggests that in order to correct the errors introduced by the Byzantine database, we should effectively consider $N-2B = 3$ honest databases. Consequently, considering any 3 databases, the number of undesired symbols is 6. We note that any $T = 2$ of them can collude, therefore, we are left with 2 undesired symbols that can be used to generate side information among the 2 colluding databases. Hence, each database should get 1 side information equation $b_{[11:15]}$. These side-information symbols can be added to new desired symbols $a_{[11:15]}$. The complete query structure is shown in Table 4.1.

Table 4.1: The query table for the case $M = 2$, $N = 5$, $T = 2$, $B = 1$.

DB 1	DB 2	DB 3	DB 4	DB 5
a_1	a_3	a_5	a_7	a_9
a_2	a_4	a_6	a_8	a_{10}
b_1	b_3	b_5	b_7	b_9
b_2	b_4	b_6	b_8	b_{10}
$a_{11} + b_{11}$	$a_{12} + b_{12}$	$a_{13} + b_{13}$	$a_{14} + b_{14}$	$a_{15} + b_{15}$

Now, we specialize the query structure in Table 4.1, and identify the specific construction of the mixtures $a_{[1:15]}$ and $b_{[1:15]}$. For the desired message W_1, considering any $N - 2B = 3$ honest databases, we see 9 distinct symbols. Therefore, the length of W_1 is $L = 9$, and we use $\mathbf{S}_1$, which is a 9×9 random mixing matrix picked uniformly from the full-rank matrices over $\mathbb{F}_q^{9\times 9}$. These 9 mixed symbols are further mapped to $a_{[1:15]}$ by a $(15, 9)$ MDS code generator matrix $\mathbf{MDS}_{15\times 9}$, therefore,

$$a_{[1:15]} = \mathbf{MDS}_{15\times 9}\mathbf{S}_1 W_1 \tag{4.16}$$

For the undesired message W_2, considering again any $N-2B = 3$ honest databases, we have 6 individual symbols from W_2 in round 1. We should be able to reconstruct the side information equations $b_{[11:15]}$ in round 2 from any 6 individual symbols, hence we get 6 random symbols from W_2. This can be done by considering the first 6 rows of the random mixing matrix $\mathbf{S}_2 \in \mathbb{F}_q^{9\times 9}$. These randomly mixed symbols are further mapped to $b_{[1:15]}$ via and MDS code with generator matrix $\mathbf{MDS}_{15\times 6}$, i.e.,

$$b_{[1:15]} = \mathbf{MDS}_{15\times 6}\mathbf{S}_2([1:6], :)W_2 \tag{4.17}$$

To see the decodability: the worst-case scenario is that the Byzantine database

commits errors in all the symbols returned to the user. This means that the database commits 2 errors in the individual symbols from W_1, 2 errors in the individual symbols from W_2, and 1 extra error in the sum of $a + b$.

Consider the codeword $b_{[1:10]}$: this codeword belongs to $(15, 6)$ MDS code with a sequence of length $z = 5$ removed. Hence, this codeword belongs to $(10, 6)$ punctured MDS code. Since $z = 5 < 15 - 6 = 9$, the $(10, 6)$ punctured MDS code is still an MDS code. Denote the minimum distance of the $(10, 6)$ punctured MDS code that results in $b_{[1:10]}$ by d_p^b. Then, $d_p^b = 10 - 6 + 1 = 5$. Consequently, from Lemma 4.3, the $(10, 6)$ punctured MDS code can tolerate errors up to τ_b, such that

$$\tau_b \leq \left\lfloor \frac{d_p^b - 1}{2} \right\rfloor = 2 \tag{4.18}$$

Therefore, this code can correct all errors that can be introduced to the individual undesired symbols $b_{[1:10]}$. Let $b^*_{[1:10]}$ be the correct codeword of $b_{[1:10]}$. Choose any 6 symbols from $b^*_{[1:10]}$. Now, since $\mathbf{MDS}_{15\times 6}$ matrix has the property that any 6×6 matrix is an invertible matrix, then from any 6 symbols from $b^*_{[1:10]}$, the *correct side information* equations $b^*_{[11:15]}$ are determined and canceled from the sums of a and b in round 2.

For the desired message W_1: after removing the interference from W_2, we are left with $\tilde{a}_{[1:15]}$. Note that this is not exactly $a_{[1:15]}$, because we canceled the correct side information and not $b_{[1:15]}$. However, the total errors in $\tilde{a}_{[1:15]}$ still is upper bounded by 3, since $\tilde{a}_{[1:15]}$ can differ from $a_{[1:15]}$ only in the positions that correspond to Byzantine databases. The desired message W_1 is coded via $(15, 9)$

MDS code. Then, the minimum distance for this code is $d^a = 15 - 9 + 1 = 7$. Consequently, this code can tolerate errors up to τ_a, such that

$$\tau_a \leq \left\lfloor \frac{d^a - 1}{2} \right\rfloor = 3 \tag{4.19}$$

Hence, all the errors in $\tilde{a}_{[1:15]}$ can be corrected, and we can obtain true $a^*_{[1:15]}$. Consider the first 9 symbols from $a^*_{[1:15]}$, without loss of generality, then

$$W_1 = (\mathbf{MDS}_{15\times 9}([1:9],:)\mathbf{S}_1)^{-1} a^*_{[1:9]} \tag{4.20}$$

since $\mathbf{MDS}_{15\times 9}([1:9],:)\mathbf{S}_1$ is a 9×9 invertible matrix.

Therefore, despite Byzantine behaviour of $B = 1$ database, we decode the desired message correctly. In addition, our achievable scheme can identify the Byzantine database as does the scheme in [7] by comparing $a^*_{[1:10]}$ with $a_{[1:10]}$, and $b^*_{[1:10]}$ with $b_{[1:10]}$ and see which database has introduced errors.

To see the privacy: we note that from any $T = 2$ databases, our achievable scheme collects 6 symbols from $a_{[1:15]}$ and 6 symbols from $b_{[1:15]}$ indexed by $\mathcal{I}$ such that $|\mathcal{I}| = 6$. For the undesired message, we collect $b_{\mathcal{I}}$,

$$b_{\mathcal{I}} = \mathbf{MDS}_{15\times 6}(\mathcal{I},:)\mathbf{S}_2([1:6],:)W_2 \tag{4.21}$$

$$\sim \mathbf{S}_2([1:6],:)W_2 \tag{4.22}$$

where (4.22) follows from Lemma 4.2 as any 6×6 matrix in $\mathbf{MDS}_{15\times 6}$ matrix is

full-rank. Therefore, the symbols $b_{\mathcal{I}}$ are independent and uniformly distributed. For $a_{\mathcal{I}}$, we have

$$a_{\mathcal{I}} = \mathbf{MDS}_{15\times 9}(\mathcal{I},:)\mathbf{S}_1 W_1 \tag{4.23}$$

$$= \Psi_{6\times 9} W_1 \tag{4.24}$$

where $\Psi = \mathbf{MDS}_{15\times 9}(\mathcal{I},:)\mathbf{S}_1$ is a full row-rank matrix as any 6 rows in $\mathbf{MDS}_{15\times 9}$ are linearly independent. Consequently, the symbols $a_{\mathcal{I}}$ are also independent and uniformly distributed, and $a_{\mathcal{I}} \sim b_{\mathcal{I}}$ for every 2 databases, where $\sim$ means that the involved random vectors are statistically identical. Thus, the proposed scheme is 2-private; that is, despite colluding behaviour of $T = 2$ databases, we have privacy.

Finally, the achievable resilient retrieval rate is $R = \frac{9}{25} = \frac{N-2B}{N} \cdot \frac{1-\frac{T}{N-2B}}{1-\left(\frac{T}{N-2B}\right)^M} = C$. In comparison, the trivial rate for this system is $\frac{1}{(2B+1)M} = \frac{1}{6}$, as the user must download the entire database from 3 different databases for correct decoding.

4.4.3 General Achievable Scheme

The general achievable scheme is performed in M rounds. The ith round includes all the $\binom{M}{i}$ combinations of the sums of any i messages. In our construction[2], we use $L = (N-2B)^M$. The construction resembles the optimal scheme for RPIR in [14]. The new key ingredient in our achievable scheme is the decoding procedure, which includes correcting the undesired symbols by punctured MDS codes, successive in-

[2]We note that we do not claim that $L = (N-2B)^M$ is the minimum message length needed to achieve the capacity. The reason we choose this specific L is that it enables us to realize our achievable scheme for general N, B, T, M. The problem of obtaining the minimum capacity-achieving L is an interesting open problem.

terference cancellation to cancel the interfering messages, and correcting the errors in the desired message by an outer layer MDS code.

4.4.3.1 General Description for the Scheme

1. *Initialization:* The scheme starts with downloading T^{M-1} mixed symbols from the desired message from the first database. The specific construction of the mixture will be specified shortly. The scheme sets the round index $i = 1$.

2. *Message symmetry:* To satisfy the privacy constraint, the user downloads the same number of mixed symbols from the undesired messages with all the possible combinations, i.e., in the ith round, the user downloads $\binom{M-1}{i}(N - 2B - T)^{i-1}T^{M-i}$ mixed symbols from the remaining $M - 1$ messages. The specific construction of the undesired mixture will be specified shortly.

3. *Database symmetry:* The user repeats the same steps at all the databases. Specifically, the user downloads $\binom{M-1}{i-1}(N - 2B - T)^{i-1}T^{M-i}$ equations in the form of a desired message mixture symbol and $i - 1$ mixed symbols from the undesired messages, and $\binom{M-1}{i}(N - 2B - T)^{i-1}T^{M-i}$ mixed symbols from the undesired messages only, from each database.

4. *Exploiting side information:* The specific construction of the undesired mixtures should be done such that in the $(i + 1)$th round, the user should be able to generate $\frac{N-2B-T}{T}$ side information equations for each undesired symbol in the ith round. This fraction is a consequence of considering $\tilde{N} = N - 2B$ honest databases only, and dividing the undesired symbols from the $\tilde{N} - T$

databases among the T colluding databases. The side information generated is added to a new mixed symbol from the desired message.

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

4.4.3.2 Specific Construction of the Symbol Mixtures

Let $W_m \in \mathbb{F}_q^{(N-2B)^M}$, $m \in \{1, \cdots, M\}$ be the message vectors, and $\mathbf{S}_m$, $m \in \{1, \cdots, M\}$ be random mixing matrices picked independently and uniformly from the full-rank matrices in $\mathbb{F}_q^{(N-2B)^M \times (N-2B)^M}$. From the general description of the scheme, we note that at the ith round, the user downloads all possible combinations of the sums of any i messages. In the following specific construction, we enumerate all the sets that contain a symbol from the desired message and assign them labels $\mathcal{L}_1, \cdots, \mathcal{L}_\delta$. For each undesired message, we further enumerate also all the sets that contain symbols from this undesired message and do not include any desired symbols and assign them labels $\mathcal{K}_1, \cdots, \mathcal{K}_\Delta$. These sets construct the undesired symbol mixtures and the corresponding side information.

For the desired message: Assume that the desired message is W_ℓ. Let δ be the number of the distinct subsets of $\{1, \cdots, M\}$ that contain ℓ, then $\delta = 2^{M-1}$. Let $\mathcal{L}_i$, $i \in \{1, \cdots, \delta\}$ be the ith subset that contains ℓ. Assume without loss of generality, that these sets are arranged in ascending order in the sizes of the sets $|\mathcal{L}_i|$. According to this order, we note that $\mathcal{L}_1 = \{\ell\}$ and belongs to round 1. Round 2 contains sets $\mathcal{L}_2, \cdots, \mathcal{L}_{\binom{M-1}{1}+1}$, and so on. Let $X^{[\ell]} \in \mathbb{F}_q^{N(N-2B)^M}$ be the vector of mixtures that should be obtained from the desired message W_ℓ. Divide $X^{[\ell]}$ into δ

partitions denoted by $x_{\mathcal{L}_i}^{[\ell]}$, each corresponds to a distinct set $\mathcal{L}_i$. Now, encode the desired message by a $\left(N(N-2B)^{M-1}, (N-2B)^M\right)$ MDS code as,

$$X^{[\ell]} = \begin{bmatrix} x_{\mathcal{L}_1}^{[\ell]} \\ x_{\mathcal{L}_2}^{[\ell]} \\ \vdots \\ x_{\mathcal{L}_\delta}^{[\ell]} \end{bmatrix} = \mathbf{MDS}_{N(N-2B)^{M-1}\times(N-2B)^M}\mathbf{S}_\ell W_l \tag{4.25}$$

where $x_{\mathcal{L}_i}^{[\ell]}$ is a vector of length $N(N-2B-T)^{|\mathcal{L}_i|-1}T^{M-|\mathcal{L}_i|}$ in $\mathbb{F}_q$.

For any other undesired message: Consider the undesired message W_k, $k \in \{1, \cdots, M\} \setminus \{\ell\}$. Let $\Delta = 2^{M-2}$ be the number of distinct subsets that contain k and do not contain ℓ. Let $\mathcal{K}_i$, $i \in \{1, \cdots, \Delta\}$ be the ith subset that contains k and does not contain ℓ with indices in ascending order in the size of set $|\mathcal{K}_i|$. Define $u_{\mathcal{K}_i}^{[k]}$ to be the undesired symbol mixtures in the $|\mathcal{K}_i|$th round corresponding to message k among the $\mathcal{K}_i$ set. Define $\sigma_{\mathcal{K}_i}^{[k]}$ to be the side information symbols from message k among the $\mathcal{K}_i$ subset of undesired messages. These side information equations are added to a desired message symbol in the $(|\mathcal{K}_i|+1)$th round. For each subset $\mathcal{K}_i$, the undesired symbols and side information symbols are related via,

$$\begin{bmatrix} u_{\mathcal{K}_i}^{[k]} \\ \sigma_{\mathcal{K}_i}^{[k]} \end{bmatrix} = \mathbf{MDS}_{\frac{N}{T}\alpha_i\times\alpha_i}\mathbf{S}_k\left(\left[\sum_{j=1}^{i-1}\alpha_j + 1 : \sum_{j=1}^{i}\alpha_j\right], :\right) W_k \tag{4.26}$$

where $\alpha_i = (N-2B)(N-2B-T)^{|\mathcal{K}_i|-1}T^{M-|\mathcal{K}_i|}$, $u_{\mathcal{K}_i}^{[k]}$ is a vector of length $\frac{N}{N-2B}\alpha_i$, and $\sigma_{\mathcal{K}_i}^{[k]}$ is a vector of length $\frac{N-2B-T}{T}\cdot\frac{N}{N-2B}\alpha_i$. This implies that the side information $\sigma_{\mathcal{K}_i}^{[k]}$

in the $(|\mathcal{K}_i|+1)$th round is completely determined by $u_{\mathcal{K}_i}^{[k]}$ in the $|\mathcal{K}_i|$th round. We note that these choices of the dimensions ensure that the same number of desired and undesired symbols exist in the $|\mathcal{K}_i|$th round, and they are both equal to $N(N-2B-T)^{|\mathcal{K}_i|-1}T^{M-|\mathcal{K}_i|}$. We further note that the $\frac{N-2B-T}{T}$ factor in the length of $\sigma_{\mathcal{K}_i}^{[k]}$, implies that we generate $\frac{N-2B-T}{T}$ side information symbols for each undesired symbol. We note that the same MDS matrix is used for all messages $k \neq \ell$ that belong to the same subset $\mathcal{K}_i$. This is critical to enable *interference alignment*, and *joint error correction*. Let $X^{[k]} \in \mathbb{F}_q^{N(N-2B)^{M-1}}$ be the vector of mixtures corresponding to message $k \neq \ell$. Then, $X^{[k]}$ is given by,

$$\begin{bmatrix} u_{\mathcal{K}_1}^{[k]} \\ \sigma_{\mathcal{K}_1}^{[k]} \\ u_{\mathcal{K}_2}^{[k]} \\ \sigma_{\mathcal{K}_2}^{[k]} \\ \vdots \\ u_{\mathcal{K}_\Delta}^{[k]} \\ \sigma_{\mathcal{K}_\Delta}^{[k]} \end{bmatrix} = \begin{bmatrix} \mathbf{MDS}_{\frac{N}{T}\alpha_1\times\alpha_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{MDS}_{\frac{N}{T}\alpha_2\times\alpha_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{MDS}_{\frac{N}{T}\alpha_\Delta\times\alpha_\Delta} \end{bmatrix} \mathbf{S}_k([1:T(N-2B)^{M-1}],:)W_k \tag{4.27}$$

Now, we are ready to specify the queries. For every non-empty set $\mathcal{M} \subseteq$

$\{1, \cdots, M\}$, define $\mathcal{Q}_{\mathcal{M}}^{[\ell]}$ to be all queries related to set $\mathcal{M}$,

$$\mathcal{Q}_{\mathcal{M}}^{[\ell]} = \begin{cases} x_{\mathcal{L}_1}^{[\ell]}, & \mathcal{M} = \mathcal{L}_1 = \{\ell\} \\ x_{\mathcal{L}_j}^{[\ell]} + \sum_{k \in \mathcal{K}_i} \sigma_{\mathcal{K}_i}^{[k]} & \exists i, j : \mathcal{M} = \mathcal{K}_i \cup \{\ell\} = \mathcal{L}_j \\ \sum_{k \in \mathcal{K}_i} u_{\mathcal{K}_i}^{[k]} & \exists i : \mathcal{M} = \mathcal{K}_i \end{cases} \tag{4.28}$$

We distribute the queries randomly and evenly among the N databases for each subset $\mathcal{M}$, and the construction is now complete.

4.4.4 Decodability, Privacy, and the Achievable Rate

First, we show how the decoding is performed. The first step is to correct the errors in the undesired symbols in the $\mathcal{K}_i$ set in the $|\mathcal{K}_i|$th round, so that we can generate the correct side information in the $(|\mathcal{K}_i|+1)$th round. Consider again the encoding,

$$\begin{bmatrix} u_{\mathcal{K}_i}^{[k]} \\ \sigma_{\mathcal{K}_i}^{[k]} \end{bmatrix} = \mathbf{MDS}_{\frac{N}{T}\alpha_i \times \alpha_i} \mathbf{S}_k \left(\mathcal{J}_i, :\right) W_k \tag{4.29}$$

where $\mathcal{J}_i = \left[\sum_{j=1}^{i-1} \alpha_j + 1 : \sum_{j=1}^{i} \alpha_j\right]$. Since the sum of linear codes is also a linear code, for the every set $\mathcal{K}_i$, $i \in \{1, \cdots, \Delta\}$, we have

$$\begin{bmatrix} \sum_{k \in \mathcal{K}_i} u_{\mathcal{K}_i}^{[k]} \\ \sum_{k \in \mathcal{K}_i} \sigma_{\mathcal{K}_i}^{[k]} \end{bmatrix} = \mathbf{MDS}_{\frac{N}{T}\alpha_i \times \alpha_i} \sum_{k \in \mathcal{K}_i} \mathbf{S}_k \left(\mathcal{J}_i, :\right) W_k \tag{4.30}$$

This enables *joint error correction* on the aligned sum. The minimum distance of this MDS code is $d^{\mathcal{K}_i} = \frac{N}{T}\alpha_i - \alpha_i + 1 = \frac{N-T}{T}\alpha_i + 1$.

Now, in the $|\mathcal{K}_i|$th round, the user downloads $\sum_{k\in\mathcal{K}_i} u_{\mathcal{K}_i}^{[k]}$ which is a vector of length $\frac{N}{N-2B}\alpha_i$ from all databases. The vector $\sum_{k\in\mathcal{K}_i} u_{\mathcal{K}_i}^{[k]}$ belongs to $\left(\frac{N}{N-2B}\alpha_i, \alpha_i\right)$ punctured MDS code with a puncturing sequence corresponding to the side information symbols, i.e., with a puncturing sequence of length $z = |\sigma_{\mathcal{K}_i}^{[k]}| = \frac{N-2B-T}{T}\cdot\frac{N}{N-2B}\alpha_i$. Therefore,

$$d^{\mathcal{K}_i} - z - 1 = \frac{N-T}{T}\alpha_i - \frac{N-2B-T}{T}\cdot\frac{N}{N-2B}\alpha_i \tag{4.31}$$

$$= \frac{2B}{N-2B}\alpha_i \tag{4.32}$$

$$= 2B(N-2B-T)^{|\mathcal{K}_i|-1}T^{M-|\mathcal{K}_i|} > 0 \tag{4.33}$$

Thus, the $\left(\frac{N}{N-2B}\alpha_i, \alpha_i\right)$ punctured MDS code remains an MDS code with a minimum distance d^{u_i}, such that

$$d^{u_i} = \frac{N}{N-2B}\alpha_i - \alpha_i + 1 \tag{4.34}$$

$$= \frac{2B}{N-2B}\alpha_i + 1 \tag{4.35}$$

Hence, the punctured code can correct upto τ_{u_i} errors, such that

$$\tau_{u_i} \leq \left\lfloor \frac{d^{u_i}-1}{2} \right\rfloor = \frac{B}{N-2B}\alpha_i \tag{4.36}$$

Each database contributes $\frac{1}{N-2B}\alpha_i$ symbols from $\sum_{k\in\mathcal{K}_i} u_{\mathcal{K}_i}^{[k]}$, hence the Byzantine databases can introduce at most $\frac{B}{N-2B}\alpha_i$ errors. Consequently, the punctured MDS code can correct all errors in $\sum_{k\in\mathcal{K}_i} u_{\mathcal{K}_i}^{[k]}$. This results in a corrected undesired

message vector $\left(\sum_{k\in\mathcal{K}_i} u_{\mathcal{K}_i}^{[k]}\right)^*$. Choose any α_i symbols from $\left(\sum_{k\in\mathcal{K}_i} u_{\mathcal{K}_i}^{[k]}\right)^*$. By the MDS property of the $(\frac{N}{T}\alpha_i, \alpha_i)$ MDS code, any $\alpha_i \times \alpha_i$ submatrix is invertible, hence a correct version of the side information vector, which is used in the $(|\mathcal{K}_i|+1)$th round, can be generated. Denote this correct version by $\left(\sum_{k\in\mathcal{K}_i} \sigma_{\mathcal{K}_i}^{[k]}\right)^*$.

Now, we cancel the correct side information successively from each set $\mathcal{K}_i$. Note that the successive correction of side information gives rise to non-linearity in the decoding. After interference cancellation, we are left with $\tilde{X}^{[\ell]}$, which is not exactly $X^{[\ell]}$, as we cancelled the correct side information from the sum and not the side information provided by the Byzantine databases. This is not a problem, because $\tilde{X}^{[\ell]}$ and $X^{[\ell]}$ differ in codeword positions if and only if these positions belong to the Byzantine databases, hence the worst-case number of errors in $\tilde{X}^{[\ell]}$ cannot increase. The desired message is encoded by $(N(N-2B)^{M-1}, (N-2B)^M)$ MDS code with minimum distance d^x, such that

$$d^x = N(N-2B)^{M-1} - (N-2B)^M + 1 \tag{4.37}$$

$$= 2B(N-2B)^{M-1} + 1 \tag{4.38}$$

Each database returns $(N-2B)^{M-1}$ symbols from the desired message. The B Byzantine databases can at most introduce $B(N-2B)^{M-1}$ errors. The outer MDS code can correct up to τ_x errors, such that

$$\tau_x \leq \left\lfloor \frac{d^x - 1}{2} \right\rfloor = B(N-2B)^{M-1} \tag{4.39}$$

Thus, the user can correct all the errors introduced by the Byzantine databases to get a correct vector $\left(X^{[\ell]}\right)^* \in \mathbb{F}_q^{N(N-2B)^{M-1}}$. Consider any $(N-2B)^M$ symbols from $\left(X^{[\ell]}\right)^*$. Denote these symbols by x_ℓ^*, and index them by $\mathcal{I}_x$. Then, the user can decode W_ℓ with zero error via

$$W_\ell = (\mathbf{MDS}_{N(N-2B)^{M-1}\times(N-2B)^M}(\mathcal{I}_x,:)\mathbf{S}_1)^{-1}x_\ell^* \tag{4.40}$$

This is true as matrix $\mathbf{MDS}_{N(N-2B)^{M-1}\times(N-2B)^M}(\mathcal{I}_x,:)\mathbf{S}_1$ is invertible by the MDS property.

In addition, the user can identify the Byzantine databases by comparing the correct versions of the undesired symbols at each cancellation step $(\sum_{k\in\mathcal{K}_i} u_{\mathcal{K}_i}^{[k]})^*$, and the desired symbols $\left(X^{[\ell]}\right)^*$ by their counterparts from the retrieval process. Any change between the correct vector and the retrieved vector implies that this database is a Byzantine database (or unsynchronized). The user can expurgate the malicious nodes in this case as in [7, 60, 61].

Next, we show how the privacy is achieved. The queries for any T colluding databases are comprised of $T(N-2B)^{M-1}$ mixed symbols from each message W_i, $i \in \{1, \cdots, M\}$. Let these symbols be indexed by $\mathcal{I}$. Denote the kth message symbols by $x_{\mathcal{I}}^{[k]}$. For the desired symbols, we have

$$x_{\mathcal{I}}^{[\ell]} = \mathbf{MDS}_{N(N-2B)^{M-1}\times(N-2B)^M}(\mathcal{I},:)\mathbf{S}_\ell W_l \tag{4.41}$$

Since $|\mathcal{I}| = T(N-2B)^{M-1} < (N-2B)^M$ as $2B+T<N$ by construction, and due

to the MDS property, the symbols $x_{\mathcal{I}}^{[\ell]}$ have full-rank. Hence, they are independent and uniformly distributed. Furthermore, for any undesired message W_k, $k \neq \ell$, we have,

$$x_{\mathcal{I}}^{[k]} = \underbrace{\begin{bmatrix} \mathbf{MDS}_{\frac{N}{T}\alpha_1 \times \alpha_1}(\mathcal{I}_1, :) & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{MDS}_{\frac{N}{T}\alpha_\Delta \times \alpha_\Delta}(\mathcal{I}_\Delta, :) \end{bmatrix}}_{\Phi} \mathbf{S}_k([1 : T(N-2B)^{M-1}], :) W_k \tag{4.42}$$

where $\mathcal{I} = \bigcup_{j=1}^{\Delta} \mathcal{I}_j$, and $|\mathcal{I}_j| = \alpha_j$. Therefore, each submatrix in Φ is an $\alpha_i \times \alpha_i$ invertible matrix by the MDS property. Hence, Φ is also an invertible matrix because it is a block-diagonal matrix. By Lemma 4.2, we have

$$x_{\mathcal{I}}^{[k]} \sim \mathbf{S}_k([1 : T(N-2B)^{M-1}], :) W_k \tag{4.43}$$

Thus, symbols $x_{\mathcal{I}}^{[k]}$ are independent and uniformly distributed, and the privacy is guaranteed.

We next calculate the achievable resilient rate. We note that the scheme operates in M rounds. At the ith round, the scheme downloads $\binom{M-1}{i-1}(N-2B-T)^{i-1}T^{M-i}$ equations in the form of one desired symbol added to $i-1$ symbols from the undesired messages, and $\binom{M-1}{i}(N-2B-T)^{i-1}T^{M-i}$ undesired symbols only. Then, the total download in the ith round is $\binom{M}{i}(N-2B-T)^{i-1}T^{M-i}$ from each

database, i.e., the total download of the scheme, D, is $D = N\sum_{i=1}^{M}\binom{M}{i}(N-2B-T)^{i-1}T^{M-i}$. The scheme decodes correctly the desired message, which has length $L = (N-2B)^M$. Thus, the resilient retrieval rate is,

$$
\begin{aligned}
R &= \frac{L}{D} && (4.44)\\
&= \frac{(N-2B)^M}{N\sum_{i=1}^{M}\binom{M}{i}(N-2B-T)^{i-1}T^{M-i}} && (4.45)\\
&= \frac{N-2B}{N}\cdot\frac{(N-2B)^{M-1}}{\sum_{i=1}^{M}\binom{M}{i}(N-2B-T)^{i-1}T^{M-i}} && (4.46)\\
&= \frac{N-2B}{N}\cdot\frac{(N-2B)^{M-1}}{\frac{1}{N-2B-T}\sum_{i=1}^{M}\binom{M}{i}(N-2B-T)^{i}T^{M-i}} && (4.47)\\
&= \frac{N-2B}{N}\cdot\frac{(N-2B)^{M-1}}{\frac{1}{N-2B-T}\left((N-2B)^M-T^M\right)} && (4.48)\\
&= \frac{N-2B}{N}\cdot\frac{(N-2B)^M-T(N-2B)^{M-1}}{(N-2B)^M-T^M} && (4.49)\\
&= \frac{N-2B}{N}\cdot\frac{1-\frac{T}{N-2B}}{1-\left(\frac{T}{N-2B}\right)^M} && (4.50)
\end{aligned}
$$

which is the expression in Theorem 4.1. We have some additional remarks about the achievable scheme.

Remark 4.8 *We note that our achievable scheme is capable of identifying the Byzantine databases by observing discrepancies between the corrected codewords of desired and undesired messages and their counterparts from the retrieval process. Therefore, if multiple-rounds are allowed in the achievable scheme, we can remove the databases that introduce errors at each retrieval round, and achieve larger retrieval rates in future rounds. For instance, assume that $\tilde{B} \leq B$ databases commit errors and are identified to be Byzantine in the kth retrieval round, then removing*

these databases from the system and downloading only from the remaining $(N-\tilde{B})$ *databases, we can achieve the following retrieval rate in the* $(k+1)th$ *round*

$$R^{(k+1)} = \frac{N-\tilde{B}-2(B-\tilde{B})}{N-\tilde{B}} \cdot \frac{1-\frac{T}{N-\tilde{B}-2(B-\tilde{B})}}{1-(\frac{T}{N-\tilde{B}-2(B-\tilde{B})})^M} \tag{4.51}$$

$$= \frac{N+\tilde{B}-2B}{N-\tilde{B}} \cdot \frac{1-\frac{T}{N+\tilde{B}-2B}}{1-(\frac{T}{N+\tilde{B}-2B})^M} \tag{4.52}$$

In particular, if all B *Byzantine databases act maliciously in the* k*th retrieval round and get identified, i.e.,* $\tilde{B}=B$*, then we can achieve the following retrieval rate in the* $(k+1)th$ *round*

$$R^{(k+1)} = \frac{1-\frac{T}{N-B}}{1-(\frac{T}{N-B})^M} \tag{4.53}$$

which is the retrieval rate if B *databases are just unresponsive.*

Remark 4.9 *Our achievable scheme can be seamlessly extended to the case of BPIR with* U *unresponsive databases (as in the case of RPIR [14]) – also known in the literature as* T*-private* B*-Byzantine* $(N-U)$*-out-of-*N *PIR as in [56]. The construction of the achievable scheme can be done by replacing every* $N-2B$ *with* $N-2B-U$ *in the general achievable scheme. Using Lemma 4.3, that states that correct decoding is possible if* $2\tau+\rho \leq d-1$*, and considering the effect of the unresponsive databases as erasures, i.e., via* ρ*, the decodability holds for the BPIR problem with*

unresponsive databases. The retrieval rate in this case is,

$$R = \frac{N-2B-U}{N-U} \cdot \frac{1-\frac{T}{N-2B-U}}{1-(\frac{T}{N-2B-U})^M} \tag{4.54}$$

The retrieval expression is the same as the BPIR capacity in (4.11) if the number of databases is $N-U$. This in turn implies that the expression in (4.54) is the capacity of the BPIR problem with unresponsive databases. The details of the construction and the analysis are omitted to avoid repetition.

4.4.5 Further Examples

In this section, we present some further simple examples with tractable parameters of M, N, T, B for better understanding of the achievable scheme. Here, we use increased number of messages ($M = 3$) and databases ($N = 6$) compared to the selections $M = 2$, $N = 5$ in the motivating example in Section 4.4.2. In the following two subsections, we choose $T = 1$, $B = 2$ and $T = 2$, $B = 1$, respectively, to show the different effects of colluding and Byzantine behavior. We assume without loss of generality that the desired message is W_1.

4.4.5.1 $M = 3$ Messages, $N = 6$, $T = 1$, $B = 2$ Databases

We denote the mixed symbols of messages W_1, W_2, W_3 by a, b, c, respectively. In this example $L = (N-2B)^M = 8$, hence we use 8×8 random mixing matrices denoted by $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3$. We have $\mathcal{L}_1 = \{1\}, \mathcal{L}_2 = \{1,2\}, \mathcal{L}_3 = \{1,3\}, \mathcal{L}_4 = \{1,2,3\}$. Also, for the undesired message W_2, we have $\mathcal{K}_1 = \{2\}, \mathcal{K}_2 = \{2,3\}$, and sim-

ilarly for W_3. The scheme starts with downloading $T^{M-1} = 1$ symbol from each message from each database. Therefore, in round 1, the scheme downloads $x^{[1]}_{\mathcal{L}_1} = a_{[1:6]}$, $u^{[2]}_{\mathcal{K}_1} = b_{[1:6]}$, and $u^{[3]}_{\mathcal{K}_1} = c_{[1:6]}$; see Table 4.2. For every undesired symbol in round 1, we generate $\frac{N-2B-T}{T} = 1$ side information symbols to be used in round 2. The scheme constructs the side information symbols $\sigma^{[2]}_{\mathcal{K}_1} = b_{[7:12]}$ based on the downloaded symbols $b_{[1:6]}$, and similarly for $\sigma^{[3]}_{\mathcal{K}_1} = c_{[7:12]}$. Round 2 contains all combinations of the sums of 2 messages. Round 2 adds one new symbol from the desired message with one symbol of the generated side information from b, c. This results in the sums $x^{[1]}_{\mathcal{L}_2} + \sigma^{[2]}_{\mathcal{K}_1} = a_{[7:12]} + b_{[7:12]}$, and the sums $x^{[1]}_{\mathcal{L}_3} + \sigma^{[3]}_{\mathcal{K}_1} = a_{[13:18]} + c_{[7:12]}$. By message symmetry, we must include the undesired symbol sum $\sum_{k\in\mathcal{K}_2} u^{[k]}_{\mathcal{K}_2} = b_{[13:18]} + c_{[13:18]}$; see Table 4.2. We note that these undesired information equation is in the form of *aligned sums*. The undesired symbols in round 2 generate the side information equations $\sum_{k\in\mathcal{K}_2} \sigma^{[k]}_{\mathcal{K}_2} = b_{[19:24]} + c_{[19:24]}$. These side information equations are added to new symbols from the desired message to have $x^{[1]}_{\mathcal{L}_4} + \sum_{k\in\mathcal{K}_2} \sigma^{[k]}_{\mathcal{K}_2} = a_{[19:24]} + b_{[19:24]} + c_{[19:24]}$. The query table is shown in Table 4.2.

Table 4.2: The query table for the case $M = 3$, $N = 6$, $T = 1$, $B = 2$.

DB 1	DB 2	DB 3	DB 4	DB 5	DB 6
a_1	a_2	a_3	a_4	a_5	a_6
b_1	b_2	b_3	b_4	b_5	b_6
c_1	c_2	c_3	c_4	c_5	c_6
$a_7 + b_7$	$a_8 + b_8$	$a_9 + b_9$	$a_{10} + b_{10}$	$a_{11} + b_{11}$	$a_{12} + b_{12}$
$a_{13} + c_7$	$a_{14} + c_8$	$a_{15} + c_9$	$a_{16} + c_{10}$	$a_{17} + c_{11}$	$a_{18} + c_{12}$
$b_{13} + c_{13}$	$b_{14} + c_{14}$	$b_{15} + c_{15}$	$b_{16} + c_{16}$	$b_{17} + c_{17}$	$b_{18} + c_{18}$
$a_{19}+b_{19}+c_{19}$	$a_{20}+b_{20}+c_{20}$	$a_{21}+b_{21}+c_{21}$	$a_{22}+b_{22}+c_{22}$	$a_{23}+b_{23}+c_{23}$	$a_{24}+b_{24}+c_{24}$

The specific construction of the symbol mixtures are,

$$a_{[1:24]} = \mathbf{MDS}_{24\times 8}\mathbf{S}_1 W_1 \tag{4.55}$$

$$b_{[1:24]} = \begin{bmatrix} u^{[2]}_{\mathcal{K}_1} \\ \sigma^{[2]}_{\mathcal{K}_1} \\ u^{[2]}_{\mathcal{K}_2} \\ \sigma^{[2]}_{\mathcal{K}_2} \end{bmatrix} = \begin{bmatrix} \mathbf{MDS}_{12\times 2} & \mathbf{0} \\ \mathbf{0} & \mathbf{MDS}_{12\times 2} \end{bmatrix} \mathbf{S}_2([1:4],:)W_2 \tag{4.56}$$

$$c_{[1:24]} = \begin{bmatrix} u^{[3]}_{\mathcal{K}_1} \\ \sigma^{[3]}_{\mathcal{K}_1} \\ u^{[3]}_{\mathcal{K}_2} \\ \sigma^{[3]}_{\mathcal{K}_2} \end{bmatrix} = \begin{bmatrix} \mathbf{MDS}_{12\times 2} & \mathbf{0} \\ \mathbf{0} & \mathbf{MDS}_{12\times 2} \end{bmatrix} \mathbf{S}_3([1:4],:)W_3 \tag{4.57}$$

For the decodability, we note that $B = 2$ Byzantine databases can introduce at most 2 errors in $b_{[1:6]}$, 2 errors in $c_{[1:6]}$, 2 errors in $b_{[13:18]} + c_{[13:18]}$, and 8 errors in $a_{[1:24]}$. We note that $b_{[1:6]}$ is encoded via $(6, 2)$ punctured MDS code, which still is an MDS code because $z = 6 < 12 - 2 = 10$. The $(6, 2)$ punctured MDS code can correct errors up to $\lfloor \frac{6-2}{2} \rfloor = 2$ errors. Then, the 2 errors in $b_{[1:6]}$ can be corrected. The same argument holds for $c_{[1:6]}$. For $b_{[13:18]} + c_{[13:18]}$, since the same generator matrix is used for $b_{[13:18]}$, $c_{[13:18]}$, and because of the linearity of the code, the *aligned sum* is a codeword from $(6, 2)$ punctured MDS code as well. Thus, we can correct all the errors in the aligned sum $b_{[13:18]} + c_{[13:18]}$. Knowing the correct undesired symbols results in decoding the correct side information symbols $b_{[7:12]}$, $c_{[7:12]}$ and $b_{[19:24]} + c_{[19:24]}$, respectively, by the MDS property. Cancelling these side information

from the answer strings, we are left with $\tilde{a}_{[1:24]}$, which are coded with an outer $(24, 8)$ MDS code, which is capable of correcting $\lfloor \frac{24-8}{2} \rfloor = 8$ errors. Hence, the user can correct all the errors introduced by the Byzantine databases and W_1 is decodable.

For the privacy, from any individual database, the user asks for 4 mixed symbols from each message. Because of the MDS property, the symbols from all messages are full-rank, and hence they are independent and uniformly distributed. Thus, the scheme is private.

The resilient achievable rate is $R = \frac{8}{42} = \frac{4}{21} = \frac{1}{3} \cdot \frac{4}{7} = \frac{N-2B}{N} \cdot \frac{1-\frac{T}{N-2B}}{1-\left(\frac{T}{N-2B}\right)^M} = C$.

4.4.5.2 $M = 3$ Messages, $N = 6$, $T = 2$, $B = 1$ Databases

In this case $L = (N-2B)^M = 64$, and we use random mixing matrices $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3$ of size 64×64. The scheme starts by downloading $T^{M-1} = 4$ symbols from each message from each database, namely, $a_{[1:24]}$, $b_{[1:24]}$, $c_{[1:24]}$; see Table 4.3. The undesired symbols from $b_{[1:24]}$ and $c_{[1:24]}$ create $\frac{N-2B-T}{T} = 1$ side information symbol for each undesired symbol in a single database. Therefore, the scheme generates the side information $b_{[25:48]}$, $c_{[25:48]}$. In round 2, these side information are added to $a_{[25:48]}, a_{[49:72]}$, respectively. Round 2 concludes by applying message symmetry, and downloads $b_{[49:72]} + c_{[49:72]}$. These undesired symbols produce $b_{[73:96]} + c_{[73:96]}$ as side information symbols for round 3. The query table is shown in Table 4.3.

The specific construction of the symbol mixtures are,

$$a_{[1:96]} = \mathbf{MDS}_{96\times 64}\mathbf{S}_1 W_1 \tag{4.58}$$

Table 4.3: The query table for the case $M = 3$, $N = 6$, $T = 2$, $B = 1$.

DB 1	DB 2	DB 3	DB 4	DB 5	DB 6
a_1, a_2, a_3, a_4	a_5, a_6, a_7, a_8	$a_9, a_{10}, a_{11}, a_{12}$	$a_{13}, a_{14}, a_{15}, a_{16}$	$a_{17}, a_{18}, a_{19}, a_{20}$	$a_{21}, a_{22}, a_{23}, a_{24}$
b_1, b_2, b_3, b_4	b_5, b_6, b_7, b_8	$b_9, b_{10}, b_{11}, b_{12}$	$b_{13}, b_{14}, b_{15}, b_{16}$	$b_{17}, b_{18}, b_{19}, b_{20}$	$b_{21}, b_{22}, b_{23}, b_{24}$
c_1, c_2, c_3, c_4	c_5, c_6, c_7, c_8	$c_9, c_{10}, c_{11}, c_{12}$	$c_{13}, c_{14}, c_{15}, c_{16}$	$c_{17}, c_{18}, c_{19}, c_{20}$	$c_{21}, c_{22}, c_{23}, c_{24}$
$a_{25} + b_{25}$	$a_{29} + b_{29}$	$a_{33} + b_{33}$	$a_{37} + b_{37}$	$a_{41} + b_{41}$	$a_{45} + b_{45}$
$a_{26} + b_{26}$	$a_{30} + b_{30}$	$a_{34} + b_{34}$	$a_{38} + b_{38}$	$a_{42} + b_{42}$	$a_{46} + b_{30}$
$a_{27} + b_{27}$	$a_{31} + b_{31}$	$a_{35} + b_{35}$	$a_{39} + b_{39}$	$a_{43} + b_{43}$	$a_{47} + b_{47}$
$a_{28} + b_{28}$	$a_{32} + b_{32}$	$a_{36} + b_{36}$	$a_{40} + b_{40}$	$a_{44} + b_{44}$	$a_{48} + b_{48}$
$a_{49} + c_{25}$	$a_{53} + c_{29}$	$a_{57} + c_{33}$	$a_{61} + c_{37}$	$a_{65} + c_{41}$	$a_{69} + c_{45}$
$a_{50} + c_{26}$	$a_{54} + c_{30}$	$a_{58} + c_{34}$	$a_{62} + c_{38}$	$a_{66} + c_{42}$	$a_{70} + c_{30}$
$a_{51} + c_{27}$	$a_{55} + c_{31}$	$a_{59} + c_{35}$	$a_{63} + c_{39}$	$a_{67} + c_{43}$	$a_{71} + c_{47}$
$a_{52} + c_{28}$	$a_{56} + c_{32}$	$a_{60} + c_{36}$	$a_{64} + c_{40}$	$a_{68} + c_{44}$	$a_{72} + c_{48}$
$b_{49} + c_{49}$	$b_{53} + c_{53}$	$b_{57} + c_{57}$	$b_{61} + c_{61}$	$b_{65} + c_{65}$	$b_{69} + c_{69}$
$b_{50} + c_{50}$	$b_{54} + c_{54}$	$b_{58} + c_{58}$	$b_{62} + c_{62}$	$b_{66} + c_{66}$	$b_{70} + c_{70}$
$b_{51} + c_{51}$	$b_{55} + c_{55}$	$b_{59} + c_{59}$	$b_{63} + c_{63}$	$b_{67} + c_{67}$	$b_{71} + c_{71}$
$b_{52} + c_{52}$	$b_{56} + c_{56}$	$b_{60} + c_{60}$	$b_{64} + c_{64}$	$b_{68} + c_{68}$	$b_{72} + c_{72}$
$a_{73}+b_{73}+c_{73}$	$a_{77}+b_{77}+c_{20}$	$a_{81}+b_{81}+c_{81}$	$a_{85}+b_{85}+c_{85}$	$a_{89}+b_{89}+c_{89}$	$a_{93}+b_{93}+c_{93}$
$a_{74}+b_{74}+c_{74}$	$a_{78}+b_{78}+c_{78}$	$a_{82}+b_{82}+c_{82}$	$a_{86}+b_{86}+c_{86}$	$a_{90}+b_{90}+c_{90}$	$a_{94}+b_{94}+c_{94}$
$a_{75}+b_{75}+c_{75}$	$a_{79}+b_{79}+c_{79}$	$a_{83}+b_{83}+c_{83}$	$a_{87}+b_{87}+c_{87}$	$a_{91}+b_{91}+c_{91}$	$a_{95}+b_{95}+c_{95}$
$a_{76}+b_{76}+c_{76}$	$a_{80}+b_{80}+c_{80}$	$a_{84}+b_{84}+c_{84}$	$a_{88}+b_{88}+c_{89}$	$a_{92}+b_{92}+c_{92}$	$a_{96}+b_{96}+c_{96}$

$$b_{[1:96]} = \begin{bmatrix} \mathbf{MDS}_{48\times16} & \mathbf{0} \\ \mathbf{0} & \mathbf{MDS}_{48\times16} \end{bmatrix} \mathbf{S}_2([1:32],:)W_2 \tag{4.59}$$

$$c_{[1:96]} = \begin{bmatrix} \mathbf{MDS}_{48\times16} & \mathbf{0} \\ \mathbf{0} & \mathbf{MDS}_{48\times16} \end{bmatrix} \mathbf{S}_3([1:32],:)W_3 \tag{4.60}$$

For the decodability, the Byzantine database can commit 4 errors in $b_{[1:24]}$, 4 errors in $c_{[1:24]}$, 4 errors in $b_{[49:72]} + c_{[49:72]}$, and 16 errors in $a_{[1:96]}$. All layers of the undesired symbols are encoded via $(24, 16)$ punctured MDS code, which is still MDS code, and can correct up to $\lfloor \frac{24-16}{2} \rfloor = 4$ errors. Therefore, all the undesired symbols can be corrected, which in turn generate the correct side information in all layers. By canceling the side information, we are left with $\tilde{a}_{[1:96]}$, which is encoded by $(96, 64)$ outer MDS code. This code can correct up to $\lfloor \frac{96-64}{2} \rfloor = 16$ errors. Hence,

the user can decode W_1 reliably.

For the privacy, from any 2 databases, the user asks for 16 symbols from each message. By the MDS property and Lemma 4.2, all these symbols are full-rank, and hence they are independent and uniformly distributed. Therefore, the scheme is 2-private.

The resilient achievable rate is $R = \frac{64}{168} = \frac{8}{21} = \frac{4}{6} \cdot \frac{4}{7} = \frac{N-2B}{N} \cdot \frac{1-\frac{T}{N-2B}}{1-\left(\frac{T}{N-2B}\right)^M} = C$.

Note that, for the same M, N, the achievable rate with $T = 1$, $B = 2$ in the previous subsection, $\frac{4}{21}$, is smaller than the achievable rate with $T = 2$, $B = 1$ in this subsection, $\frac{8}{21}$, which signifies that Byzantine behavior is a more severe adversarial behavior to cope with compared to colluding behavior.

4.5 Converse Proof

In this section, we develop an upper bound for the BPIR problem. We adapt the cut-set upper bound proof in [60, 61] to the PIR setting. The upper bound can be thought of as a network version of the Singleton bound [62]. The upper bound intuitively asserts that the effect of the Byzantine databases on the retrieval rate is harmful as if $2B$ databases are removed from the retrieval process, but the user still needs to access them. The settings of PIR and network coding problem in [60] share that they are both planar networks, and they both lack backward edges, as we consider here a single-round retrieval, and hence the answer strings from the honest databases are not affected by the answers of the Byzantine databases. However, some technical differences arise in the PIR setting:

1. Unlike the adversarial nodes in [60, 61], the Byzantine databases in PIR are not fully omniscient, since they do not know which message the user wishes to retrieve (by definition of PIR). Consequently, we assume in the following that the Byzantine databases alter the contents of the entire database.

2. In the PIR setting, the user does not know the entire codebook in advance, in contrast to the network coding problem in [60].

For sake of deriving an upper bound, we make the following simplifications:

1. We assume that the actions of the Byzantine databases are restricted to altering the contents of the entire database, i.e., the nth Byzantine database changes its contents Ω_n from $\mathcal{W}$ to $\tilde{\mathcal{W}}$, where $\tilde{\mathcal{W}} \neq \mathcal{W}$. This restriction is valid from the converse point of view, since it potentially results in a weaker adversary, which in turn results in a higher rate. Note that, in this sense the Byzantine databases are reduced to being unsynchronized databases (with unknown number of mis-synchronized messages).

2. We further restrict the answering string from the nth database to be a deterministic function $f_n(\cdot)$, i.e., $A_n^{[i]} = f_n(\Omega_n, Q_n^{[i]})$, of the altered database Ω_n. This restriction also limits the capabilities of the Byzantine databases. This results in a further upper bound on rate. Since we restrict the actions of the Byzantine databases to altering Ω_n only, we signify this dependence on Ω_n by writing the answering string $A_n^{[i]}$ as $A_n^{[i]}(\Omega_n)$.

3. We can assume that the retrieval scheme is symmetric. This is without loss

of generality, since any asymmetric PIR scheme can be made symmetric by proper time sharing without changing the retrieval rate [12, 117, 123], i.e.,

$$H(A_1^{[i]}|\mathcal{Q}) = H(A_2^{[i]}|\mathcal{Q}) = \cdots = H(A_N^{[i]}|\mathcal{Q}) \tag{4.61}$$

This assumption remains true in the BPIR problem, because if the nth Byzantine database returned an answering string which has $H(A_n^{[i]}|\mathcal{Q}) \neq H(A_j^{[i]}|\mathcal{Q})$ for some honest database j, i.e., the answering string has a different length as a response to a symmetric retrieval scheme, this database will be identified as a Byzantine database. Hence, the errors introduced by the Byzantine databases can be mitigated and these databases will be removed from the system afterwards. In addition, the restrictions in assumptions 1 and 2 above imply that the Byzantine databases answer truthfully to the queries based on their own (altered) Ω_n. Therefore, the lengths of the answer strings will be symmetric in response to a symmetric scheme.

The main argument of the converse proof is summarized in the following lemma.

Lemma 4.4 *Fix a set of honest databases $\mathcal{U} \subset \{1, \cdots, N\}$ such that $|\mathcal{U}| = N - 2B$, and $\Omega_n = \mathcal{W}$, for every $n \in \mathcal{U}$. Then, for correct decoding of W_i, the answer strings $A_{\mathcal{U}}^{[i]}(\mathcal{W})$ is unique for every realization of $\mathcal{W}$, i.e., there cannot exist two realizations of the message set $\mathcal{W}, \tilde{\mathcal{W}}$, such that $\mathcal{W} \neq \tilde{\mathcal{W}}$, and $A_{\mathcal{U}}^{[i]}(\mathcal{W}) = A_{\mathcal{U}}^{[i]}(\tilde{\mathcal{W}})$.*

We have this following remark about Lemma 4.4 first, before we give its proof next.

Remark 4.10 *Lemma 4.4 implies that the answer strings from any $N-2B$ honest databases are enough to reconstruct the desired message, since every realization of the message set produces different answering strings from any $N-2B$ databases. This argument was previously used by [60, Theorem 1] and [61, Theorem 6], as they show that the capacity of the adversarial network coding problem and the adversarial distributed storage problem, respectively, is upper bounded by the capacity of the edges of any cut in the network after removing $2B$ edges from this cut. These edges correspond to the set $\mathcal{U}$ in our problem. The proof in [60, 61] relies on the fact that in the presence of an adversary controlling B nodes, and for any distinct messages $w_1 \neq w_2$, a necessary condition for the receiver to not make a decoding error is to have $X_{\mathcal{U}}(w_1) \neq X_{\mathcal{U}}(w_2)$.*

Proof: Divide the set $\bar{\mathcal{U}} = \{1, \cdots, N\} \setminus \mathcal{U}$ into two sets $\mathcal{B}_1$, $\mathcal{B}_2$ such that $|\mathcal{B}_1| = |\mathcal{B}_2| = B$. In the BPIR problem, we must guarantee correct decoding if the Byzantine databases are any subset $\mathcal{B} \subset \{1, \cdots, N\}$, such that $|\mathcal{B}| = B$, in particular, if the Byzantine databases are either $\mathcal{B}_1$ or $\mathcal{B}_2$.

Now, assume for sake of contradiction, that there exists a valid retrieval scheme that achieves correct decoding of W_i, and there exist two realizations of the message set $\mathcal{W}$, $\tilde{\mathcal{W}}$ such that $\mathcal{W} \neq \tilde{\mathcal{W}}$, and

$$A_{\mathcal{U}}^{[i]}(\mathcal{W}) = A_{\mathcal{U}}^{[i]}(\tilde{\mathcal{W}}) \tag{4.62}$$

Two scenarios can arise:

1. The true realization of the database contents is $\mathcal{W}$. In this case, if the adversarial nodes are the databases indexed by $\mathcal{B}_2$, and they flip their contents $\Omega_{\mathcal{B}_2}$ into $\tilde{\mathcal{W}}$, the user collects the answer strings $\left(A_{\mathcal{B}_1}^{[i]}(\mathcal{W}), A_{\mathcal{B}_2}^{[i]}(\tilde{\mathcal{W}}), A_{\mathcal{U}}^{[i]}(\mathcal{W})\right)$.

2. The true realization of the database contents is $\tilde{\mathcal{W}}$. In this case, if the adversarial nodes are the databases indexed by $\mathcal{B}_1$, and they flip their contents $\Omega_{\mathcal{B}_1}$ into $\mathcal{W}$, the user collects the answer strings $\left(A_{\mathcal{B}_1}^{[i]}(\mathcal{W}), A_{\mathcal{B}_2}^{[i]}(\tilde{\mathcal{W}}), A_{\mathcal{U}}^{[i]}(\tilde{\mathcal{W}})\right)$.

Since $A_{\mathcal{U}}^{[i]}(\mathcal{W}) = A_{\mathcal{U}}^{[i]}(\tilde{\mathcal{W}})$, there is no way for the user to differentiate between the two scenarios. Hence, the user commits an error either directly (if $\mathcal{W}$ and $\tilde{\mathcal{W}}$ differ in W_i) or indirectly (if $\mathcal{W}$ and $\tilde{\mathcal{W}}$ differ in any message other than W_i, as the user fails in canceling the interference from the answer strings). This is a contradiction to the reliability constraint $H(W_i|A_{1:N}^{[i]}, Q_{1:N}^{[i]}) = 0$. ■

Now, we continue with the main body of the converse proof. From Lemma 4.4, the answers $A_{\mathcal{U}}^{[i]}(\mathcal{W})$ are unique for every $\mathcal{W}$, hence restricting the decoding function to these answers uniquely determine W_i, i.e., there exists no further confusion about the correct database contents $\mathcal{W}$, and the answering strings are designed to retrieve W_i from this $\mathcal{W}$. Consequently, if the true realization of the database is $\mathcal{W}$, we can write

$$R = \frac{L}{\sum_{n=1}^{N} H(A_n^{[i]})} \tag{4.63}$$

$$\leq \frac{L}{\sum_{n=1}^{N} H(A_n^{[i]}|\mathcal{Q})} \tag{4.64}$$

$$= \frac{N-2B}{N} \cdot \frac{L}{(N-2B)H(A_1^{[i]}|\mathcal{Q})} \tag{4.65}$$

$$= \frac{N-2B}{N} \cdot \frac{L}{\sum_{n\in\mathcal{U}} H(A_n^{[i]}(\mathcal{W})|\mathcal{Q})} \tag{4.66}$$

$$\leq \frac{N-2B}{N} \cdot C_T(|\mathcal{U}|) \tag{4.67}$$

$$= \frac{N-2B}{N} \cdot C_T(N-2B) \tag{4.68}$$

$$= \frac{N-2B}{N} \cdot \frac{1-\frac{T}{N-2B}}{1-\left(\frac{T}{N-2B}\right)^M} \tag{4.69}$$

where $C_T(\cdot)$ is the capacity of the PIR problem with T colluding databases as a function of the number of databases. Here, (4.65) follows from the symmetry assumption, (4.66) follows from the fact that $A_{\mathcal{U}}^{[i]}(\mathcal{W})$ can decode W_i correctly and then $\frac{L}{\sum_{n\in\mathcal{U}} H(A_n^{[i]}(\mathcal{W})|\mathcal{Q})}$ is a valid upper bound on the retrieval rate under the T-privacy constraint if the accessed databases are restricted to $\mathcal{U}$, which is further upper bounded by the TPIR capacity $C_T(|\mathcal{U}|)$ in (4.67) as $C_T(|\mathcal{U}|)$ is the supremum of all rates that can be achieved using the set of databases $\mathcal{U}$ under the T-privacy constraint, and (4.69) follows from the capacity expression in [14].

4.6 Conclusions

In this chapter, we investigated the PIR problem from N replicated databases in the presence of B Byzantine databases, and T-colluding databases from an information-theoretic perspective. We determined the exact capacity of the BPIR problem to be $C = \frac{N-2B}{N} \cdot \frac{1-\frac{T}{N-2B}}{1-(\frac{T}{N-2B})^M}$. The capacity expression shows the severe degradation in the retrieval rate in the presence of Byzantine databases. The expression shows that in order to correct the errors introduced by the adversarial databases, the system needs to have $2B$ redundant storage nodes. The retrieval rate is further penalized by the

factor $\frac{N-2B}{N}$, which reflects the ignorance of the user which $N-2B$ databases are honest. The BPIR capacity converges to $C \to 1-2\gamma$ as B, $N \to \infty$, $B = \gamma N$, where γ is the fraction of Byzantine databases. For large enough number of messages, the BPIR capacity approaches $C \to 1 - \frac{2B+T}{N}$. We extended the optimal scheme for the RPIR problem to permit *error correction* of any error pattern introduced by the Byzantine databases. The new key ingredients in the achievable scheme are: encoding the undesired messages via a punctured MDS code, successive interference cancellation to remove the interfering messages, and encoding the desired message by an outer-layer MDS code. For the converse, we adapted the cut-set bound, which was originally derived for the network coding problem against adversarial nodes, for the PIR setting.

The BPIR problem can be extended in several interesting directions. According to our formulation here, the capacities of unsynchronized and Byzantine PIR problems are the same. However, in the unsynchronized PIR problem, if the user knows in advance that at most S messages are mis-synchronized, and if S is small with respect to M, the user can potentially achieve higher rates than our formulation here, in particular, if it uses a multi-round scheme as in [7]. In addition, in modeling the mis-synchronization, if we consider some specific attack/error patterns (e.g., during mis-synchronization the stored data goes through a noisy channel with a known model), then the user can tailor an error mitigation procedure that fits these attack/error models explicitly, in contrast to our formulation here, where we assumed that the user is prepared for the worst-case errors of any structure. Finally, while we assumed that the B Byzantine databases can be any one of the $\binom{N}{B}$ pos-

sible subsets, the problem can be extended to the case where only a certain subset of all possible $\binom{N}{B}$ Byzantine configurations is possible as in [15] which considered a limited collusion model.

CHAPTER 5

Private Information Retrieval Under Asymmetric Traffic Constraints

5.1 Introduction

In this chapter, we consider the classical setting of PIR of a single message (file) out of M messages from N distributed databases under the new constraint of *asymmetric traffic* from databases. In this problem, the *ratios between the traffic* from the databases are constrained, i.e., the ratio of the length of the answer string that the user (retriever) receives from the nth database to the total length of all answer strings from all databases is constrained to be τ_n. This may happen if the user's access to the databases is restricted due to database availability, channel quality to the databases, and other factors. For this problem, for fixed M, N, we develop a general upper bound $\bar{C}(\boldsymbol{\tau})$, which generalizes the converse proof of Sun-Jafar [12], where database symmetry was inherently used. Our converse bound is a piece-wise affine function in the traffic ratio vector $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_N)$. For the lower bound, we explicitly show the achievability of $\binom{M+N-1}{M}$ corner points. For the remaining traffic ratio vectors, we perform time-sharing between these corner points. The recursive

structure of our achievability scheme is captured via a system of difference equations. The upper and lower bounds exactly match for $M = 2$ and $M = 3$ for any N and any $\boldsymbol{\tau}$. The results show strict loss of PIR capacity due to the asymmetric traffic constraints compared with the symmetric case of Sun-Jafar [12] which implicitly uses $\tau_n = \frac{1}{N}$ for all n.

5.2 System Model

Consider a classical PIR model with N non-communicating and replicated databases storing M messages (or files). Each database stores the same set of messages $W_{1:M} = \{W_1, \cdots, W_M\}$. Messages $W_{1:M}$ are independent and identically distributed over all vectors of size L picked from a finite field $\mathbb{F}_q^L$, i.e.,

$$H(W_i) = L, \quad i \in \{1, \cdots, M\} \tag{5.1}$$

$$H(W_1, \cdots, W_M) = ML, \quad (q\text{-ary units}) \tag{5.2}$$

In the PIR problem, a user wants to retrieve a message $W_i \in W_{1:M}$ correctly without revealing any information about the identity of the message i to any individual database. To that end, the user submits a query $Q_n^{[i]}$ to the nth database. The messages and the queries are statistically independent due to the fact that the user does not know the message realizations in advance, i.e.,

$$I(W_{1:M}; Q_{1:N}^{[i]}) = 0 \tag{5.3}$$

where $Q_{1:N}^{[i]} = \{Q_1^{[i]}, \cdots, Q_N^{[i]}\}$. The nth database responds truthfully by an answer string $A_n^{[i]}$. The answer string $A_n^{[i]}$ is a deterministic function of the query $Q_n^{[i]}$ and all the messages $W_{1:M}$, hence

$$H(A_n^{[i]}|Q_n^{[i]}, W_{1:M}) = 0, \quad n \in \{1, \cdots, N\} \tag{5.4}$$

In the PIR model with asymmetric traffic constraints, the lengths of the answer strings are different (see Fig. 5.1). More specifically, we assume that the nth database responds with a t_n-length answer string, such that $t_n = \lambda_n t_1$, where λ_n is the ratio between the traffic from the nth database to the traffic from the first database. Without loss of generality, we assume that the first database has the highest traffic and the remaining databases are ordered descendingly in λ_n. Hence, $\{\lambda_n\}_{n=1}^N$ is a *non-increasing monotone* sequence with $\lambda_1 = 1$, and $\lambda_n \in [0, 1]$, i.e.,

$$H(A_n^{[i]}) \leq \lambda_n t_1, \quad i \in \{1, \cdots, M\},\ n \in \{1, \cdots, N\},\ 1 \geq \lambda_2 \geq \cdots \geq \lambda_N \tag{5.5}$$

We define the *traffic ratio* of the nth database τ_n as the ratio between the traffic from the nth database and the total traffic from all databases, i.e.,

$$\tau_n = \frac{\lambda_n}{\sum_{j=1}^N \lambda_j} \tag{5.6}$$

We note that there is a one-to-one transformation between the vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_N)$ and the vector $\boldsymbol{\tau} = (\tau_1, \tau_2, \cdots, \tau_N)$. Thus, $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$ are used interchangeably within the context of this chapter.

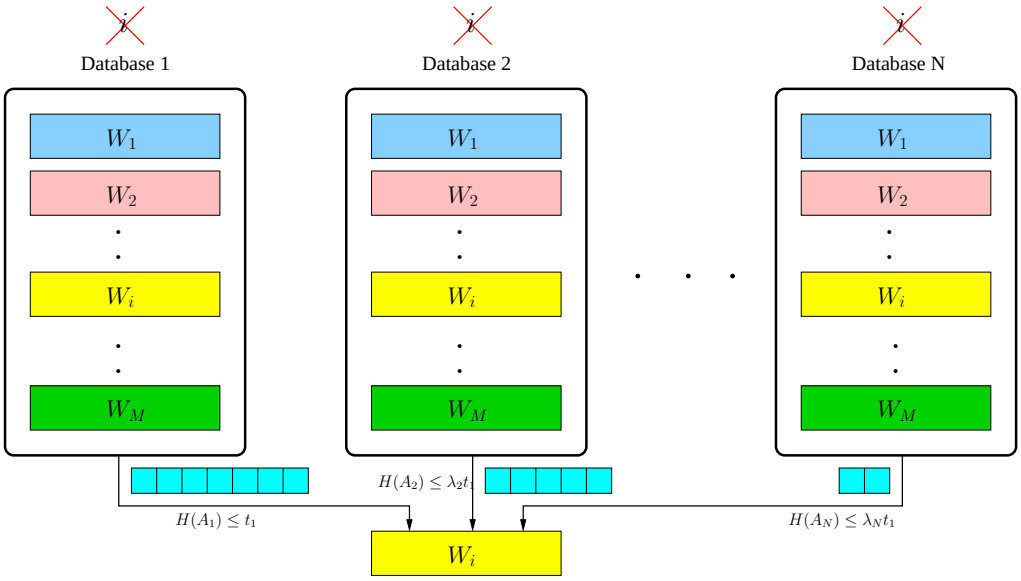


Figure 5.1: PIR under asymmetric traffic constraints.

In order to ensure the privacy, at the nth database, the query $Q_n^{[i]}$ designed to retrieve W_i should be indistinguishable from the queries designed to retrieve any other message, i.e.,

$$(Q_n^{[i]}, A_n^{[i]}, W_{1:M}) \sim (Q_n^{[j]}, A_n^{[j]}, W_{1:M}), \quad \forall j \in \{1, \cdots, M\} \tag{5.7}$$

where $\sim$ denotes statistical equivalence.

In addition, the user should be able to reconstruct W_i from the collected answer strings $A_{1:N}^{[i]}$ with arbitrarily small probability of error. By Fano's inequality, we have the following reliability constraint,

$$H(W_i|Q_{1:N}^{[i]}, A_{1:N}^{[i]}) = o(L) \tag{5.8}$$

where $\frac{o(L)}{L} \to 0$ as $L \to \infty$.

For a fixed N, M, and a traffic ratio vector $\boldsymbol{\tau}$, a retrieval rate $R(\boldsymbol{\tau})$ is achievable if there exists a PIR scheme which satisfies the privacy constraint (5.7) and the reliability constraint (5.8) for some message lengths $L(\boldsymbol{\tau})$ and answer strings of lengths $\{t_n(\boldsymbol{\tau})\}_{n=1}^N$ that satisfy the asymmetric traffic constraint (5.5), such that

$$R(\boldsymbol{\tau}) = \frac{L(\boldsymbol{\tau})}{\sum_{n=1}^N t_n(\boldsymbol{\tau})} \tag{5.9}$$

We note that in this problem, we do not constrain either the message length $L(\boldsymbol{\tau})$ or the lengths of the answer strings $t_n(\boldsymbol{\tau})$, but we rather constrain the ratios of the traffic of each database with respect to the traffic of the first database. The pair $(L(\boldsymbol{\tau}), t_1(\boldsymbol{\tau}))$ can grow arbitrarily large to conform with the information-theoretic framework.

The capacity of the PIR problem under asymmetric traffic constraints $C(\boldsymbol{\tau})$ is defined as the supremum of all achievable retrieval rates, i.e., $C(\boldsymbol{\tau}) = \sup\ R(\boldsymbol{\tau})$.

5.3 Main Results and Discussions

Our first result is an upper bound on $C(\boldsymbol{\tau})$ as a function of $\boldsymbol{\tau}$ for any fixed M, N.

Theorem 5.1 (Upper bound) *For the PIR problem under monotone non-increasing asymmetric traffic constraints $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_N)$, the PIR capacity $C(\boldsymbol{\tau})$*

is upper bounded by

$$C(\boldsymbol{\tau}) \leq \bar{C}(\boldsymbol{\tau}) = \min_{n_1,\cdots,n_{M-1}\in\{1,\cdots,N\}} \frac{1+\frac{\sum_{n=n_1+1}^{N}\tau_n}{n_1}+\frac{\sum_{n=n_2+1}^{N}\tau_n}{n_1n_2}+\cdots+\frac{\sum_{n=n_{M-1}+1}^{N}\tau_n}{n_0n_1\cdots n_{M-1}}}{1+\frac{1}{n_1}+\frac{1}{n_1n_2}+\cdots+\frac{1}{n_0n_1\cdots n_{M-1}}} \tag{5.10}$$

The proof of this upper bound is given in Section 5.4. We have the following remarks.

Remark 5.1 *The minimization in (5.10) is performed to obtain the tightest bound, i.e., the bound in* (5.10) *is valid for any sequence of* $\{n_i\}_{i=1}^{N} \subset \{1,\cdots,N\}^{M-1}$. *In particular, restricting the minimization in the bound in* (5.10) *to monotone non-decreasing sequences* $\{n_i\}_{i=1}^{M-1} \subset \{1,\cdots,N\}^{M-1}$ *such that* $n_1 \leq n_2 \leq \cdots \leq n_{M-1}$ *is still a valid upper bound. For fixed* M, N, *the number of such monotone bounds is* $\binom{M+N-2}{M-1}$.

Remark 5.2 *The upper bound for the capacity function* $\bar{C}(\boldsymbol{\tau})$ *in (5.10) is a piece-wise affine function in the traffic ratio vector* $\boldsymbol{\tau}$.

Remark 5.3 *The upper bound in* (5.10) *generalizes the known results about the PIR problem. By picking* $n_1 = \cdots = n_{M-1} = N$, (5.10) *leads to*

$$C(\boldsymbol{\tau}) \leq \frac{1}{1+\frac{1}{N}+\frac{1}{N^2}+\cdots+\frac{1}{N^{M-1}}} \tag{5.11}$$

which is the capacity of PIR with symmetric traffic (no traffic constraints) in [12]. On the other hand, if $\boldsymbol{\tau} = (1,0,0,\cdots,0)$, *which implies that no traffic is returned*

from any database except for the first one, by picking $n_1 = \cdots = n_{M-1} = 1$*, the upper bound in* (5.10) *leads to* $\frac{1}{M}$*, which is the capacity of the PIR problem with one database [1].*

The following corollary is a direct consequence of Theorem 5.1. The corollary asserts that there is a strict capacity loss due to the asymmetric traffic constraints if the traffic ratio of the weakest link falls below a certain threshold.

Corollary 5.1 (Asymmetry hurts) *For the PIR problem under monotone non-increasing asymmetric traffic constraints* $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_N)$*, if* $\tau_N < \tau^*$*, such that*

$$\tau^* = \frac{N^{M-1} - 1}{N^M - 1}, \quad N > 1 \tag{5.12}$$

then $C(\boldsymbol{\tau}) < C$*, where* $C = \frac{1}{1+\frac{1}{N}+\cdots+\frac{1}{N^{M-1}}}$ *is the PIR capacity without the asymmetric traffic constraints in [12].*

Proof: From Theorem 5.1, the upper bound corresponding to $n_1 = N - 1$, and $n_2 = \cdots = n_{M-1} = N$ is strictly tighter than the capacity without asymmetric traffic constraints C if

$$\frac{1 + \frac{\tau_N}{N-1}}{1 + \frac{1}{N-1} + \frac{1}{(N-1)N} + \cdots + \frac{1}{(N-1)N^{M-1}}} < C \tag{5.13}$$

which leads to

$$\frac{\tau_N}{N-1}\left(1+\frac{1}{N}+\cdots+\frac{1}{N^{M-1}}\right)<\left(\frac{1}{N-1}-\frac{1}{N}\right)\left(1+\frac{1}{N}+\cdots+\frac{1}{N^{M-2}}\right) \tag{5.14}$$

which further simplifies to

$$\tau_N<\frac{\frac{1}{N}\left(1+\frac{1}{N}+\cdots+\frac{1}{N^{M-2}}\right)}{\left(1+\frac{1}{N}+\cdots+\frac{1}{N^{M-1}}\right)}=\frac{\sum_{i=0}^{M-2}N^i}{\sum_{i=0}^{M-1}N^i}=\tau^* \tag{5.15}$$

which implies that the upper bound for the capacity under the asymmetric traffic constraint is strictly less than C, which in turn implies that any achievable rate is strictly less than the unconstrained capacity. ■

Remark 5.4 *As the number of messages M becomes large enough, i.e., as $M\to\infty$, the traffic ratio threshold in (5.12) $\tau^*\to\frac{1}{N}$. This implies that as $M\to\infty$, any asymmetric traffic constraint incurs strict capacity loss.*

Our second result is a lower bound on $C(\boldsymbol{\tau})$ as a function of $\boldsymbol{\tau}$ for any fixed M, N.

Theorem 5.2 (Lower bound) *For the PIR problem under asymmetric traffic constraints, for a monotone non-decreasing sequence $\mathbf{n}=\{n_i\}_{i=0}^{M-1}\subset\{1,\cdots,N\}^M$, let $n_{-1}=0$, and $\mathcal{S}=\{i\geq 0: n_i-n_{i-1}>0\}$. Denote $y_\ell[k]$ as the number of stages of the achievable scheme that downloads k-sums from the nth database, such that $n_{\ell-1}\leq n\leq n_\ell$, and $\ell\in\mathcal{S}$. Let $\xi_\ell=\prod_{s\in\mathcal{S}\setminus\{\ell\}}\binom{M-2}{s-1}$. The number of stages $y_\ell[k]$ is*

characterized by the following system of difference equations:

$$
\begin{aligned}
y_0[k] &= (n_0-1)y_0[k-1] + \sum_{j\in\mathcal{S}\setminus\{0\}} (n_j-n_{j-1})y_j[k-1] \\
y_1[k] &= (n_1-n_0-1)y_1[k-1] + \sum_{j\in\mathcal{S}\setminus\{1\}} (n_j-n_{j-1})y_j[k-1] \\
y_\ell[k] &= n_0\xi_\ell\delta[k-\ell-1] + (n_\ell-n_{\ell-1}-1)y_\ell[k-1] + \sum_{j\in\mathcal{S}\setminus\{\ell\}} (n_j-n_{j-1})y_j[k-1], \quad \ell \geq 2
\end{aligned}
\tag{5.16}
$$

where $\delta[\cdot]$ denotes the Kronecker delta function. The initial conditions of (5.16) *are* $y_0[1] = \prod_{s\in\mathcal{S}} \binom{M-2}{s-1}$*, and* $y_j[k] = 0$ *for* $k \leq j$*. Consequently, the traffic ratio vector* $\boldsymbol{\tau}(\mathbf{n}) = (\tau_1(\mathbf{n}), \cdots, \tau_N(\mathbf{n}))$ *corresponding to the sequence* $\mathbf{n} = \{n_i\}_{i=0}^{M-1}$ *is given by:*

$$
\tau_n(\mathbf{n}) = \frac{\sum_{k=1}^{M} \binom{M}{k} y_j[k]}{\sum_{\ell\in\mathcal{S}} \sum_{k=1}^{M} \binom{M}{k} y_\ell[k](n_\ell - n_{\ell-1})}, \quad n_{j-1}+1 \leq n \leq n_j \tag{5.17}
$$

and the achievable rate corresponding to $\boldsymbol{\tau}(\mathbf{n})$ *is given by:*

$$
R(\boldsymbol{\tau}(\mathbf{n})) = \frac{\sum_{\ell\in\mathcal{S}} \sum_{k=1}^{M} \binom{M-1}{k-1} y_\ell[k](n_\ell - n_{\ell-1})}{\sum_{\ell\in\mathcal{S}} \sum_{k=1}^{M} \binom{M}{k} y_\ell[k](n_\ell - n_{\ell-1})} \tag{5.18}
$$

Moreover, for $\boldsymbol{\tau} = \sum_{i=1}^{N} \alpha_i \boldsymbol{\tau}(\mathbf{n}_i)$ *for* $\alpha_i \geq 0$*, for all* i*, and* $\sum_{i=1}^{N} \alpha_i = 1$*, the following is a lower bound on* $C(\boldsymbol{\tau})$*,*

$$
C(\boldsymbol{\tau}) \geq R(\boldsymbol{\tau}) = \sum_{i=1}^{N} \alpha_i R(\boldsymbol{\tau}(\mathbf{n}_i)) \tag{5.19}
$$

The proof of Theorem 5.2 can be found in Section 5.5. The theorem characterizes an achievable rate for the corner points $\boldsymbol{\tau}(\mathbf{n})$ corresponding to any monotone non-decreasing sequence $\mathbf{n} = \{n_i\}_{i=0}^{M-1} \subset \{1, \cdots, N\}^M$. For any other traffic ratio vector $\boldsymbol{\tau}$, the achievability scheme is obtained by time-sharing between the nearest corner points. We note that due to the large number of corner points, we do not provide an explicit achievable rate for each corner point but we rather describe the achievable rate by a system of difference equations. The solution of this system of difference equations specifies the traffic ratio vector $\boldsymbol{\tau}(\mathbf{n})$ and the achievable rate $R(\boldsymbol{\tau}(\mathbf{n}))$ corresponding to the monotone non-decreasing sequence $\{n_i\}_{i=0}^{M-1}$. We have the following remarks.

Remark 5.5 *If $n_i = N$ for all $i \in \{0, \cdots, M-1\}$, then $\mathcal{S} = \{0\}$ and the number of stages of k-sums is described by the following difference equation for any database*

$$y[k] = (N-1)y[k-1] \tag{5.20}$$

with initial condition of $y[1] = 1$. In this case $\tau_n = \frac{1}{N}$ for all n, and $R = \frac{1}{1+\frac{1}{N}+\cdots+\frac{1}{N^{M-1}}}$, i.e., the scheme in Theorem 5.2 reduces to the symmetric scheme in [12] if the sequence $\mathbf{n} = (N, N, \cdots, N)$ is used.

Remark 5.6 *We note that the sequence $\{n_i\}_{i=0}^{M-1}$ suffices to completely specify the traffic ratio vector $\boldsymbol{\tau}(\mathbf{n})$ for every corner point as a consequence of the monotonicity*

of the sequence, i.e.,

$$\{n_i\}_{i=0}^{M-1} \Rightarrow (\underbrace{\tilde{\tau}_0, \cdots, \tilde{\tau}_0}_{n_0 \text{ elements}}, \underbrace{\tilde{\tau}_1, \cdots, \tilde{\tau}_1}_{(n_1-n_0) \text{ elements}}, \cdots, \underbrace{\tilde{\tau}_{M-1}, \cdots, \tilde{\tau}_{M-1}}_{(n_{M-1}-n_{M-2}) \text{ elements}}) \tag{5.21}$$

where $\tilde{\tau}_j = \frac{\sum_{k=1}^{M} \binom{M}{k} y_j[k]}{\sum_{\ell \in \mathcal{S}} \sum_{k=1}^{M} \binom{M}{k} y_j[k](n_\ell - n_{\ell-1})}$.

Remark 5.7 *For fixed M, N, the number of corner points in Theorem 5.2 corresponds to the number of monotone non-decreasing sequences $\mathbf{n} = \{n_i\}_{i=0}^{M-1}$, which is $\binom{M+N-1}{M}$.*

The next corollary asserts that the achievable scheme in Theorem 5.2 is optimal for $M = 2$ and $M = 3$ messages for any traffic ratio vector $\boldsymbol{\tau}$ and any number of databases N.

Corollary 5.2 (Capacity for $M = 2$ and $M = 3$ messages) *For the PIR problem with asymmetric traffic constraints $\boldsymbol{\tau}$, the capacity $C(\boldsymbol{\tau})$ for $M = 2$ and $M = 3$, and for any arbitrary N is given by:*

$$C(\boldsymbol{\tau}) = \begin{cases} \min_{n_0 \in \{1,\cdots,N\}} \frac{1+\frac{1}{n_0}\sum_{n=n_0+1}^{N} \tau_n}{1+\frac{1}{n_0}}, & M = 2 \\ \min_{n_0 \le n_1 \in \{1,\cdots,N\}} \frac{1+\frac{1}{n_0}\sum_{n=n_0+1}^{N} \tau_n + \frac{1}{n_0 n_1}\sum_{n=n_1+1}^{N} \tau_n}{1+\frac{1}{n_0}+\frac{1}{n_0 n_1}}, & M = 3 \end{cases} \tag{5.22}$$

The proof of Corollary 5.2 is given in Section 5.6.

Fig. 5.2 shows the PIR capacity under asymmetric constraints $C(\lambda_2)$ as a function of λ_2 (which is bijective to $\boldsymbol{\tau}$) for the case of $M = 3$ messages and $N = 2$ databases. We note that the capacity $C(\lambda_2)$ is a piece-wise monotone curve in λ_2, which consists of $\binom{M+N-2}{M-1} = 3$ regimes. There exist $\binom{M+N-1}{M} = 4$ corner points.

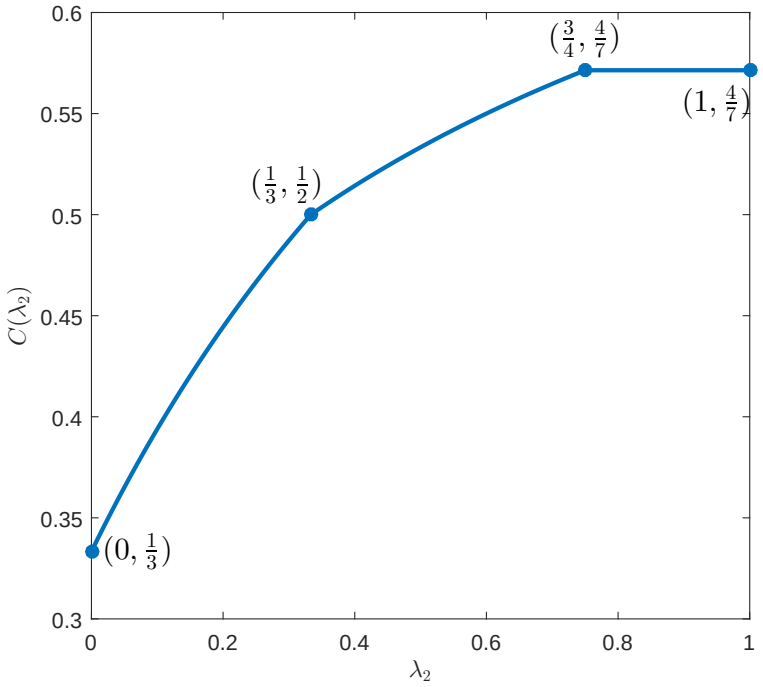


Figure 5.2: Capacity function $C(\lambda_2)$ for $M = 3$, $N = 2$.

Specific achievable schemes for the case of $M = 3$ and $N = 2$ are given in Section 5.5.1. Each corner point shown in Fig. 5.2 corresponds to an explicit achievable scheme given in Section 5.5.1.1. For any other point, time-sharing between corner points is used to achieve these points as shown in Section 5.5.1.2.

Fig. 5.3 shows the capacity region $C(\lambda_2, \lambda_3)$ for the case of $M = 3$ messages and $N = 3$ databases as a function of the pair (λ_2, λ_3) (which is bijective to $\boldsymbol{\tau}$). Fig. 5.3 shows that there exist $\binom{M+N-1}{M} = 10$ corner points, and $\binom{M+N-2}{M-1} = 6$ regions. We show the capacity regions in terms of the triple $(\lambda_2, \lambda_3, C(\lambda_2, \lambda_3))$. Furthermore, for every region we show the corresponding (n_0, n_1) to be plugged in (5.22). The capacity for any point (λ_2, λ_3) other than the corner points is obtained by time-sharing between the corner points that enclose (λ_2, λ_3). Specific achievable

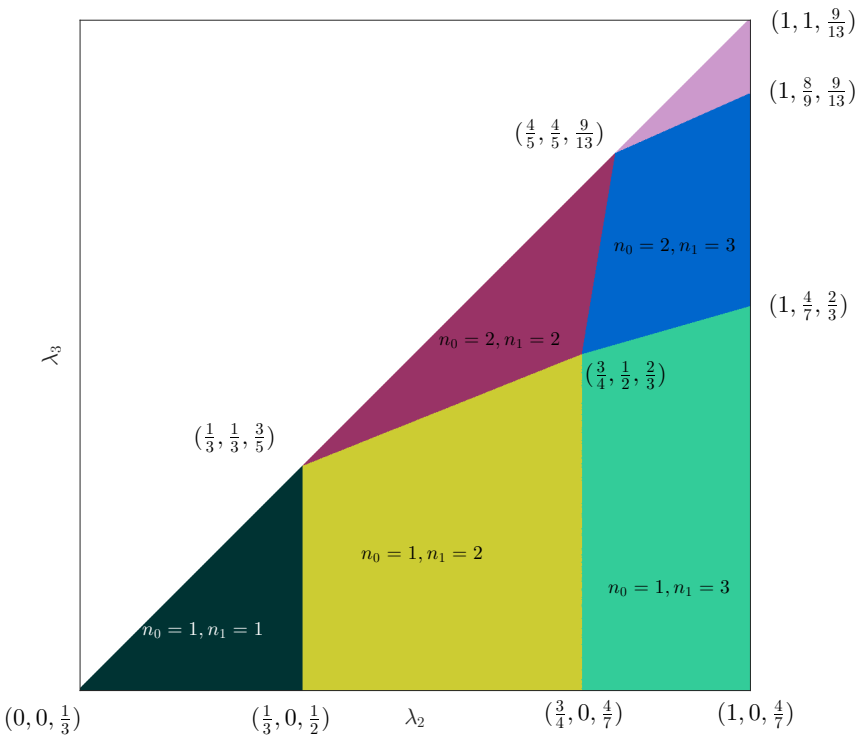


Figure 5.3: Illustration of corner points and regions of $C(\lambda_2, \lambda_3)$ for $M = 3$, $N = 3$.

schemes for $M = 3$, $N = 3$ are given in Section 5.8.2.

Finally, in the following corollary, we specialize the achievable scheme in Theorem 5.2 to the case of $N = 2$ for any arbitrary M.

Corollary 5.3 (Achievable traffic versus retrieval rate tradeoff for $N = 2$) *For the PIR problem with $N = 2$ and an arbitrary M under asymmetric traffic constraints $\boldsymbol{\tau} = (1 - \tau_2, \tau_2)$, $\tau_2 \leq \frac{1}{2}$, let $s_2 \in \{1, \cdots, M-1\}$, for the traffic ratio $\tau_2(s_2)$, where*

$$\tau_2(s_2) = \frac{\sum_{i=0}^{\lfloor \frac{M-s_2-1}{2} \rfloor} \binom{M}{s_2+2i+1}}{M\binom{M-2}{s_2-1} + \sum_{i=0}^{M-s_2-1} \binom{M}{s_2+1+i}} \tag{5.23}$$

the PIR capacity $C(\tau_2(s_2))$ is lower bounded by:

$$C(\tau_2(s_2)) \geq R(\tau_2(s_2)) = \frac{\binom{M-2}{s_2-1} + \sum_{i=0}^{M-s_2-1} \binom{M-1}{s_2+i}}{M\binom{M-2}{s_2-1} + \sum_{i=0}^{M-s_2-1} \binom{M}{s_2+1+i}} \tag{5.24}$$

Moreover, if $\tau_2(s_2) < \tau_2 < \tau_2(s_2+1)$, and $\alpha \in (0,1)$, such that $\tau_2 = \alpha\tau_2(s_2) + (1-\alpha)\tau_2(s_2+1)$, then

$$C(\tau_2) \geq R(\tau_2) = \alpha R(\tau_2(s_2)) + (1-\alpha)R(\tau_2(s_2+1)) \tag{5.25}$$

The proof of Corollary 5.3 is given in Section 5.7.

Remark 5.8 *Fig. 5.4 shows the tradeoff between the traffic ratio τ_2 and the achievable retrieval rate $R(\tau_2)$. We note that as M increases $R(\tau_2)$ decreases pointwise. We observe that as $M \to \infty$, the rate-traffic tradeoff converges to $R(\tau_2) = \tau_2$. This implies that for large enough M, our achievable scheme reduces to time-sharing between the trivial achievable scheme of downloading all the messages from database 1 which achieves a rate of $\frac{1}{M}$, and the asymptotically-optimal achievable scheme in [6] which achieves $R = 1 - \frac{1}{N}$.*

5.4 Converse Proof

In this section, we derive an upper bound for the PIR problem with asymmetric traffic constraints. We extend the converse techniques introduced in [12] to account for the asymmetry of the returned answer strings.

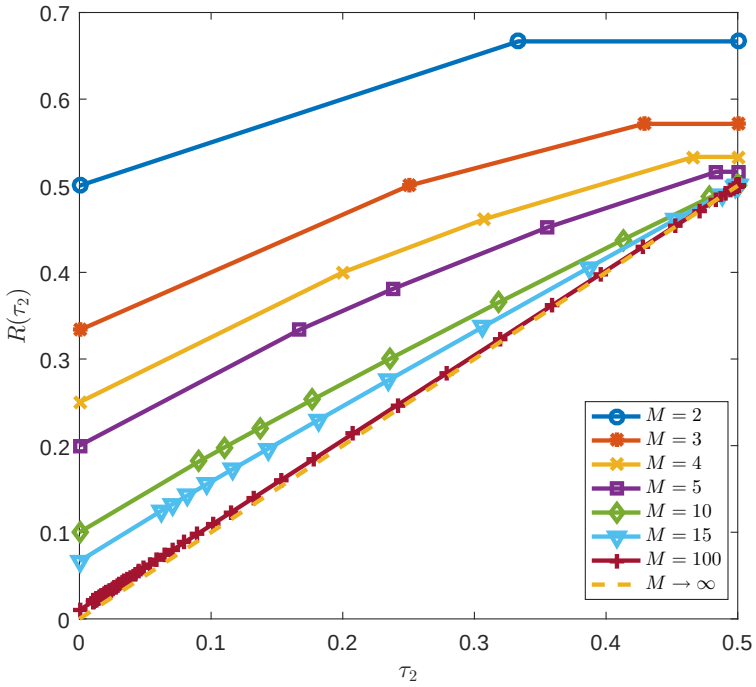


Figure 5.4: Achievable rate-traffic ratio tradeoff for $N = 2$.

We first need the following lemma, which characterizes a fundamental lower bound on the interference from the undesired messages within the answer strings, i.e., a lower bound on $\sum_{n=1}^{N} t_n - L$, as a consequence of the privacy constraint. The proof of this lemma can be found in [12, Lemma 5]. The proof follows for our case since the privacy constraint does not change in the PIR with asymmetric traffic constraints, and the fact that the proof in [12, Lemma 5] deals with the length of the entire downloaded answer strings $A_{1:N}^{[1]}$ and not the individual lengths of each answer string, see [12, equations (46)-(47)].

Lemma 5.1 (Interference lower bound) *For the PIR problem under asymmetric traffic constraints* $\{t_n\}_{n=1}^{N}$*, the interference from undesired messages within the*

answer strings $\sum_{n=1}^{N} t_n - L$ *is lower bounded as,*

$$\sum_{n=1}^{N} t_n - L + o(L) \geq I\left(W_{2:M}; Q_{1:N}^{[1]}, A_{1:N}^{[1]} | W_1\right) \tag{5.26}$$

In the following lemma, we prove an inductive relation for the mutual information term on the right hand side of (5.26). In this lemma, the interference lower bound in (5.26) is expanded into two parts. The first part, which contains the answer strings from the first n_{m-1} databases $A_{1:n_{m-1}}^{[m]}$, is dealt with as in the proof of [12, Lemma 6]. For the second part, which contains the remaining answer strings $A_{n_{m-1}+1:N}^{[m]}$, each answer string $A_n^{[m]}$ is bounded trivially by the length of the answer string t_n.

Lemma 5.2 (Induction lemma) *For all* $m \in \{2, \ldots, M\}$ *and for an arbitrary* $n_{m-1} \in \{1, \cdots, N\}$*, the mutual information term in Lemma 5.1 can be inductively lower bounded as,*

$$I\left(W_{m:M}; Q_{1:N}^{[m-1]}, A_{1:N}^{[m-1]} | W_{1:m-1}\right)$$
$$\geq \frac{1}{n_{m-1}} I\left(W_{m+1:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m}\right) + \frac{1}{n_{m-1}}\left(L - t_1 \sum_{n=n_{m-1}+1}^{N} \lambda_n\right) - \frac{o(L)}{n_{m-1}} \tag{5.27}$$

We note that [12, Lemma 6] can be interpreted as a special case of Lemma 5.2 with setting $n_{m-1} = N$. Intuitively, n_{m-1} represents the number of databases that can apply the optimal symmetric scheme in [12] if the user wants to retrieve message W_{m-1} from the set of $W_{m-1:M}$ messages (i.e., conditioned on $W_{1:m-1}$).

Proof: We start with the left hand side of (5.27) after multiplying by n_{m-1},

$$
\begin{aligned}
& n_{m-1}\, I\left(W_{m:M}; Q_{1:N}^{[m-1]}, A_{1:N}^{[m-1]} | W_{1:m-1}\right) \\
& \quad \geq n_{m-1}\, I\left(W_{m:M}; Q_{1:n_{m-1}}^{[m-1]}, A_{1:n_{m-1}}^{[m-1]} | W_{1:m-1}\right) && (5.28) \\
& \quad \geq \sum_{n=1}^{n_{m-1}} I\left(W_{m:M}; Q_{n}^{[m-1]}, A_{n}^{[m-1]} | W_{1:m-1}\right) && (5.29) \\
& \quad \overset{(5.7)}{=} \sum_{n=1}^{n_{m-1}} I\left(W_{m:M}; Q_{n}^{[m]}, A_{n}^{[m]} | W_{1:m-1}\right) && (5.30) \\
& \quad \overset{(5.3)}{=} \sum_{n=1}^{n_{m-1}} I\left(W_{m:M}; A_{n}^{[m]} | Q_{n}^{[m]}, W_{1:m-1}\right) && (5.31) \\
& \quad \overset{(5.4)}{=} \sum_{n=1}^{n_{m-1}} H\left(A_{n}^{[m]} | Q_{n}^{[m]}, W_{1:m-1}\right) && (5.32) \\
& \quad \geq \sum_{n=1}^{n_{m-1}} H\left(A_{n}^{[m]} | A_{1:n-1}^{[m]}, Q_{1:n_{m-1}}^{[m]}, W_{1:m-1}\right) && (5.33) \\
& \quad \overset{(5.4)}{=} \sum_{n=1}^{n_{m-1}} I\left(W_{m:M}; A_{n}^{[m]} | A_{1:n-1}^{[m]}, Q_{1:n_{m-1}}^{[m]}, W_{1:m-1}\right) && (5.34) \\
& \quad = I\left(W_{m:M}; A_{1:n_{m-1}}^{[m]} | Q_{1:n_{m-1}}^{[m]}, W_{1:m-1}\right) && (5.35) \\
& \quad \overset{(5.3)}{=} I\left(W_{m:M}; Q_{1:n_{m-1}}^{[m]}, A_{1:n_{m-1}}^{[m]} | W_{1:m-1}\right) && (5.36) \\
& \quad \overset{(5.3),(5.4)}{=} I\left(W_{m:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m-1}\right) \\
& \qquad - I\left(W_{m:M}; A_{n_{m-1}+1:N}^{[m]} | Q_{1:N}^{[m]}, A_{1:n_{m-1}}^{[m]}, W_{1:m-1}\right) && (5.37) \\
& \quad \overset{(5.4)}{=} I\left(W_{m:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m-1}\right) - H\left(A_{n_{m-1}+1:N}^{[m]} | Q_{1:N}^{[m]}, A_{1:n_{m-1}}^{[m]}, W_{1:m-1}\right) \\
& && (5.38) \\
& \quad \overset{(5.5)}{\geq} I\left(W_{m:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m-1}\right) - t_1 \sum_{n=n_{m-1}+1}^{N} \lambda_n && (5.39) \\
& \quad \overset{(5.8)}{=} I\left(W_{m:M}; W_m, Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m-1}\right) - t_1 \sum_{n=n_{m-1}+1}^{N} \lambda_n - o(L) && (5.40)
\end{aligned}
$$

$$= I\left(W_{m:M}; W_m | W_{1:m-1}\right) + I\left(W_{m:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m}\right) - t_1 \sum_{n=n_{m-1}+1}^{N} \lambda_n - o(L) \tag{5.41}$$

$$= I\left(W_{m+1:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m}\right) + \left(L - t_1 \sum_{n=n_{m-1}+1}^{N} \lambda_n\right) - o(L) \tag{5.42}$$

where (5.28), (5.29) follow from the non-negativity of mutual information, (5.30) follows from the privacy constraint, (5.31) follows from the independence of $\left(W_{m:M}, Q_n^{[m]}\right)$, (5.32),(5.34) follow from the fact that the answer string $A_n^{[m]}$ is a deterministic function of $\left(Q_n^{[m]}, W_{1:M}\right)$, (5.33) follows from conditioning reduces entropy, (5.36) follows from the independence of $\left(W_{m:M}, Q_{1:n_{m-1}}^{[m]}\right)$, (5.37) follows from the chain rule, the independence of the queries and the messages, and the fact that $Q_{1:N}^{[m]} \to Q_{1:n_{m-1}}^{[m]} \to A_{1:n_{m-1}}^{[m]}$ forms a Markov chain by (5.4), (5.38) follows from the fact that the answer strings $A_{1:n_{m-1}}^{[m]}$ are fully determined from $\left(Q_{1:N}^{[m]}, W_{1:M}\right)$, (5.39) follows from the fact that conditioning reduces entropy and $H(A_{n_{m-1}+1:N}) \leq \sum_{n=n_{m-1}+1}^{N} t_n$ which is equal to $t_1 \sum_{n=n_{m-1}+1}^{N} \lambda_n$ from the asymmetric traffic constraints, (5.40) follows from the reliability constraint. Finally, dividing both sides by n_{m-1} leads to (5.27). ■

Now, we are ready to derive an explicit upper bound for the retrieval rate under asymmetric traffic constraints. Applying Lemma 5.1 and Lemma 5.2 successively for an arbitrary sequence $\{n_i\}_{i=1}^{M-1} \subset \{1, \cdots, N\}^{M-1}$ and observing that $\sum_{n=1}^{N} t_n = t_1 \sum_{n=1}^{N} \lambda_n$ under the asymmetric traffic constraints, we have the following

$$t_1 \sum_{n=1}^{N} \lambda_n - L + \tilde{o}(L)$$

$$\overset{(5.26)}{\geq} I\left(W_{2:M}; Q^{[1]}_{1:N}, A^{[1]}_{1:N} | W_1\right) \tag{5.43}$$

$$\overset{(5.27)}{\geq} \frac{1}{n_1}\left(L - t_1 \sum_{n=n_1+1}^{N} \lambda_n\right) + \frac{1}{n_1} I\left(W_{3:M}; Q^{[2]}_{1:N}, A^{[2]}_{1:N} | W_{1:2}\right) \tag{5.44}$$

$$\overset{(5.27)}{\geq} \frac{1}{n_1}\left(L - t_1 \sum_{n=n_1+1}^{N} \lambda_n\right) + \frac{1}{n_1 n_2}\left(L - t_1 \sum_{n=n_2+1}^{N} \lambda_n\right) + \frac{1}{n_2} I\left(W_{4:M}; Q^{[3]}_{1:N}, A^{[3]}_{1:N} | W_{1:3}\right) \tag{5.45}$$

$$\overset{(5.27)}{\geq} \cdots$$

$$\overset{(5.27)}{\geq} \frac{1}{n_1}\left(L - t_1 \sum_{n=n_1+1}^{N} \lambda_n\right) + \frac{1}{n_1 n_2}\left(L - t_1 \sum_{n=n_2+1}^{N} \lambda_n\right) + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}\left(L - t_1 \sum_{n=n_{M-1}+1}^{N} \lambda_n\right) \tag{5.46}$$

where $\tilde{o}(L) = \left(1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}\right) o(L)$, (5.43) follows from Lemma 5.1, and the remaining bounding steps follow from successive application of Lemma 5.2.

Ordering terms, we have,

$$\left(1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}\right) L \leq \left(1 + \frac{\gamma(n_1)}{n_1} + \cdots + \frac{\gamma(n_{M-1})}{\prod_{i=1}^{M-1} n_i}\right) t_1 \sum_{n=1}^{N} \lambda_n + \tilde{o}(L) \tag{5.47}$$

where $\gamma(\ell) = \frac{\sum_{n=\ell+1}^{N} \lambda_n}{\sum_{n=1}^{N} \lambda_n} = \sum_{n=\ell+1}^{N} \tau_n$ corresponds to the sum of the traffic ratios from databases $[\ell + 1 : N]$.

We conclude the proof by taking $L \to \infty$. Thus, for an arbitrary sequence

$\{n_i\}_{i=1}^{M-1}$, we have

$$R(\boldsymbol{\tau}) = \frac{L}{t_1 \sum_{n=1}^{N} \lambda_n} \leq \frac{1 + \frac{\gamma(n_1)}{n_1} + \frac{\gamma(n_2)}{n_1 n_2} + \cdots + \frac{\gamma(n_{M-1})}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{5.48}$$

Finally, we get the tightest bound by minimizing over the sequence $\{n_i\}_{i=1}^{M-1}$ over the set $\{1, \cdots, N\}$, as

$$R(\boldsymbol{\tau}) \leq \min_{n_1, \cdots, n_{M-1} \in \{1, \cdots, N\}} \frac{1 + \frac{\gamma(n_1)}{n_1} + \frac{\gamma(n_2)}{n_1 n_2} + \cdots + \frac{\gamma(n_{M-1})}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{5.49}$$

$$= \min_{n_1, \cdots, n_{M-1} \in \{1, \cdots, N\}} \frac{1 + \frac{\sum_{n=n_1+1}^{N} \tau_n}{n_1} + \frac{\sum_{n=n_2+1}^{N} \tau_n}{n_1 n_2} + \cdots + \frac{\sum_{n=n_{M-1}+1}^{N} \tau_n}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{5.50}$$

Remark 5.9 *From the converse proof, we note that we can intuitively interpret n_i as the number of databases that can apply the symmetric traffic scheme in [12] if the number of messages is reduced to be $M - i + 1$. We point out that in the absence of asymmetric traffic constraints as in [12], all databases can apply symmetric schemes, therefore $n_i = N$ for all $i \in \{1, \cdots, M-1\}$. Now, since reducing the number of messages cannot decrease the number of databases that apply the symmetric scheme as the problem would be less constrained (in terms of the privacy constraint), which leads to more flexibility in terms of satisfying the traffic constraints, it suffices to evaluate the bound in* (5.10) *for monotone non-decreasing sequences $\{n_i\}_{i=1}^{M-1} \subset \{1, \cdots, N\}^{M-1}$ such that $n_1 \leq n_2 \leq \cdots \leq n_{M-1}$.*

5.5 Achievability Proof

The achievability scheme for the PIR problem under asymmetric traffic constraints is inspired by the PIR schemes in [12, 29]. Our achievable scheme applies message symmetry, and side information exploitation as in [12, 29]. However, due to the asymmetric traffic constraints, database symmetry cannot be applied. In an alternative view, we use the side information in an asymmetric fashion among the databases. The most relevant achievable scheme to our achievable scheme here is the scheme in [29], in which the bits stored in the user's cache is exploited differently depending on the caching ratio. We begin the discussion by presenting the $M = 3$, $N = 2$ case as a concrete example to illustrate the main concepts of the scheme.

5.5.1 Motivating Example: $M = 3$ Messages, $N = 2$ Databases

In this section, we show the achievability scheme for $M = 3$, $N = 2$. We first carry out the minimization in (5.10) over $n_1, n_2 \in \{1, 2\}$. In this case, we have 4 upper bounds (or effectively 3 bounds if $n_1 \leq n_2$ restriction is applied). By taking the minimum of these bounds for every $\lambda_2 \in [0, 1]$, we have the following explicit upper bound on the capacity as a function of λ_2 (which is in bijection to τ_2)

$$C(\lambda_2) \leq \begin{cases} \frac{1+3\lambda_2}{3(1+\lambda_2)}, & 0 \leq \lambda_2 \leq \frac{1}{3} \\ \frac{2(1+2\lambda_2)}{5(1+\lambda_2)}, & \frac{1}{3} \leq \lambda_2 \leq \frac{3}{4} \\ \frac{4}{7}, & \frac{3}{4} \leq \lambda_2 \leq 1 \end{cases} \tag{5.51}$$

To show the achievability of the upper bound in (5.51), let a_i, b_i, c_i denote randomly and independently permuted symbols of messages W_1, W_2, W_3, respectively. Define $s_2 \in \{0, 1, 2\}$ to be the number of side information symbols that are used simultaneously in database 2 within the initial round of downloads. First, we show the achievability of the corner points, i.e., the achievability of the points corresponding to $\lambda_2 \in \{0, \frac{1}{3}, \frac{3}{4}, 1\}$.

5.5.1.1 Achievability of the Corner Points

The $\lambda_2 = 0$ Corner Point: $\lambda_2 = 0$ means that the second database does not return any answer strings. The optimal achievable scheme is to download all files from the first database (see Table 5.1). This scheme achieves $R = \frac{1}{3} = C(0)$.

Table 5.1: The query table for $M = 3$, $N = 2$, $\lambda_2 = 0$.

Database 1	Database 2
a_1, b_1, c_1	

The $\lambda_2 = 1$ Corner Point: Since $\lambda_1 = 1$ by definition, $\lambda_2 = 1$ means that a symmetric scheme can be applied to both databases. Thus, the optimal achievable scheme is the optimal symmetric scheme in [12] (see Table 5.2). We present the scheme here for completeness. In this scheme, the user starts with downloading the individual symbols a_1, b_1, c_1 from database 1. Since $\lambda_2 = 1$, database symmetry can be applied, hence the user downloads a_2, b_2, c_2 from database 2. Note that in this case, the user does not exploit the side information generated from database 1 in the first round of downloads, but rather downloads new individual symbols, hence

$s_2 = 0$ in this case. The undesired symbols $b_i, c_i,\ i = 1, 2$ can be exploited in the other database. This can be done by downloading $a_3 + b_2$, $a_4 + c_2$ from database 1, and similarly by applying database symmetry, the user downloads $a_5 + b_1$, $a_6 + c_1$ from database 2. In order to satisfy the privacy constraint, the user applies the message symmetry and downloads $b_3 + c_3$ from database 1, and $b_4 + c_4$ from database 2. Finally, the user exploits the newly generated side information by downloading $a_7 + b_4 + c_4$ from database 1, and $a_8 + b_3 + c_3$ from database 2. Consequently, the user downloads $L = 8$ symbols in 14 downloads which results in $R = \frac{8}{14} = \frac{4}{7} = C(1)$.

Table 5.2: The query table for $M = 3$, $N = 2$, $\lambda_2 = 1$.

Database 1	Database 2
a_1, b_1, c_1	a_2, b_2, c_2
$a_3 + b_2$	$a_5 + b_1$
$a_4 + c_2$	$a_6 + c_1$
$b_3 + c_3$	$b_4 + c_4$
$a_7 + b_4 + c_4$	$a_8 + b_3 + c_3$

The $\lambda_2 = \frac{3}{4}$ Corner Point: The user can cut the first round of downloads in database 2 and exploit the side information generated from database 1 directly in the form of sums of 2, i.e., the user downloads a_1, b_1, c_1 from database 1 and then exploits the undesired symbols as side information by downloading $a_2 + b_1$, $a_3 + c_1$ from database 2. The user then applies message symmetry and downloads $b_2 + c_2$. Since the user uses 1 bit of side information in the initial download round from database 2, $s_2 = 1$ in this case. Finally, the user exploits the undesired sum $b_2 + c_2$ from database 2 as a side information in database 1 and downloads $a_4 + b_2 + c_2$. Using this scheme the user downloads 4 symbols from database 1 and 3 symbols from database 2,

hence $\lambda_2 = \frac{3}{4}$. The user downloads $L = 4$ desired symbols out of 7 downloads, thus $R = \frac{4}{7} = C(\frac{3}{4})$. The privacy is satisfied since W_1, W_2, W_3 are independently and randomly permuted, and since the scheme includes all the possible combinations of the sums in any round. The query table for this scheme is given in Table 5.3. We note that this scheme is exactly the asymmetric achievable scheme presented in [18].

Table 5.3: The query table for $M = 3$, $N = 2$, $\lambda_2 = \frac{3}{4}$.

Database 1	Database 2
a_1, b_1, c_1	
	$a_2 + b_1$ $a_3 + c_1$ $b_2 + c_2$
$a_4 + b_2 + c_2$	

The $\lambda_2 = \frac{1}{3}$ Corner Point: In this case, the user downloads a_1, b_1, c_1 from database 1. In database 2, the user exploits the side information b_1, c_1 simultaneously and downloads $a_2 + b_1 + c_1$. Due to the fact that 2 side information symbols are used simultaneously in the initial round of download from database 2, $s_2 = 2$ in this case. Using this scheme the user downloads 3 symbols from database 1 and 1 symbol from database 2, therefore $\lambda_2 = \frac{1}{3}$. The user downloads $L = 2$ desired symbols in 4 downloads, hence $R = \frac{1}{2} = C(\frac{1}{3})$. The privacy follows by the same argument as in the previous case. The query table for this case is given in Table 5.4.

Table 5.4: The query table for $M = 3$, $N = 2$, $\lambda_2 = \frac{1}{3}$.

Database 1	Database 2
a_1, b_1, c_1	
	$a_2 + b_1 + c_1$

5.5.1.2 Achievability of Non-Corner Points

In the following, we show that by combining the achievable schemes of the corner points over different symbols, the upper bound in (5.51) is tight for any λ_2. We note that the privacy constraint is still satisfied after this combination as each scheme operates over different sets of symbols and the fact that each scheme satisfies the privacy constraint individually. A formal argument for proving that combination of private schemes remains private can be found in [18, Theorem 4]. Let ν_{s_2}, where $s_2 = 0, 1, 2$, denote the number of repetitions of the scheme that uses s_2 side information symbols simultaneously in the first round of download in database 2. By convention, let ν_3 denote the number of repetitions of the trivial retrieval scheme, i.e., when the retrieval is solely done from database 1.

The $0 \leq \lambda_2 \leq \frac{1}{3}$ Regime: We combine the achievable scheme of $\lambda_2 = 0$ corner point with the achievable scheme of $\lambda_2 = \frac{1}{3}$ corner point. The achievable scheme of $\lambda_2 = 0$ downloads 3 symbols from database 1 and 0 symbols from database 2. We perform this scheme ν_3 repetitions. The achievable scheme of $\lambda_2 = \frac{1}{3}$ downloads 3 symbols from database 1 and 1 symbol from database 2. We perform this scheme ν_2 repetitions. Under the asymmetric traffic constraints, this results in the following system of equations

$$3\nu_3 + 3\nu_2 = t_1 \tag{5.52}$$

$$\nu_2 = \lambda_2 t_1 \tag{5.53}$$

This system has a unique solution (parametrized by t_1) of $\nu_2 = \lambda_2 t_1$ and $\nu_3 = \frac{1-3\lambda_2}{3} t_1$. Note that $\nu_3 \geq 0$ in the regime of $0 \leq \lambda_2 \leq \frac{1}{3}$. Since the scheme of $\lambda_2 = 0$ downloads 1 symbol from the desired message and the scheme of $\lambda_2 = \frac{1}{3}$ downloads 2 symbols from the desired message. The achievable rate $R(\lambda_2)$ is given by

$$R(\lambda_2) = \frac{2\nu_2 + \nu_3}{(1+\lambda_2)t_1} = \frac{1+3\lambda_2}{3(1+\lambda_2)} = C(\lambda_2), \qquad 0 \leq \lambda_2 \leq \frac{1}{3} \tag{5.54}$$

The $\frac{1}{3} \leq \lambda_2 \leq \frac{3}{4}$ Regime: Similarly, the user combines the achievable schemes of $\lambda_2 = \frac{1}{3}$ and $\lambda_2 = \frac{3}{4}$ corner points. The user applies the scheme of $\lambda_2 = \frac{1}{3}$ for ν_2 repetitions, which downloads 3 symbols from database 1 and 1 symbol from database 2 and has $L = 2$. The user applies the scheme of $\lambda_2 = \frac{3}{4}$ for ν_1 repetitions, which downloads 4 symbols from database 1 and 3 symbols from database 2 and has $L = 4$. This results in the following system of equations

$$4\nu_1 + 3\nu_2 = t_1 \tag{5.55}$$

$$3\nu_1 + \nu_2 = \lambda_2 t_1 \tag{5.56}$$

which has the following solution: $\nu_1 = \frac{-1+3\lambda_2}{5} t_1 \geq 0$ and $\nu_2 = \frac{3-4\lambda_2}{5} t_1 \geq 0$ in the regime of $\frac{1}{3} \leq \lambda_2 \leq \frac{3}{4}$. Consequently, the achievable rate is given by

$$R(\lambda_2) = \frac{4\nu_1 + 2\nu_2}{(1+\lambda_2)t_1} = \frac{2(1+2\lambda_2)}{5(1+\lambda_2)} = C(\lambda_2), \qquad \frac{1}{3} \leq \lambda_2 \leq \frac{3}{4} \tag{5.57}$$

The $\frac{3}{4} \leq \lambda_2 \leq 1$ Regime: The user combines the achievable schemes of $\lambda_2 = \frac{3}{4}$ and $\lambda_2 = 1$ corner points. The user repeats the scheme of $\lambda_2 = \frac{3}{4}$ for ν_1 repetitions, and the scheme of $\lambda_2 = 1$ for ν_0 repetitions. This results in the following system of equations

$$4\nu_1 + 7\nu_0 = t_1 \tag{5.58}$$

$$3\nu_1 + 7\nu_0 = \lambda_2 t_1 \tag{5.59}$$

The solution for this system is given by: $\nu_1 = (1 - \lambda_2)t_1 \geq 0$ and $\nu_0 = \frac{-3+4\lambda_2}{7} t_1 \geq 0$ in the regime of $\frac{3}{4} \leq \lambda_2 \leq 1$. The corresponding rate is given by

$$R(\lambda_2) = \frac{4\nu_1 + 8\nu_0}{(1 + \lambda_2)t_1} = \frac{4}{7} = C(\lambda_2), \qquad \frac{3}{4} \leq \lambda_2 \leq 1 \tag{5.60}$$

Specific Example for Non-Corner Points, $\lambda_2 = \frac{1}{2}$: The query table for this case is given in Table 5.5. The user applies the scheme of $\lambda_2 = \frac{3}{4}$ for $\nu_1 = \frac{-1+3\lambda_2}{5} t_1 = \frac{1}{10} t_1$ repetitions, and the scheme of $\lambda_2 = \frac{1}{3}$ for $\nu_2 = \frac{3-4\lambda_2}{5} t_1 = \frac{1}{5} t_1$ repetitions. Choosing $t_1 = 10$, we have $\nu_1 = 1$ and $\nu_2 = 2$. The scheme downloads 10 symbols from database 1 and 5 symbols from database 2, thus, $\lambda_2 = \frac{1}{2}$. The scheme downloads 8 symbols in 15 downloads, hence $R(\frac{1}{2}) = \frac{8}{15} = \frac{2(1+2\lambda_2)}{5(1+\lambda_2)} = C(\frac{1}{2})$.

5.5.2 Description of the General Scheme

In this section, we describe the general achievable scheme that achieves the retrieval rates in Theorem 5.2. We first show explicitly the achievability schemes for corner

Table 5.5: The query table for $M = 3$, $N = 2$, $\lambda_2 = \frac{1}{2}$.

Database 1	Database 2
a_1, b_1, c_1	$a_2 + b_1$ $a_3 + c_1$ $b_2 + c_2$
$a_4 + b_2 + c_2$	
a_5, b_3, c_3	$a_6 + b_3 + c_3$
a_7, b_4, c_4	$a_8 + b_4 + c_4$

points, i.e., the achievability scheme for every monotone non-decreasing sequence $\{n_i\}_{i=0}^{M-1} \subset \{1, \cdots, N\}^M$. We note that our achievability scheme is different in two key steps: First regarding the database symmetry, we note that it is not applied over all databases directly as in [12], but rather it is applied over groups of databases, such as, group 0 includes databases 1 through n_0, group 1 includes databases $n_0 + 1$ through n_1, etc. Second, regarding the exploitation of side information step, we note that each group of databases exploits side information differently in the *initial* round of downloading. More specifically, we note that group 0 of databases do not exploit any side information in the initial round of the download, group 1 exploits 1 side information symbol in the initial round of the download, group 2 exploits sums of 2 side information symbols in the initial round of the download, and so on.

Next, we show that for non-corner points, time-sharing between corner points is achievable and this concludes the achievability proof of Theorem 5.2.

5.5.2.1 Achievability Scheme for the Corner Points

Let $s_n \in \{0, 1, \cdots, M-1\}$ denote the number of side information symbols that are used simultaneously in the initial round of downloads at the nth database.

For a given non-decreasing sequence $\{n_i\}_{i=0}^{M-1} \subset \{1, \cdots, N\}^M$, let $s_n = i$ for all $n_{i-1} + 1 \leq n \leq n_i$ with $n_{-1} = 0$ by convention. Denote $\mathcal{S} = \{i : s_n = i \text{ for some } n \in \{1, \cdots, N\}\}$. We follow the round and stage definitions in [123]. The kth round is the download queries that admit a sum of k different messages (k-sum in [12]). A stage of the kth round is a query block of the kth round that exhausts all $\binom{M}{k}$ combinations of the k-sum. Denote $y_\ell[k]$ to be the number of stages in round k downloaded from the nth database, such that $n_{\ell-1} + 1 \leq n \leq n_\ell$. The details of the achievable scheme are as follows:

1. *Initialization:* The user permutes each message independently and uniformly using a random interleaver, i.e.,

$$x_m(i) = W_m(\pi_m(i)), \quad i \in \{1, \cdots, L\} \tag{5.61}$$

   where $x_m(i)$ is the ith symbol of the permuted W_m, $\pi_m(\cdot)$ is a random interleaver for the mth message that is chosen independently, uniformly, and privately at the user's side. From the nth database where $1 \leq n \leq n_0$, the user downloads $\prod_{s\in\mathcal{S}} \binom{M-2}{s-1}$ symbols from the desired message. The user sets the round index $k = 1$. I.e., the user starts downloading the desired symbols from $y_0[1] = \prod_{s\in\mathcal{S}} \binom{M-2}{s-1}$ different stages.

2. *Message symmetry:* To satisfy the privacy constraint, for each stage initiated in the previous step, the user completes the stage by downloading the remaining $\binom{M-1}{k-1}$ k-sum combinations that do not include the desired symbols, in

particular, if $k = 1$, the user downloads $\prod_{s\in\mathcal{S}} \binom{M-2}{s-1}$ individual symbols from each undesired message.

3. *Database symmetry:* Due to the asymmetric traffic constraints, the original database symmetry step in [12] cannot be applied directly to our problem. Instead, we divide the databases into groups. Group $\ell \in \mathcal{S}$ corresponds to databases $n_{\ell-1} + 1$ to n_ℓ. Database symmetry is applied within each group only. Consequently, the user repeats step 2 over each group of databases, in particular, if $k = 1$, the user downloads $\prod_{s\in\mathcal{S}} \binom{M-2}{s-1}$ individual symbols from each message from the first n_0 databases (group 1).

4. *Exploitation of side information:* This step is also different from [12] because of the asymmetric traffic constraints. In order to create different lengths of the answer strings, the initial exploitation of side information is group-dependent as well. More specifically, the undesired symbols downloaded within the kth round (the k-sums that do not include the desired message) are used as side information in the $(k + 1)$th round. This exploitation of side information is performed by downloading $(k + 1)$-sum consisting of 1 desired symbol and a k-sum of undesired symbols only that were generated in the kth round. However, the main difference from [12] is that for the nth database, if $s_n > k$, then this database does not exploit the side information generated in the kth round. Thus, the nth database belonging to the ℓth group exploits the side information generated in the kth round from all databases except itself if $s_n \leq k$. Moreover, for $s_n = k$, extra side information can be used in the nth

database. This is because the user can form $n_0 \prod_{s\in\mathcal{S}\setminus\{s_n\}} \binom{M-2}{s-1}$ extra stages of side information by constructing k-sums of the undesired symbols in round 1 from the databases in group 0.

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

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

5.5.2.2 Achievability Scheme for Non-Corner Points

In this section, we show that achievability schemes for non-corner points can be derived by time-sharing between the nearest corner points, i.e., the achievable scheme under $\boldsymbol{\tau}$ constraints is performed by time-sharing between the corner points of an N-dimensional polytope that enclose the traffic vector $\boldsymbol{\tau}$. The following lemma formalizes the time-sharing argument. Lemma 5.3 can be thought of as an adaptation of [18, Theorem 4] and [28, Lemma 1] to the PIR problem under asymmetric traffic constraints.

Lemma 5.3 (Time-sharing) *For the PIR problem under asymmetric traffic constraints $\boldsymbol{\tau}$, let the retrieval rate $R(\boldsymbol{\tau}_i)$ be achievable for the traffic ratio vector $\boldsymbol{\tau}_i$ for all $i \in \{1, \cdots, N\}$. Moreover, assume that $\boldsymbol{\tau} = \sum_{i=1}^{N} \alpha_i \boldsymbol{\tau}_i$ for some $\{\alpha_i\}_{i=1}^{N}$ such that $\alpha_i \geq 0$, for all i, and $\sum_{i=1}^{N} \alpha_i = 1$. Then, the following retrieval rate $R(\boldsymbol{\tau})$ is*

achievable,

$$R(\boldsymbol{\tau}) = \sum_{i=1}^{N} \alpha_i R(\boldsymbol{\tau}_i) \tag{5.62}$$

Proof: Let PIR_i denote the PIR scheme that achieves retrieval rate $R(\boldsymbol{\tau}_i)$ for a traffic ratio vector $\boldsymbol{\tau}_i$. Denote the total download of PIR_i by D_i and the corresponding message length by L_i.

Now, construct the following PIR scheme with total download D and message length L. For each database, concatenate the queries from the N PIR schemes with ensuring that each symbol is queried by one PIR scheme only. Hence, $D = \sum_{i=1}^{N} D_i$, such that $D_i = \alpha_i D$, for $i \in \{1, \cdots, N\}$, and the download from the nth database is $t_n(\boldsymbol{\tau}) = \sum_{i=1}^{N} t_n(\boldsymbol{\tau}_i)$. This concatenation of the achievable schemes is feasible under asymmetric traffic constraints since $\boldsymbol{\tau} = \sum_{i=1}^{N} \alpha_i \boldsymbol{\tau}_i$. To see this, we note that the nth element of the traffic ratio vector τ_n is given by

$$\tau_n = \frac{t_n(\boldsymbol{\tau})}{D} = \frac{\sum_{i=1}^{N} t_n(\boldsymbol{\tau}_i)}{D} = \frac{\sum_{i=1}^{N} \tau_n^{(i)} D_i}{D} = \frac{\sum_{i=1}^{N} \tau_n^{(i)} \alpha_i D}{D} = \sum_{i=1}^{N} \alpha_i \tau_n^{(i)} \tag{5.63}$$

where $\tau_n^{(i)}$ denotes the nth element in $\boldsymbol{\tau}_i$. Since these implications are true for each element in $\boldsymbol{\tau}$, we have $\boldsymbol{\tau} = \sum_{i=1}^{N} \alpha_i \boldsymbol{\tau}_i$ as required.

PIR_i scheme downloads L_i symbols from the desired messages, such that

$$L_i = R(\boldsymbol{\tau}_i) D_i = \alpha_i R(\boldsymbol{\tau}_i) D \tag{5.64}$$

Hence, the total message length by concatenating all the achievable schemes together is

$$L = \sum_{i=1}^{N} L_i = \sum_{i=1}^{N} \alpha_i R(\boldsymbol{\tau}_i) D \tag{5.65}$$

and the corresponding achievable rate is given by

$$R(\boldsymbol{\tau}) = \frac{L}{D} = \sum_{i=1}^{N} \alpha_i R(\boldsymbol{\tau}_i) \tag{5.66}$$

The reliability constraint follows from the reliability of each PIR scheme. The privacy constraint is satisfied due to the fact that each PIR scheme operates on a different portion of the messages and these portions are picked uniformly and independently. Hence, the privacy constraint for the concatenated scheme follows from the privacy of each PIR scheme. A formal treatment of the privacy constraint of concatenated schemes can be found in [18]. ■

Thus, Lemma 5.3 provides an achievability proof for any traffic ratio vector $\boldsymbol{\tau}$ that is not a corner point. Finally, we have the following remark regarding this time-sharing lemma.

Remark 5.10 *We note that although the vector* $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_N)$ *is in bijection with* $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_N)$*, the time-sharing argument in Lemma 5.3 does not hold for* $R(\boldsymbol{\lambda})$*. This is due to the fact that* $R(\boldsymbol{\lambda})$ *is a non-linear function of* $\boldsymbol{\lambda}$ *whereas* $R(\boldsymbol{\tau})$ *is an affine function of* $\boldsymbol{\tau}$*.*

5.5.3 Decodability, Privacy, and Calculation of the Achievable Rate

In this section, we prove the decodability, privacy and the achievable rate in Theorem 5.2. We note that it suffices to consider the corner points only, as Lemma 5.3 settles the decodability, privacy and achievable rate for non-corner points based on the existence of feasible PIR schemes that achieve the corner points.

Decodability: By construction, in the $(k+1)$th round at the nth database, the user exploits the side information generated in the kth round in the remaining active databases by adding 1 symbol of the desired message with $(k-1)$-sum of undesired messages which was downloaded previously in the kth round. Moreover, for the nth database belonging to the ℓth group at the $(\ell+1)$th round, the user adds every ℓ symbols of the undesired symbols downloaded from group 0 to make one side information symbol. Since the user downloads $\prod_{\ell \in \mathcal{S}} \binom{M-2}{\ell-1}$ symbols from every database in the first n_0 databases (group 0), the user can exploit such side information to initiate $n_0 \prod_{\ell \in \mathcal{S} \setminus \{\ell\}} \binom{M-2}{\ell-1}$ stages in the $(\ell+1)$th round from every database in group ℓ. Since all side information symbols used in the $(k+1)$th round are decodable in the kth round or from round 1, the user cancels out these side information symbols and is left with symbols from the desired message.

Privacy: For every stage of the kth round initiated in the exploitation of the side information step, the user completes the stage by including all the remaining $\binom{M-1}{k-1}$ undesired symbols. This implies that all $\binom{M}{k}$ combinations of the k-sum are included at each round. Thus, the structure of the queries is the same for any desired

message. The privacy constraint in (5.7) is satisfied by the random and independent permutation of each message and the random shuffling of the order of the queries. This ensures that all queries are equally likely independent of the desired message index.

Calculation of the Achievable Rate: For a corner point characterized by the non-decreasing sequence $\{n_i\}_{i=0}^{M-1}$, as mentioned before, we denote $y_\ell[k]$ to be the number of stages that admit k-sums downloaded from any database belonging to the ℓth group, i.e., the nth database such that $n_{\ell-1}+1 \leq n \leq n_\ell$. By construction, we observe that all databases belonging to the ℓth group are inactive until the $(\ell+1)$th round as the user initiates download in such databases by exploiting ℓ bits of side information simultaneously by definition of the group. Consequently, we have the initial condition $y_\ell[k]=0$ for $k \leq \ell$. Since the user downloads $\prod_{s\in\mathcal{S}} \binom{M-2}{s-1}$ individual symbols (i.e., from round 1) from group 0, we have the following initial condition $y_0[1]=\prod_{s\in\mathcal{S}} \binom{M-2}{s-1}$.

Now, we note from the side information exploitation step that the user initiates new stages in the kth round from the nth database depending on the number of stages of the $(k-1)$th round for group 0 and group 1 (i.e., for $1 \leq n \leq n_1$). More specifically, for the nth database belonging to group 0, the user considers all the undesired symbols downloaded from all databases (except the nth database) in the $(k-1)$th round as side information. Since database symmetry applies over each group, and from the fact that each stage in the $(k-1)$th round initiates a stage in

the kth round, we have

$$y_0[k] = (n_0 - 1)y_0[k-1] + \sum_{\ell \in \mathcal{S}\setminus\{0\}} (n_\ell - n_{\ell-1})y_\ell[k-1] \tag{5.67}$$

where the left side is the total number of stages in the $(k-1)$th round from all the $N-1$ databases (i.e., except for the nth database that belongs to group 0). The same argument holds for group 1 as well, hence

$$y_1[k] = (n_1 - n_0 - 1)y_1[k-1] + \sum_{\ell \in \mathcal{S}\setminus\{1\}} (n_\ell - n_{\ell-1})y_\ell[k-1] \tag{5.68}$$

where $(n_1 - n_0 - 1)$ denotes the number of databases in group 1 other than the nth database.

For a database belonging to the ℓth group such that $\ell \geq 2$, the user can generate extra stages by exploiting the symbols downloaded in round 1. To initiate one stage in the $(\ell+1)$th round, the user needs to combine symbols from $\frac{\binom{M-1}{\ell}\ell}{M-1} = \binom{M-2}{\ell-1}$ stages. Therefore, the number of stages initiated in the $(\ell+1)$th round as a consequence of the side information in round 1 is $\xi_\ell = \frac{y_0[1]}{\binom{M-2}{\ell-1}} = \prod_{s\in\mathcal{S}\setminus\{\ell\}} \binom{M-2}{s-1}$. Since these extra side information can be used once (at the $(\ell+1)$th round only) and after that for the kth round, the database exploits the side information generated in the $(k-1)$th round only. We represent this one-time exploitation of side information in the $(\ell+1)$th round by the Kronecker delta function $\delta[k-\ell-1]$. Consequently, the number of stages for the ℓth group, $\ell \geq 2$ is related via the following difference

equation:

$$y_\ell[k] = n_0\xi_\ell\delta[k-\ell-1] + (n_\ell - n_{\ell-1} - 1)y_\ell[k-1] + \sum_{j\in\mathcal{S}\setminus\{\ell\}} (n_j - n_{j-1})y_j[k-1] \quad (5.69)$$

Now, we are ready to characterize $\boldsymbol{\tau}(\mathbf{n})$ and $R(\boldsymbol{\tau}(\mathbf{n}))$ in terms of $y_\ell[k]$, where $\ell \in \mathcal{S}$ and $k = 1, \cdots, M$. For any stage in the kth round, the user downloads $\binom{M-1}{k-1}$ desired symbols from a total of $\binom{M}{k}$ downloads. Therefore, from a database belonging to the ℓth group, the user downloads $\sum_{k=1}^{M}\binom{M-1}{k-1}y_\ell[k]$ desired symbols from a total of $\sum_{k=1}^{M}\binom{M}{k}y_\ell[k]$. The number of databases belonging to the ℓth group is given by $n_\ell - n_{\ell-1}$. Therefore, the total download is given by,

$$\sum_{n=1}^{N} t_n(\boldsymbol{\tau}(\mathbf{n})) = \sum_{\ell\in\mathcal{S}}\sum_{k=1}^{M}\binom{M}{k}y_\ell[k](n_\ell - n_{\ell-1}) \quad (5.70)$$

Thus, the traffic ratio of the nth database belonging to the ℓth group (i.e., $n_{\ell-1}+1 \le n \le n_\ell$) corresponding to $\mathbf{n} = \{n_i\}_{i=0}^{M-1}$ is given by

$$\tau_n(\mathbf{n}) = \tilde{\tau}_\ell = \frac{\sum_{k=1}^{M}\binom{M}{k}y_\ell[k]}{\sum_{\ell\in\mathcal{S}}\sum_{k=1}^{M}\binom{M}{k}y_\ell[k](n_\ell - n_{\ell-1})}, \quad n_{\ell-1}+1 \le n \le n_\ell \quad (5.71)$$

Furthermore, the total desired symbols from all databases is given by

$$L(\boldsymbol{\tau}(\mathbf{n})) = \sum_{\ell\in\mathcal{S}}\sum_{k=1}^{M}\binom{M-1}{k-1}y_\ell[k](n_\ell - n_{\ell-1}) \quad (5.72)$$

which further leads to the following achievable rate

$$R(\boldsymbol{\tau}(\mathbf{n})) = \frac{\sum_{\ell\in\mathcal{S}}\sum_{k=1}^{M}\binom{M-1}{k-1}y_\ell[k](n_\ell - n_{\ell-1})}{\sum_{\ell\in\mathcal{S}}\sum_{k=1}^{M}\binom{M}{k}y_\ell[k](n_\ell - n_{\ell-1})} \tag{5.73}$$

5.6 Optimality of $M = 2$ and $M = 3$ Cases

In this section, we prove Corollary 5.2, i.e., we prove that the capacity of the PIR problem under asymmetric traffic constraints $C(\boldsymbol{\tau})$ for $M = 2, 3$ is given by (5.22). We note that since the upper bound in Theorem 5.1 is affine in $\boldsymbol{\tau}$ and time-sharing rates are achievable from Lemma 5.3, it suffices to prove the optimality of all corner points to settle the PIR capacity $C(\boldsymbol{\tau})$ for $M = 2, 3$. In the following, we use Theorem 5.1 and Theorem 5.2 to show the optimality of these corner points.

5.6.1 $M = 2$ Messages

We start the proof from the achievability side. From Theorem 5.2, the corner points are specified by the non-decreasing sequence $\mathbf{n} = (n_0, n_1)$. In this case, the system of difference equations in (5.16) is reduced to

$$y_0[k] = (n_0 - 1)y_0[k-1] \tag{5.74}$$

$$y_1[k] = n_0 y_0[k-1] \tag{5.75}$$

for $k = 1, 2$, where $y_0[1] = 1$, and $y_1[1] = 0$. Hence, $y_0[2] = n_0 - 1$, and $y_1[2] = n_0$. Hence, the total downloads for the corner point $\mathbf{n} = (n_0, n_1)$ is

$$\sum_{n=1}^{N} t_n(\boldsymbol{\tau}(\mathbf{n})) = \sum_{\ell=0}^{1} \sum_{k=1}^{2} \binom{2}{k} y_\ell[k](n_\ell - n_{\ell-1}) = n_0(n_1 + 1) \tag{5.76}$$

Thus, from Theorem 5.2, the traffic-ratio vector $\boldsymbol{\tau}(\mathbf{n})$ is given by

$$\tilde{\tau}_0 = \frac{\binom{2}{1} y_0[1] + \binom{2}{2} y_0[2]}{\sum_{n=1}^{N} t_n(\boldsymbol{\tau}(\mathbf{n}))} = \frac{n_0 + 1}{n_0(n_1 + 1)} \tag{5.77}$$

$$\tilde{\tau}_1 = \frac{\binom{2}{1} y_1[1] + \binom{2}{2} y_1[2]}{\sum_{n=1}^{N} t_n(\boldsymbol{\tau}(\mathbf{n}))} = \frac{1}{n_1 + 1} \tag{5.78}$$

where $\tau_n = \tilde{\tau}_0$, for $1 \leq n \leq n_0$, and $\tau_n = \tilde{\tau}_1$, for $n_0 + 1 \leq n \leq n_1$, and $\tau_n = 0$ otherwise. For the desired symbols, the user downloads $L_0(\boldsymbol{\tau}(\mathbf{n}))$ symbols from the nth database when $1 \leq n \leq n_0$, and $L_1(\boldsymbol{\tau}(\mathbf{n}))$ symbols from the nth database when $n_0 + 1 \leq n \leq n_1$

$$L_0(\boldsymbol{\tau}(\mathbf{n})) = y_0[1] + y_0[2] = n_0 \tag{5.79}$$

$$L_1(\boldsymbol{\tau}(\mathbf{n})) = y_1[1] + y_1[2] = n_0 \tag{5.80}$$

Consequently, $L = n_0 L_0 + (n_1 - n_0) L_1 = n_0 n_1$, and the achievable retrieval rate $R(\boldsymbol{\tau}(\mathbf{n}))$ is given by

$$R(\boldsymbol{\tau}(\mathbf{n})) = \frac{L(\boldsymbol{\tau}(\mathbf{n}))}{\sum_{n=1}^{N} t_n(\boldsymbol{\tau}(\mathbf{n}))} = \frac{n_1}{n_1 + 1} \tag{5.81}$$

For the converse, we evaluate the bound in (5.10) (without the minimization) for $n_1 = n_0$, i.e., we substitute with n_0 in the argument of the upper bound. Then, we have the following upper bound

$$R(\boldsymbol{\tau}(\mathbf{n})) \leq \frac{1 + \frac{\sum_{n=n_0+1}^{N} \tau_n}{n_0}}{1 + \frac{1}{n_0}} \tag{5.82}$$

$$= \frac{1 + \frac{(n_1 - n_0)\tilde{\tau}_1}{n_0}}{1 + \frac{1}{n_0}} \tag{5.83}$$

$$= \frac{n_1}{n_1 + 1} \tag{5.84}$$

This concludes the optimality of our achievable scheme for $M = 2$.

5.6.2 $M = 3$ Messages

Similarly, for the corner point specified by the non-decreasing sequence $\mathbf{n} = (n_0, n_1, n_2)$, we have the following system of difference equations for $k = 1, 2, 3$

$$y_0[k] = (n_0 - 1)y_0[k-1] + (n_1 - n_0)y_1[k-1] + (n_2 - n_1)y_2[k-1] \tag{5.85}$$

$$y_1[k] = n_0 y_0[k-1] + (n_1 - n_0 - 1)y_1[k-1] + (n_2 - n_1)y_2[k-1] \tag{5.86}$$

$$y_2[k] = n_0\delta[k-3] + n_0 y_0[k-1] + (n_1 - n_0)y_1[k-1] + (n_2 - n_1 - 1)y_2[k-1] \tag{5.87}$$

with the initial conditions $y_0[1] = 1$, $y_1[1] = 0$, and $y_2[1] = y_2[2] = 0$. Evaluating $y_\ell[k]$, for $\ell = 0, 1, 2$, and $k = 1, 2, 3$ recursively leads to $y_0[2] = n_0 - 1$, $y_1[2] = n_0$, $y_0[3] = n_1 n_0 - 2n_0 + 1$, $y_1[3] = n_1 n_0 - 2n_0$, and $y_2[3] = n_1 n_0$. This leads to the

following total download

$$\sum_{n=1}^{N} t_n(\boldsymbol{\tau}(\mathbf{n})) = \sum_{\ell=0}^{2}\sum_{k=1}^{3}\binom{3}{k} y_\ell[k](n_\ell - n_{\ell-1}) = n_0(n_1 n_2 + n_1 + 1) \tag{5.88}$$

The sequence $\mathbf{n} = (n_0, n_1, n_2)$ specifies the traffic ratio vector $\boldsymbol{\tau}(\mathbf{n})$ such that

$$\tilde{\tau}_0 = \frac{n_0 n_1 + n_0 + 1}{n_0(n_2 n_1 + n_1 + 1)} \tag{5.89}$$

$$\tilde{\tau}_1 = \frac{n_1 + 1}{n_2 n_1 + n_1 + 1} \tag{5.90}$$

$$\tilde{\tau}_2 = \frac{n_1}{n_2 n_1 + n_1 + 1} \tag{5.91}$$

where $\tau_n = \tilde{\tau}_0$ for $1 \le n \le n_0$, $\tau_n = \tilde{\tau}_1$ for $n_0+1 \le n \le n_1$, $\tau = \tilde{\tau}_2$ for $n_1+1 \le n \le n_2$, and $\tau_n = 0$ otherwise.

For the desired symbols, the user downloads $L_0(\boldsymbol{\tau}(\mathbf{n}))$ symbols from the nth database if $1 \le n \le n_0$, $L_1(\boldsymbol{\tau}(\mathbf{n}))$ symbols if $n_0 + 1 \le n \le n_1$, and $L_2(\boldsymbol{\tau}(\mathbf{n}))$ symbols if $n_1 + 1 \le n \le n_2$, hence

$$L_\ell(\boldsymbol{\tau}(\mathbf{n})) = \sum_{k=1}^{3}\binom{2}{k-1} y_\ell[k] = n_0 n_1, \quad \ell = 0, 1, 2 \tag{5.92}$$

Consequently, the following rate is achievable

$$R(\boldsymbol{\tau}(\mathbf{n})) = \frac{n_1 n_2}{n_1 n_2 + n_1 + 1} \tag{5.93}$$

For the converse, pick (n_1, n_2) in the converse bound to be (n_0, n_1), which

leads to the following bound

$$R(\boldsymbol{\tau}(\mathbf{n})) \leq \frac{1+\frac{\sum_{n=n_0+1}^{N}\tau_n}{n_0}+\frac{\sum_{n=n_1+1}^{N}\tau_n}{n_0 n_1}}{1+\frac{1}{n_0}+\frac{1}{n_0 n_1}} \tag{5.94}$$

$$= \frac{1+\frac{(n_1-n_0)\tilde{\tau}_1}{n_0}+\frac{(n_2-n_1)\tilde{\tau}_2}{n_0}+\frac{(n_2-n_1)\tilde{\tau}_2}{n_0 n_1}}{1+\frac{1}{n_0}+\frac{1}{n_0 n_1}} \tag{5.95}$$

$$= \frac{n_1 n_2}{n_1 n_2 + n_1 + 1} \tag{5.96}$$

This concludes the optimality of our achievable scheme for $M = 3$.

Remark 5.11 *We note that, surprisingly, for the corner points of the cases $M = 2$ and $M = 3$, the number of desired symbols downloaded from each active database is the same irrespective to the traffic ratio of the database; see* (5.79)-(5.80) *for $M = 2$ and* (5.92) *for $M = 3$. This suggests that at these corner points, the optimal scheme performs* combinatorial water-filling *for the undesired symbols first, i.e., the nth active database downloads $t_n - n_0$ undesired symbols for $M = 2$ and $t_n - n_0 n_1$ undesired symbols for $M = 3$, and then downloads the same number of desired symbols from all active databases.*

5.7 Achievable Tradeoff for $N = 2$ and Arbitrary M

For the special case of $N = 2$, and an arbitrary M, the retrieval rate calculation in Theorem 5.2 is significantly simplified. Let $s_2 \in \{0, \cdots, M-1\}$ be the number of side information symbols that are used simultaneously in the initial round of download at the second database. Note that there is a bijection between s_2 and the non-decreasing sequence $\mathbf{n}$ as $n_0 = n_1 = \cdots = n_{s_2-1} = 1$, and $n_{s_2} = 2$ for

any corner point other than the corner point corresponding to the trivial scheme of downloading the contents of the first database.

The user starts with downloading $\binom{M-2}{s_2-1}$ stages of individual symbols (i.e., the user downloads $M\binom{M-2}{s_2-1}$ symbols in round 1 from all messages) from the first database to create 1 stage in the (s_2+1)th round. After the initial exploitation of side information, the two databases exchange side information. More specifically, from database 1 in the (s_2+2k)th round, where $k=1,\cdots,\lfloor\frac{M-s_2}{2}\rfloor$, the user exploits the side information generated in database 2 in the (s_2+2k-1)th round to download $\binom{M-1}{s_2+2k-1}$ desired symbols (by adding one symbol of the desired symbols to the (s_2+2k-1)-sum of undesired symbols generated in database 2) from total download in the (s_2+2k)th round of $\binom{M}{s_2+2k}$. Similarly from database 2, in the (s_2+2k+1)th round, where $k=0,\cdots,\lfloor\frac{M-s_2-1}{2}\rfloor$, the user exploits the side information generated in database 1 in the (s_2+2k)th round, and downloads $\binom{M-1}{s_2+2k}$ desired symbols from total of $\binom{M}{s_2+2k+1}$ downloads in the (s_2+2k+1)th round.

Consequently, we have

$$t_1(s_2) = M\binom{M-2}{s_2-1} + \sum_{k=1}^{\lfloor\frac{M-s_2}{2}\rfloor}\binom{M}{s_2+2k} \tag{5.97}$$

$$t_2(s_2) = \sum_{k=0}^{\lfloor\frac{M-s_2-1}{2}\rfloor}\binom{M}{s_2+2k+1} \tag{5.98}$$

which further leads to the following total download

$$t_1(s_2)+t_2(s_2) = M\binom{M-2}{s_2-1} + \sum_{k=0}^{M-s_2-1}\binom{M}{s_2+k+1} \tag{5.99}$$

Thus, the traffic ratio $\tau_2(s_2)$ is given by

$$\tau_2(s_2) = \frac{t_2(s_2)}{t_1(s_2) + t_2(s_2)} = \frac{\sum_{k=0}^{\lfloor \frac{M-s_2-1}{2} \rfloor} \binom{M}{s_2+2k+1}}{M\binom{M-2}{s_2-1} + \sum_{k=0}^{M-s_2-1} \binom{M}{s_2+k+1}} \tag{5.100}$$

The total number of desired symbols is given by

$$L(s_2) = \binom{M-2}{s_2-1} + \sum_{k=1}^{\lfloor \frac{M-s_2}{2} \rfloor} \binom{M-1}{s_2+2k-1} + \sum_{k=0}^{\lfloor \frac{M-s_2-1}{2} \rfloor} \binom{M-1}{s_2+2k} \tag{5.101}$$

$$= \binom{M-2}{s_2-1} + \sum_{k=0}^{M-s_2-1} \binom{M-1}{s_2+k} \tag{5.102}$$

Thus, the following rate is achievable for $N = 2$ and arbitrary M

$$R(s_2) = \frac{L(s_2)}{t_1(s_2) + t_2(s_2)} = \frac{\binom{M-2}{s_2-1} + \sum_{k=0}^{M-s_2-1} \binom{M-1}{s_2+k}}{M\binom{M-2}{s_2-1} + \sum_{k=0}^{M-s_2-1} \binom{M}{s_2+k+1}} \tag{5.103}$$

5.8 Further Examples

In this section, we present further examples to clarify the achievable scheme for some additional tractable values of M, N.

5.8.1 $M = 4$ Messages, $N = 2$ Databases

In this example, we show that the achievable rate $R(\tau_2)$ does not match the upper bound $\bar{C}(\tau_2)$ for all traffic ratios τ_2. For $M = 4$, we have $M + 1 = 5$ corner points, corresponding to $s_2 = \{0, 1, 2, 3\}$ and another corner point corresponding to the trivial scheme of downloading the contents of database 1. Let a_i, b_i, c_i, d_i

denote the randomly permuted symbols of messages W_1, W_2, W_3, W_4, respectively. Then, $R(0) = \frac{1}{4}$ by trivially downloading a_1, b_1, c_1, d_1 from database 1. In addition, $R(\frac{1}{2}) = \frac{1-\frac{1}{2}}{1-(\frac{1}{2})^4} = \frac{8}{15}$ using the symmetric scheme in [12].

Corner Point $s_2 = 1$: (See the query table in Table 5.6.) The user uses 1 bit of side information in database 2, hence the user starts downloading from round 2 (that admits 2-sums). The user exploits the side information generated in round 1 by downloading $a_2 + b_1$, $a_3 + c_1$, and $a_4 + d_1$. The user completes the stage by downloading undesired symbols consisting of 2-sums that do not include a_i, hence the user downloads $b_2 + c_2$, $b_3 + d_2$, $c_3 + d_3$. The undesired symbols are exploited in database 1, thus the user downloads $a_5 + b_2 + c_2$, $a_6 + b_3 + d_2$, and $a_7 + c_3 + d_3$. The user completes the stage by downloading $b_4 + c_4 + d_4$, which can be exploited in database 2 by downloading $a_8 + b_4 + c_4 + d_4$. In this case, the user downloads 8 symbols from database 1 and 7 symbols from database 2, hence we have $\tau_2 = \frac{7}{15}$. Since the user downloads $L = 8$ desired symbols, the achievable rate $R(\frac{7}{15}) = \frac{8}{15}$.

Table 5.6: The query table for $M = 4$, $N = 2$, $s_2 = 1$ (corresponding to $\tau_2 = \frac{7}{15}$).

Database 1	Database 2
a_1, b_1, c_1, d_1	
	$a_2 + b_1$ $a_3 + c_1$ $a_4 + d_1$ $b_2 + c_2$ $b_3 + d_2$ $c_3 + d_3$
$a_5 + b_2 + c_2$ $a_6 + b_3 + d_2$ $a_7 + c_3 + d_3$ $b_4 + c_4 + d_4$	
	$a_8 + b_4 + c_4 + d_4$

Corner Point $s_2 = 2$: (See the query table in Table 5.7.) The user downloads $\binom{M-2}{s_2-1} = 2$ stages of individual symbols (1-sum) from database 1, so that the user forms 2-sums that can be used in database 2 as side information to start round 3 directly, i.e., by forming 2-sums as side information from the individual symbols, the user effectively skips round 2. More specifically, the user downloads $a_3 + b_1 + c_1$, $a_4 + b_2 + d_1$, $a_5 + c_2 + d_2$ from database 2 taking into considerations that all these undesired symbols are decodable from database 1. The user completes the stage by downloading $b_3 + c_3 + d_3$ that can be further exploited in database 1 by downloading $a_6 + b_3 + c_3 + d_3$. In this case, the user downloads 9 symbols from database 1 and 4 symbols from database 2, therefore $\tau_2 = \frac{4}{13}$. The user downloads $L = 6$ desired symbols, thus, $R(\frac{4}{13}) = \frac{6}{13}$.

Table 5.7: The query table for $M = 4$, $N = 2$, $s_2 = 2$ (corresponding to $\tau_2 = \frac{4}{13}$).

Database 1	Database 2
a_1, b_1, c_1, d_1 a_2, b_2, c_2, d_2	
	$a_3 + b_1 + c_1$ $a_4 + b_2 + d_1$ $a_5 + c_2 + d_2$ $b_3 + c_3 + d_3$
$a_6 + b_3 + c_3 + d_3$	

Corner Point $s_2 = 3$: (See the query table in Table 5.8.) In this case, the user skips rounds 2, 3 and jumps directly to round 4 at database 2. Therefore, the user downloads $a_2 + b_1 + c_1 + d_1$ from database 2, which uses $b_1 + c_1 + d_1$ as side information which is decodable from database 1. Thus, we have $\tau_2 = \frac{1}{5}$, and the corresponding rate $R(\frac{1}{5}) = \frac{2}{5}$.

Table 5.8: The query table for $M = 4$, $N = 2$, $s_2 = 3$ (corresponding to $\tau_2 = \frac{1}{5}$).

Database 1	Database 2
a_1, b_1, c_1, d_1	
	$a_2 + b_1 + c_1 + d_1$

Comparison with the Upper Bound: The upper bound in Theorem 5.1 can be explicitly expressed as:

$$R(\tau_2) \leq \begin{cases} \frac{1}{4} + \frac{3\tau_2}{4}, & 0 \leq \tau_2 \leq \frac{1}{5} \\ \frac{2}{7} + \frac{4\tau_2}{7}, & \frac{1}{5} \leq \tau_2 \leq \frac{3}{8} \\ \frac{4}{11} + \frac{4\tau_2}{11}, & \frac{3}{8} \leq \tau_2 \leq \frac{7}{15} \\ \frac{8}{15}, & \frac{7}{15} \leq \tau_2 \leq \frac{1}{2} \end{cases} \tag{5.104}$$

We observe that for all the corner points of the achievable scheme, the upper and lower bounds match. However, the upper bound has an extra corner point $(\frac{3}{8}, \frac{1}{2})$ which is not achievable using time-sharing. This is illustrated in Fig. 5.5

5.8.2 $M = 3$ Messages, $N = 3$ Databases

In this example, we show the capacity-achieving scheme for $M = 3$, $N = 3$ (the capacity region is illustrated in Fig. 5.3 as a function of $C(\lambda_2, \lambda_3)$). Let a_i, b_i, c_i denote the permuted symbols of messages W_1, W_2, W_3, respectively. We show here only the query tables for achieving non-trivial corner points. In this case, we have $\binom{M+N-1}{M} = 10$ corner points corresponding to non-decreasing sequences (n_0, n_1, n_2).

For the pair $(\tau_2, \tau_3) = (0, 0)$, the achievable scheme is the trivial scheme that

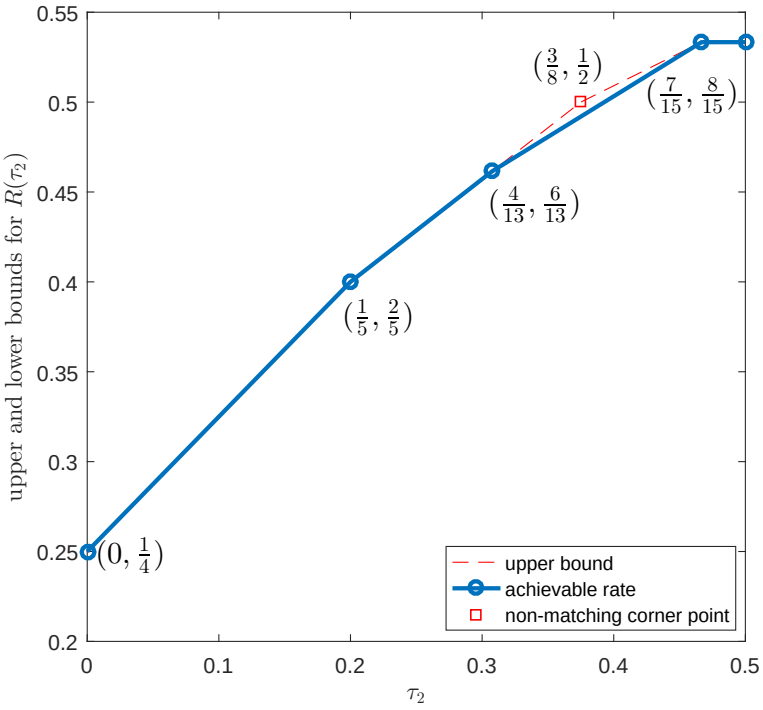


Figure 5.5: Upper and lower bounds for $R(\tau_2)$ for $M = 4$, $N = 2$.

downloads a_1, b_1, c_1 from the first database only achieving $R(0,0) = \frac{1}{3}$. For the corner point $(\frac{1}{4}, 0)$, this is exactly the same corner point presented in Section 5.5.1.1 (for $\lambda_2 = \frac{1}{3}$) as $\tau_3 = 0$, which effectively reduces the problem to $N = 2$ databases. The achievable scheme for this corner point is illustrated in Table 5.4, hence $R(\frac{1}{4}, 0) = \frac{1}{2}$. For the corner point $(\frac{3}{7}, 0)$, again this point reduces to 2 databases. The achievable scheme is given in Table 5.3, and $R(\frac{3}{7}, 0) = \frac{4}{7}$. For the corner point $(\frac{1}{3}, \frac{1}{3})$, which is the symmetric-traffic point, the achievable scheme is the symmetric scheme in [12], which achieves $R(\frac{1}{3}, \frac{1}{3}) = \frac{9}{13}$. For the corner point $(\frac{1}{2}, 0)$, we can apply the symmetric achievable scheme for $N = 2$ databases only as $\tau_3 = 0$ in this case, hence $R(\frac{1}{2}, 0) = \frac{4}{7}$.

Now, we focus on the non-trivial corner points. As mentioned previously, the

pair (s_2, s_3) is in bijection with the sequence (n_0, n_1, n_2). Therefore, we enumerate the remaining cases using the pair (s_2, s_3).

Corner Point $(s_2, s_3) = (0, 1)$: In this case, the user does not use the side information generated in database 1 within the initial download of database 2 ($s_2 = 0$), hence the user downloads new individual symbols from database 2. The user uses 1 bit of side information in database 3 in its round of download (round 2). These side information symbols come from database 1 and database 2. The query table for this case is shown in Table 5.9. In this case, we have $(\tau_2, \tau_3) = (\frac{9}{26}, \frac{4}{13})$, and the achievable rate is $R(\frac{9}{26}, \frac{4}{13}) = \frac{9}{13}$.

Table 5.9: The query table for $M = 3$, $N = 3$, $(s_2, s_3) = (0, 1)$ (i.e., $(\tau_2, \tau_3) = (\frac{9}{26}, \frac{4}{13})$).

Database 1	Database 2	Database 3
a_1, b_1, c_1	a_2, b_2, c_2	
$a_3 + b_2$ $a_4 + c_2$ $b_3 + c_3$	$a_5 + b_1$ $a_6 + c_1$ $b_4 + c_4$	$a_7 + b_1$ $a_8 + c_1$ $b_5 + c_5$ $a_9 + b_2$ $a_{10} + c_2$ $b_6 + c_6$
$a_{11} + b_4 + c_4$ $a_{12} + b_5 + c_5$ $a_{13} + b_6 + c_6$	$a_{14} + b_3 + c_3$ $a_{15} + b_5 + c_5$ $a_{16} + b_6 + c_6$	$a_{17} + b_3 + c_3$ $a_{18} + b_4 + c_4$

Corner Point $(s_2, s_3) = (0, 2)$: The user does not exploit the side information generated from database 1 in the first round of download at database 2. The user uses 2 side information symbols simultaneously in the initial round (round 3) of download at database 3. Note that in round 3 database 3 receives side information from rounds 1 and 2 of databases 1 and 2. The query table for this case is shown

in Table 5.10. In this case, we have $(\tau_2, \tau_3) = (\frac{7}{18}, \frac{2}{9})$, and the achievable rate is $R(\frac{7}{18}, \frac{2}{9}) = \frac{2}{3}$.

Table 5.10: The query table for $M = 3$, $N = 3$, $(s_2, s_3) = (0, 2)$ (i.e., $(\tau_2, \tau_3) = (\frac{7}{18}, \frac{2}{9})$).

Database 1	Database 2	Database 3
a_1, b_1, c_1	a_2, b_2, c_2	
$a_3 + b_2$ $a_4 + c_2$ $b_3 + c_3$	$a_5 + b_1$ $a_6 + c_1$ $b_4 + c_4$	
$a_7 + b_4 + c_4$	$a_8 + b_3 + c_3$	$a_9 + b_1 + c_1$ $a_{10} + b_2 + c_2$ $a_{11} + b_3 + c_3$ $a_{12} + b_4 + c_4$

Corner Point $(s_2, s_3) = (1, 1)$: In this case, both databases 2 and 3 exploit the side information generated from database 1 in their initial round of download (round 1). The query table for this case is shown in Table 5.11. In this case, we have $(\tau_2, \tau_3) = (\frac{4}{13}, \frac{4}{13})$, and the achievable rate is $R(\frac{4}{13}, \frac{4}{13}) = \frac{9}{13}$.

Table 5.11: The query table for $M = 3$, $N = 3$, $(s_2, s_3) = (1, 1)$ (i.e., $(\tau_2, \tau_3) = (\frac{4}{13}, \frac{4}{13})$).

Database 1	Database 2	Database 3
a_1, b_1, c_1		
	$a_2 + b_1$ $a_3 + c_1$ $b_2 + c_2$	$a_4 + b_1$ $a_5 + c_1$ $b_3 + c_3$
$a_6 + b_2 + c_2$ $a_7 + b_3 + c_3$	$a_8 + b_3 + c_3$	$a_9 + b_2 + c_2$

Corner Point $(s_2, s_3) = (1, 2)$: In this case, database 2 exploits 1 side information in its initial download (round 2), while database 3 skips to round 3 directly. Database 3 receives side information from the round 1 of database 1 and round 2 of database

2. The query table for this case is shown in Table 5.12. In this case, we have $(\tau_2, \tau_3) = (\frac{1}{3}, \frac{2}{9})$, and the achievable rate is $R(\frac{1}{3}, \frac{2}{9}) = \frac{2}{3}$.

Table 5.12: The query table for $M = 3$, $N = 3$, $(s_2, s_3) = (1, 2)$ (i.e., $(\tau_2, \tau_3) = (\frac{1}{3}, \frac{2}{9})$).

Database 1	Database 2	Database 3
a_1, b_1, c_1		
	$a_2 + b_1$ $a_3 + c_1$ $b_2 + c_2$	
$a_4 + b_2 + c_2$		$a_5 + b_1 + c_1$ $a_6 + b_2 + c_2$

Corner Point $(s_2, s_3) = (2, 2)$: Both databases 2 and 3 skip round 1 and 2 of downloads and go directly to round 3, in which they exploits 2 side information symbols simultaneously. The query table for this case is shown in Table 5.13. In this case, we have $(\tau_2, \tau_3) = (\frac{1}{5}, \frac{1}{5})$, and the achievable rate is $R(\frac{1}{5}, \frac{1}{5}) = \frac{3}{5}$.

Table 5.13: The query table for $M = 3$, $N = 3$, $(s_2, s_3) = (2, 2)$ (i.e., $(\tau_2, \tau_3) = (\frac{1}{5}, \frac{1}{5})$).

Database 1	Database 2	Database 3
a_1, b_1, c_1		
	$a_2 + b_1 + c_1$	$a_3 + b_1 + c_1$

5.9 Conclusions

In this chapter, we introduced the PIR problem under asymmetric traffic constraints $\boldsymbol{\tau}$. We investigated the fundamental limits of this problem by developing the novel upper bound $\bar{C}(\boldsymbol{\tau}) = \min_{n_1,\cdots,n_{M-1}\in\{1,\cdots,N\}} \frac{1+\frac{\sum_{n=n_1+1}^{N}\tau_n}{n_1}+\frac{\sum_{n=n_2+1}^{N}\tau_n}{n_1 n_2}+\cdots+\frac{\sum_{n=n_{M-1}+1}^{N}\tau_n}{n_0 n_1\cdots n_{M-1}}}{1+\frac{1}{n_1}+\frac{1}{n_1 n_2}+\cdots+\frac{1}{n_0 n_1\cdots n_{M-1}}}$, for some integer sequence $\{n_i\}_{i=1}^{N} \subset \{1, \cdots, N\}^{M-1}$. The upper bound generalizes

the converse proof in [12], which inherently utilizes database symmetry. The upper bound is a piece-wise affine function in $\boldsymbol{\tau}$. The upper bound implies a strict capacity loss due to the asymmetric traffic constraints for certain cases. We developed explicit achievable schemes for $\binom{M+N-1}{M}$ corner points, and achieved the remaining points by time-sharing. We described the achievable scheme by means of a system of difference equations. We explicitly derived the achievable rate for $N = 2$ and arbitrary M. We proved that the upper bound and the lower bound exactly match for every $\boldsymbol{\tau}$ for the cases of $M = 2$ and $M = 3$ for any N.

CHAPTER 6

Noisy Private Information Retrieval

6.1 Introduction

In this chapter, we consider the problem of noisy private information retrieval (NPIR) from N non-communicating databases, each storing the same set of M messages. In this model, the answer strings are not returned through noiseless bit pipes, but rather through *noisy* memoryless channels. We aim at characterizing the PIR capacity for this model as a function of the statistical information measures of the noisy channels such as entropy and mutual information. We derive a general upper bound for the retrieval rate in the form of a max-min optimization. We use the achievable schemes for the PIR problem under asymmetric traffic constraints and random coding arguments to derive a general lower bound for the retrieval rate. The upper and lower bounds match for $M = 2$ and $M = 3$, for any N, and any noisy channel. The results imply that separation between channel coding and retrieval is optimal except for adapting the traffic ratio from the databases. We refer to this as *almost separation*.

Next, we consider the private information retrieval problem from multiple

access channels (MAC-PIR). In MAC-PIR, the database responses reach the user through a multiple access channel (MAC) that mixes the responses together in a stochastic way. We show that for the additive MAC and the conjunction/disjunction MAC, channel coding and retrieval scheme are *inseparable* unlike in NPIR. We show that the retrieval scheme depends on the properties of the MAC, in particular on the linearity aspect. For both cases, we provide schemes that achieve the full capacity without any loss due to the privacy constraint, which implies that the user can exploit the nature of the channel to improve privacy. Finally, we show that the full unconstrained capacity is not always attainable by determining the capacity of the selection channel.

6.2 System Model

We consider a classical PIR model with N replicated and non-communicating databases storing M messages. Each database stores the same set of messages $W_{1:M} = \{W_1, \cdots, W_M\}$. The mth message W_m is an L-length binary (without loss of generality) vector picked uniformly from $\mathbb{F}_2^L$. The messages $W_{1:M}$ are independent and identically distributed, i.e.,

$$H(W_m) = L, \quad m \in \{1, \cdots, M\} \tag{6.1}$$

$$H(W_{1:M}) = ML \tag{6.2}$$

In PIR, a user wants to retrieve a message W_i reliably and privately. To that end, the user submits N queries $Q_{1:N}^{[i]} = \{Q_1^{[i]}, \cdots, Q_N^{[i]}\}$, one for each database.

Since the user does not have any information about the message set in advance, the queries and the messages are statistically independent,

$$I(W_{1:M}; Q_{1:N}^{[i]}) = 0, \quad i \in \{1, \cdots, M\} \tag{6.3}$$

The nth database responds to $Q_n^{[i]}$ with a t_n-length answer string $A_n^{[i]} = (X_{n,1}^{[i]}, \cdots, X_{n,t_n}^{[i]})$. The nth answer string is a deterministic function of the messages $W_{1:M}$ and the query $Q_n^{[i]}$, hence,

$$H(A_n^{[i]}|W_{1:M}, Q_n^{[i]}) = 0, \quad n \in \{1, \cdots, N\}, \ i \in \{1, \cdots, M\} \tag{6.4}$$

In noisy PIR with orthogonal links (NPIR, see Fig. 6.1), the user receives the nth answer string via a discrete memoryless channel (response channel) with a transition probability $p(y_n|x_n)$. In this model, the noisy channels are *orthogonal*, in the sense that the noisy answer strings do not interact (mix). Thus, the user receives a noisy answer string $\tilde{A}_n^{[i]} = (Y_{n,1}^{[i]}, \cdots, Y_{n,t_n}^{[i]})$. Therefore, we have,

$$P\left(\tilde{A}_n^{[i]} = (y_{n,1}^{[i]}, \cdots, y_{n,t_n}^{[i]})|A_n^{[i]} = (x_{n,1}^{[i]}, \cdots, x_{n,t_n}^{[i]})\right) = \prod_{\eta_n=1}^{t_n} p\left(y_{n,\eta_n}^{[i]}|x_{n,\eta_n}^{[i]}\right) \tag{6.5}$$

Consequently, $(W_{1:M}, Q_n^{[i]}) \to A_n^{[i]} \to \tilde{A}_n^{[i]}$ forms a Markov chain. Let us denote the channel capacity of the nth response channel by C_n, denote,

$$C_n = \max_{p(x_n)} \ I(X_n; Y_n) \tag{6.6}$$

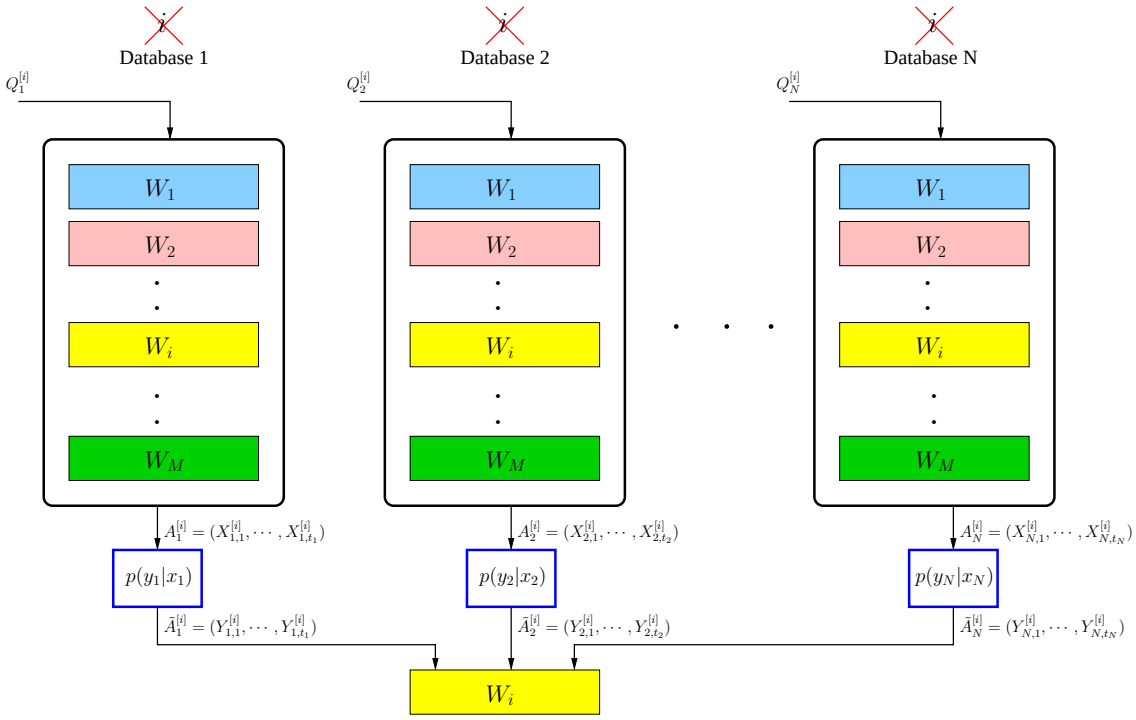


Figure 6.1: The noisy PIR (NPIR) problem.

where X_n, Y_n are the single-letter input and output pair for the nth response channel. Without loss of generality, assume that the channel capacities are ordered such that $C_1 \geq C_2 \geq \cdots \geq C_N$, i.e., the channel capacities form a non-increasing sequence. Let $\mathbf{C} = (C_1, \cdots, C_N)$ be the vector of the channel capacities.

We note that, in general, the user and the databases can agree on suitable lengths $\{t_n\}_{n=1}^N$ for the answer strings, which may not be equal in general, such that they maximize the retrieval rate. Let us define the traffic ratio vector $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_N)$ as,

$$\tau_n = \frac{t_n}{\sum_{j=1}^N t_j}, \quad n \in \{1, \cdots, N\} \tag{6.7}$$

To ensure privacy, the queries $Q_{1:N}^{[i]}$ should be designed such that the query to

the nth database does not reveal any information about i. We can write the privacy constraint as

$$(Q_n^{[i]}, A_n^{[i]}, W_{1:M}) \sim (Q_n^{[j]}, A_n^{[j]}, W_{1:M}), \quad \forall i, j \in \{1, \cdots, M\} \tag{6.8}$$

We note that from privacy constraint and due to the Markov chain $(W_{1:M}, Q_n^{[i]}) \rightarrow A_n^{[i]} \rightarrow \tilde{A}_n^{[i]}$, we may write that $(Q_n^{[i]}, A_n^{[i]}, \tilde{A}_n^{[i]}, W_{1:M}, W_{1:M}) \sim (Q_n^{[j]}, A_n^{[j]}, \tilde{A}_n^{[j]}, W_{1:M}), \quad \forall i, j \in \{1, \cdots, M\}$.

In addition, the user should be able to reconstruct the desired message W_i by observing the noisy answer strings $\tilde{A}_{1:N}^{[i]}$ with arbitrarily small probability of error $P_e(L)$, i.e., $P_e(L) \rightarrow 0$ as $L \rightarrow \infty$. Hence, from Fano's inequality, we have,

$$H(W_i | Q_{1:N}^{[i]}, \tilde{A}_{1:N}^{[i]}) \leq 1 + P_e(L) \cdot L = o(L) \tag{6.9}$$

where $\frac{o(L)}{L} \rightarrow 0$ as $L \rightarrow \infty$.

For a fixed traffic ratio vector $\boldsymbol{\tau}$, the retrieval rate $R(\boldsymbol{\tau}, \mathbf{C})$ is achievable if there exists a sequence of retrieval schemes, indexed by the message length L, that satisfy the privacy constraint (6.8) and the reliability constraint (6.9) with answer string lengths $\{t_n\}_{n=1}^N$ that conform with (6.7), thus,

$$R(\boldsymbol{\tau}, \mathbf{C}) = \lim_{L \rightarrow \infty} \frac{L}{\sum_{n=1}^N t_n} \tag{6.10}$$

Consequently, the retrieval rate $R(\mathbf{C})$ is the supremum of $R(\boldsymbol{\tau}, \mathbf{C})$ over all traffic ratio vectors in $\mathbb{T} = \{(\tau_1, \cdots, \tau_N) : \tau_n \geq 0 \ \forall n, \sum_{n=1}^N \tau_n = 1\}$. The PIR

capacity for this model $C_{\text{PIR}}(\mathbf{C})$ is given by

$$C_{\text{PIR}}(\mathbf{C}) = \sup \ \ R(\mathbf{C}) \tag{6.11}$$

where the supermum is over all achievable retrieval schemes.

6.3 Main Results and Discussions on NPIR

In this section, we present the main results of the NPIR problem. The first result gives an upper bound for the NPIR problem.

Theorem 6.1 (Upper bound) *For NPIR with noisy links of capacities* $\mathbf{C} = (C_1, \cdots, C_N)$*, the retrieval rate is upper bounded by,*

$$C_{PIR}(\mathbf{C}) \leq \bar{C}_{PIR}(\mathbf{C}) = \max_{\boldsymbol{\tau} \in \mathbb{T}} \min_{n_i \in \{1, \cdot, N\}} \frac{\sum_{n=1}^{N} \tau_n C_n + \frac{\sum_{n=n_1+1}^{N} \tau_n C_n}{n_1} + \cdots + \frac{\sum_{n=n_{M-1}+1}^{N} \tau_n C_n}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{6.12}$$

where $\mathbb{T} = \left\{ \boldsymbol{\tau} : \tau_n \geq 0 \quad \forall n \in [1:N], \quad \sum_{n=1}^{N} \tau_n = 1 \right\}$.

The proof of this upper bound is given in Section 6.4. The second result gives an achievability scheme for the NPIR problem.

Theorem 6.2 (Lower bound) *For NPIR with noisy links of capacities* $\mathbf{C} = (C_1, \cdots, C_N)$*, for a monotone non-decreasing sequence* $\mathbf{n} = \{n_i\}_{i=0}^{M-1} \subset \{1, \cdots, N\}^M$*, let* $n_{-1} = 0$*, and* $\mathcal{S} = \{i \geq 0 : n_i - n_{i-1} > 0\}$*. Denote* $y_\ell[k]$ *to be the number of stages of the achievable scheme that downloads* k*-sums from the*

nth database in one repetition of the scheme, such that $n_{\ell-1} \leq n \leq n_\ell$, *and* $\ell \in \mathcal{S}$. *Let* $\xi_\ell = \prod_{s\in\mathcal{S}\setminus\{\ell\}} \binom{M-2}{s-1}$. *The number of stages* $y_\ell[k]$ *is characterized by the following system of difference equations:*

$$
\begin{aligned}
y_0[k] &= (n_0-1)y_0[k-1] + \sum_{j\in\mathcal{S}\setminus\{0\}} (n_j-n_{j-1})y_j[k-1] \\
y_1[k] &= (n_1-n_0-1)y_1[k-1] + \sum_{j\in\mathcal{S}\setminus\{1\}} (n_j-n_{j-1})y_j[k-1] \\
y_\ell[k] &= n_0\xi_\ell\delta[k-\ell-1] + (n_\ell-n_{\ell-1}-1)y_\ell[k-1] + \sum_{j\in\mathcal{S}\setminus\{\ell\}} (n_j-n_{j-1})y_j[k-1], \quad \ell \geq 2
\end{aligned}
\tag{6.13}
$$

where $\delta[\cdot]$ *is the Kronecker delta function. The initial conditions of* (6.13) *are* $y_0[1] = \prod_{s\in\mathcal{S}} \binom{M-2}{s-1}$, *and* $y_j[k] = 0$ *for* $k \leq j$. *Then, the achievable rate corresponding to* $\mathbf{n}$ *is given by:*

$$
R(\mathbf{n}, \mathbf{C}) = \frac{\sum_{\ell\in\mathcal{S}} \sum_{k=1}^{M} \binom{M-1}{k-1} y_\ell[k](n_\ell - n_{\ell-1})}{\sum_{\ell\in\mathcal{S}} \sum_{n=n_{\ell-1}+1}^{n_\ell} \frac{\sum_{k=1}^{M}\binom{M}{k} y_\ell[k]}{C_n}} \tag{6.14}
$$

Consequently, the capacity $C_{PIR}(\mathbf{C})$ *is lower bounded by:*

$$
C_{PIR}(\mathbf{C}) \geq R(\mathbf{C}) = \max_{n_0\leq\cdots\leq n_{M-1}\in\{1,\cdots,N\}} R(\mathbf{n}, \mathbf{C}) \tag{6.15}
$$

$$
= \max_{n_0\leq\cdots\leq n_{M-1}\in\{1,\cdots,N\}} \frac{\sum_{\ell\in\mathcal{S}} \sum_{k=1}^{M} \binom{M-1}{k-1} y_\ell[k](n_\ell - n_{\ell-1})}{\sum_{\ell\in\mathcal{S}} \sum_{n=n_{\ell-1}+1}^{n_\ell} \frac{\sum_{k=1}^{M}\binom{M}{k} y_\ell[k]}{C_n}} \tag{6.16}
$$

The proof of this lower bound is given in Section 6.5. We have the following remarks.

Remark 6.1 *The upper and lower bounds for the retrieval rate are similar to the corresponding bounds for the PIR-WTC-II problem [124] after replacing the secrecy capacity of WTC-II, $1-\mu_n$, with the capacity of the noisy link C_n. Thus, the NPIR problem inherits all the structural remarks of the PIR-WTC-II problem.*

Remark 6.2 *The upper and lower bounds for the retrieval rate do not depend explicitly on the transition probabilities of the noisy channels $p(y_n|x_n)$, but rather depend on the capacities of the noisy channels C_n.*

Remark 6.3 *Theorem 6.1 and Theorem 6.2 imply that the channel coding needed for combating channel errors is "almost seperable" from the retrieval scheme. The channel coding problem and the retrieval problem are coupled only through agreeing on a traffic ratio vector $\boldsymbol{\tau}$. Other than $\boldsymbol{\tau}$, the channel coding acts as an outer code for the responses of the databases to the user queries. Interestingly, the result implies that our schemes work even for heterogeneous channels, e.g., if $N = 2$, the channel from one database can be a BSC, and the channel from the other database can be a BEC.*

Remark 6.4 *Our results imply that randomized strategies for PIR cannot increase the retrieval rate. We can view the noisy channel between the user and the database as a randomizer for the actions of the databases, which is available to the databases but not available to the user. Since the capacity expression does not depend on $p(y_n|x_n)$ and is always maximized by $C_n = 1$, any randomizing strategy $p(y_n|x_n)$ cannot enhance the retrieval rate.*

Corollary 6.1 (Exact capacity for $M = 2$ and $M = 3$ messages) *For NPIR, the capacity $C_{PIR}(\mathbf{C})$ for $M = 2, 3$, and an arbitrary N is given by:*

$$C_{PIR}(\mathbf{C}) = \begin{cases} \max_{n_0,n_1\in\{1,\cdots,N\}} \frac{n_0 n_1}{\sum_{n=1}^{n_0} \frac{n_0+1}{C_n} + \sum_{n=n_0+1}^{n_1} \frac{n_0}{C_n}}, & M = 2 \\ \max_{n_0,n_1,n_2\in\{1,\cdots,N\}} \frac{n_0 n_1 n_2}{\sum_{n=1}^{n_0} \frac{n_0 n_1+n_0+1}{C_n} + \sum_{n=n_0+1}^{n_1} \frac{n_0 n_1+n_0}{C_n} + \sum_{n=n_1+1}^{n_2} \frac{n_0 n_1}{C_n}}, & M = 3 \end{cases} \tag{6.17}$$

The proof of Corollary 6.1 follows from the optimality of the PIR-WTC-II scheme in [124] for $M = 2$ and $M = 3$ messages by replacing $1 - \mu_n$ by C_n.

Example: The capacity for NPIR from BSC(p_1), BSC(p_2), $N = 2$, $M = 3$: To show how Theorem 6.1 reduces to Corollary 6.1 for $M = 3$, we apply Theorem 6.1 to the case of $M = 3$, $N = 2$, and the links to the user are BSC(p_1), and BSC(p_2). From Theorem 6.1, we can write the upper bound for the achievable retrieval rate as:

$$R(\mathbf{C}) \leq \max_{\boldsymbol{\tau}\in\mathbb{T}} \min_{n_i\in\{1,2\}} \frac{\sum_{n=1}^{N} \tau_n C_n + \frac{\sum_{n=n_1+1}^{N} \tau_n C_n}{n_1} + \frac{\sum_{n=n_2+1}^{N} \tau_n C_n}{n_1 n_2}}{1 + \frac{1}{n_1} + \frac{1}{n_1 n_2}} \tag{6.18}$$

where $C_n = 1 - H(p_n)$.

By observing $\tau_2 = 1 - \tau_1$ and the fact that C_n is monotonically decreasing in p_n for $p_n \in (0, \frac{1}{2})$ (which implies that $p_1 \leq p_2$ satisfies $C_1 \geq C_2$), (6.18) can be explicitly written as the following linear program:

$$\max_{\tau_2, R} \quad R$$

$$\begin{aligned}
\text{s.t.} \quad & R \leq \frac{1}{3}(1-H(p_1)) + \left[(1-H(p_2)) - \frac{1}{3}(1-H(p_1))\right]\tau_2 \\
& R \leq \frac{2}{5}(1-H(p_1)) + \left[\frac{4}{5}(1-H(p_2)) - \frac{2}{5}(1-H(p_1))\right]\tau_2 \\
& R \leq \frac{4}{7}(1-H(p_1)) + \left[\frac{4}{7}(1-H(p_2)) - \frac{4}{7}(1-H(p_1))\right]\tau_2 \\
& 0 \leq \tau_2 \leq 1
\end{aligned} \tag{6.19}$$

The bound corresponding to $n_1 = 2$, $n_2 = 1$ is inactive for all values of (p_1, p_2). Since (6.19) is a linear program, its solution resides at the corner points of the feasible region. The first corner point occurs at $\tau_2^{(1)} = 0$, which corresponds to the upper bound $R \leq \frac{1-H(p_1)}{3}$. The second corner point is at the intersection of the first two constraints, i.e.,

$$\begin{aligned}
\frac{1}{3}(1-H(p_1)) + \left[(1-H(p_2)) - \frac{1}{3}(1-H(p_1))\right]\tau_2^{(2)} \\
= \frac{2}{5}(1-H(p_1)) + \left[\frac{4}{5}(1-H(p_2)) - \frac{2}{5}(1-H(p_1))\right]\tau_2^{(2)}
\end{aligned} \tag{6.20}$$

which leads to,

$$\tau_2^{(2)} = \frac{1-H(p_1)}{3(1-H(p_2)) + (1-H(p_1))} \tag{6.21}$$

which corresponds to the upper bound $R \leq \frac{2}{\frac{3}{1-H(p_1)} + \frac{1}{1-H(p_2)}}$. Similarly, by observing the intersection between the last two constraints, we have the following upper bound $R \leq \frac{4}{\frac{4}{1-H(p_1)} + \frac{3}{1-H(p_2)}}$, which is achieved at $\tau_2^{(3)} = \frac{3(1-H(p_1))}{4(1-H(p_2))+3(1-H(p_1))}$. Consequently,

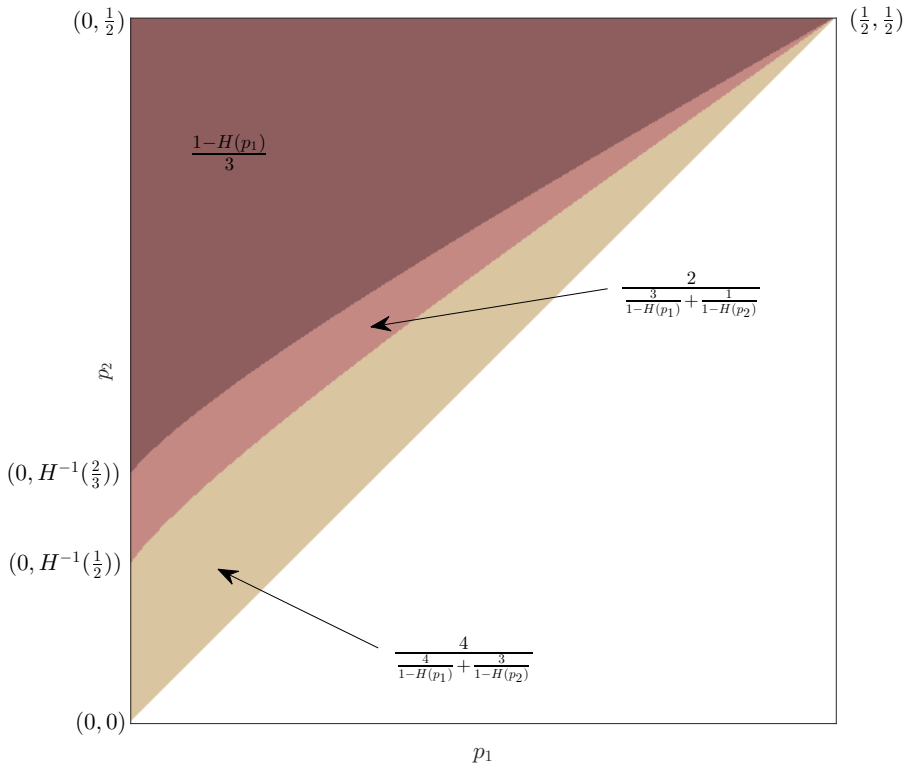


Figure 6.2: Partitions of (p_1, p_2) space according to retrieval rate expression for $M = 3, N = 2$.

an explicit upper bound for the retrieval rate is:

$$R \leq \max \left\{ \frac{1 - H(p_1)}{3}, \frac{2}{\frac{3}{1-H(p_1)} + \frac{1}{1-H(p_2)}}, \frac{4}{\frac{4}{1-H(p_1)} + \frac{3}{1-H(p_2)}} \right\} \tag{6.22}$$

In Section 6.5.1, we will show how these rates can be achieved, hence (6.22) is the exact capacity. This capacity result is illustrated in Fig. 6.2. The figure shows the partitioning of the (p_1, p_2) (by convention $p_1 \leq p_2$) space according to the active capacity expression. When the ratio $2 < \frac{1-H(p_1)}{1-H(p_2)} \leq 3$, $C_{\text{PIR}}(p_1, p_2) = \frac{2}{\frac{3}{1-H(p_1)}+\frac{1}{1-H(p_2)}}$. When the ratio $\frac{1-H(p_1)}{1-H(p_2)} \leq 2$, $C_{\text{PIR}}(p_1, p_2) = \frac{4}{\frac{4}{1-H(p_1)}+\frac{3}{1-H(p_2)}}$, otherwise, $C_{\text{PIR}}(p_1, p_2) = \frac{1-H(p_1)}{3}$. Interestingly, Fig. 6.2 shows that the dominant strategy for most (p_1, p_2)

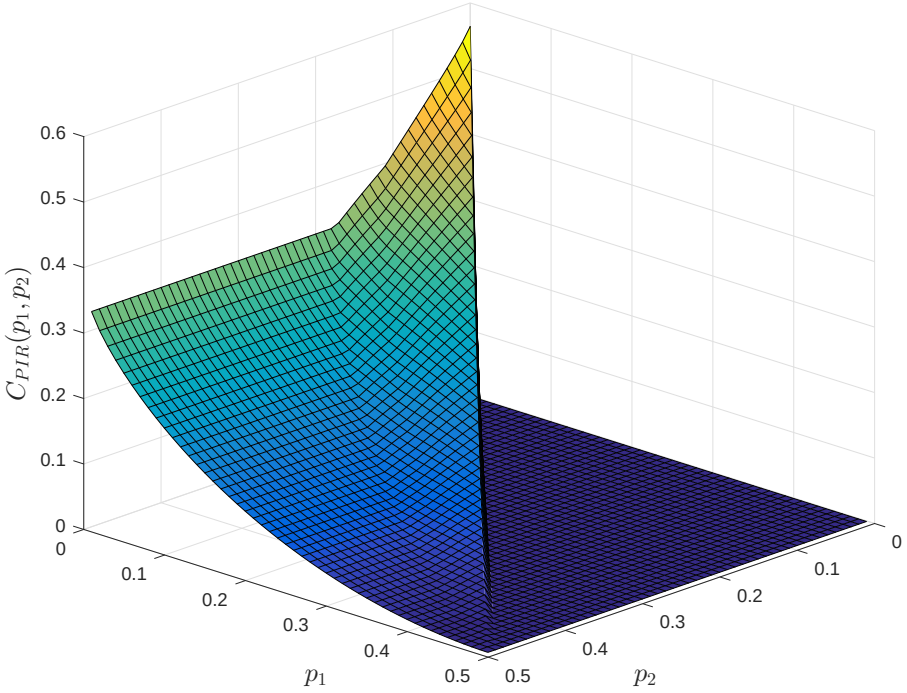
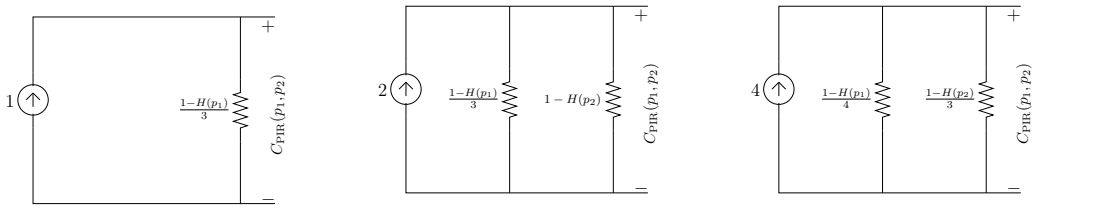


Figure 6.3: Capacity function $C_{\text{PIR}}(p_1, p_2)$ for $M = 3$, $N = 2$.

pairs is to rely only on database 1 for the retrieval process. The capacity function $C_{\text{PIR}}(p_1, p_2)$ is shown in Fig. 6.3. The figure shows that the maximum value for the capacity is $C_{\text{PIR}}(0, 0) = \frac{4}{7}$, which is consistent with [12]. The figure also shows that $C_{\text{PIR}}(0.5, 0.5) = 0$, as the answer strings become independent of the user queries. We observe that $C_{\text{PIR}}(0, p_2) = \frac{1}{3}$ for $p_2 \geq H^{-1}(\frac{2}{3}) = 0.1737$, since the retrieval is performed only from database 1, which is connected to the user via a noiseless link.

Remark 6.5 *We will show in Section 6.5 that channel coding and retrieval schemes for NPIR are almost separable. Nevertheless, the final capacity expression couples the capacity of the noisy channels and the retrieval rates from databases with noiseless links in a non-trivial way. We illustrate the capacity expression in* (6.22) *by*



(a) $C_{\mathrm{PIR}} = \frac{1-H(p_1)}{3}$ (b) $C_{\mathrm{PIR}} = \frac{2}{\frac{3}{1-H(p_1)}+\frac{1}{1-H(p_2)}}$ (c) $C_{\mathrm{PIR}} = \frac{4}{\frac{4}{1-H(p_1)}+\frac{3}{1-H(p_2)}}$

Figure 6.4: Circuit analogy for the capacity expression of PIR from BSC(p_1), BSC(p_2).

means of circuit theory analogy in Fig. 6.4. The current from the current source represents the number of desired bits, the voltage across the current source corresponds to the achievable retrieval rate, and the channel effect of the link connected to the nth database is abstracted via a parallel resistor, whose value depends on the capacity of the channel and the total download from the nth database. Intuitively, to maximize the retrieval rate, the user chooses one of the three circuits in Fig. 6.4. The circuits are arranged ascendingly in the number of the desired bits (namely, 1, 2, 4 bits), while the values of the resistors decrease, as the total download increases and/or due to adding extra parallel branch. This results in a tension between conveying more desired bits and decreasing the equivalent resistor of the circuit. The capacity-achieving scheme is the one which maximizes the product of these contradictory effects (i.e., the voltage).

6.4 Converse Proof for NPIR

In this section, we derive a general upper bound for the NPIR problem. The main idea of the converse hinges on the fact that the traffic from the databases should be dependent on the relative channel qualities (i.e., channel capacities) of the response

channels. Thus, we extend the converse proof in Chapter 5 to account for the noisy observations.

We will need the following lemma, which characterizes the channel effect on the noisy answer strings. The lemma states that the remaining uncertainty on a subset of answer strings after revealing the queries and the message set is a sum of single-letter conditional entropies of the noisy channels over the lengths of the answer strings. The lemma is a consequence of the Markov chain $(W_{1:M}, Q_{1:N}^{[m]}, \tilde{A}_{1:n-1}^{[m]}) \to A_n^{[m]} \to \tilde{A}_n^{[m]}$.

Lemma 6.1 (Channel effect) *For any subset $\mathcal{S} \subseteq \{1, \cdots, N\}$ for all $m \in \{1, \cdots, M\}$, the remaining uncertainty on the noisy answer strings $\tilde{A}_{\mathcal{S}}^{[m]}$ given $(W_{1:M}, Q_{1:N}^{[m]})$ is given by,*

$$H(\tilde{A}_{\mathcal{S}}^{[m]}|W_{1:M}, Q_{1:N}^{[m]}) = \sum_{n\in\mathcal{S}} \sum_{\eta_n=1}^{t_n} H\left(Y_{n,\eta_n}^{[m]}|X_{n,\eta_n}^{[m]}\right) \tag{6.23}$$

Furthermore, (6.23) *is true if conditioned on the complementary subset of the noisy answer strings $\tilde{A}_{\bar{\mathcal{S}}}^{[m]}$, i.e.,*

$$H(\tilde{A}_{\mathcal{S}}^{[m]}|W_{1:M}, Q_{1:N}^{[m]}, \tilde{A}_{\bar{\mathcal{S}}}^{[m]}) = \sum_{n\in\mathcal{S}} \sum_{\eta_n=1}^{t_n} H\left(Y_{n,\eta_n}^{[m]}|X_{n,\eta_n}^{[m]}\right) \tag{6.24}$$

where $\bar{\mathcal{S}} = \{1, \cdots, N\} \setminus \mathcal{S}$.

Proof: We start with the left hand side of (6.23),

$$H(\tilde{A}_{\mathcal{S}}^{[m]}|W_{1:M},Q_{1:N}^{[m]}) = \sum_{n\in\mathcal{S}} H(\tilde{A}_n^{[m]}|\tilde{A}_{1:n-1}^{[m]},W_{1:M},Q_{1:N}^{[m]}) \tag{6.25}$$

$$\overset{(6.4)}{=} \sum_{n\in\mathcal{S}} H(\tilde{A}_n^{[m]}|\tilde{A}_{1:n-1}^{[m]},W_{1:M},Q_{1:N}^{[m]},A_n^{[m]}) \tag{6.26}$$

$$= \sum_{n\in\mathcal{S}} H(\tilde{A}_n^{[m]}|A_n^{[m]}) \tag{6.27}$$

$$= \sum_{n\in\mathcal{S}} \sum_{\eta_n=1}^{t_n} H(Y_{n,\eta_n}^{[m]}|X_{n,1}^{[m]},\cdots,X_{n,t_n}^{[m]},Y_{n,1},\cdots,Y_{n,\eta_n-1}^{[m]}) \tag{6.28}$$

$$\overset{(6.5)}{=} \sum_{n\in\mathcal{S}} \sum_{\eta_n=1}^{t_n} H(Y_{n,\eta_n}^{[m]}|X_{n,\eta_n}^{[m]}) \tag{6.29}$$

where (6.26) follows from the fact that $A_n^{[m]}$ is a deterministic function of $(W_{1:M},Q_n^{[m]})$, (6.27) follows from the fact that $(W_{1:M},Q_{1:N}^{[m]},\tilde{A}_{1:n-1}^{[m]}) \to A_n^{[m]} \to \tilde{A}_n^{[m]}$ is a Markov chain, (6.29) follows from the fact that the channel is memoryless.

The proof of (6.24) follows similarly by observing that $(W_{1:M},Q_{1:N}^{[m]},\tilde{A}_{1:n-1}^{[m]},\tilde{A}_{\bar{\mathcal{S}}}^{[m]}) \to A_n^{[m]} \to \tilde{A}_n^{[m]}$ is a Markov chain as well. ■

We need the following lemma which upper bounds the mutual information between the noisy answer strings and the interfering messages with a linear function of the channel capacities.

Lemma 6.2 (Noisy interference bound) *For NPIR, the mutual information between the interfering messages $W_{2:M}$ and the noisy answer strings $\tilde{A}_{1:N}^{[1]}$ given the desired message W_1 is upper bounded by,*

$$I\left(W_{2:M};Q_{1:N}^{[1]},\tilde{A}_{1:N}^{[1]}|W_1\right) \leq \sum_{n=1}^{N} t_n C_n - L + o(L) \tag{6.30}$$

Proof: We start with the left hand side of (6.30),

$$
\begin{aligned}
&I(W_{2:M}; Q_{1:N}^{[1]}, \tilde{A}_{1:N}^{[1]} | W_1) \\
&\quad \overset{(6.2)}{=} I\left(W_{2:M}; W_1, Q_{1:N}^{[1]}, \tilde{A}_{1:N}^{[1]}\right) && (6.31)\\
&\quad = I\left(W_{2:M}; Q_{1:N}^{[1]}, \tilde{A}_{1:N}^{[1]}\right) + I\left(W_{2:M}; W_1 | Q_{1:N}^{[1]}, \tilde{A}_{1:N}^{[1]}\right) && (6.32)\\
&\quad \overset{(6.9)}{\leq} I\left(W_{2:M}; Q_{1:N}^{[1]}, \tilde{A}_{1:N}^{[1]}\right) + o(L) && (6.33)\\
&\quad \overset{(6.3)}{=} I\left(W_{2:M}; \tilde{A}_{1:N}^{[1]} | Q_{1:N}^{[1]}\right) + o(L) && (6.34)\\
&\quad = H\left(\tilde{A}_{1:N}^{[1]} | Q_{1:N}^{[1]}\right) - H\left(\tilde{A}_{1:N}^{[1]} | W_{2:M}, Q_{1:N}^{[1]}\right) + o(L) && (6.35)\\
&\quad = H\left(\tilde{A}_{1:N}^{[1]} | Q_{1:N}^{[1]}\right) - H\left(\tilde{A}_{1:N}^{[1]}, W_1 | W_{2:M}, Q_{1:N}^{[1]}\right) \\
&\qquad + H\left(W_1 | W_{2:M}, Q_{1:N}^{[1]}, \tilde{A}_{1:N}^{[1]}\right) + o(L) && (6.36)\\
&\quad \overset{(6.9)}{\leq} H\left(\tilde{A}_{1:N}^{[1]} | Q_{1:N}^{[1]}\right) - H\left(\tilde{A}_{1:N}^{[1]}, W_1 | W_{2:M}, Q_{1:N}^{[1]}\right) + o(L) && (6.37)\\
&\quad = H\left(\tilde{A}_{1:N}^{[1]} | Q_{1:N}^{[1]}\right) - H\left(W_1 | W_{2:M}, Q_{1:N}^{[1]}\right) - H\left(\tilde{A}_{1:N}^{[1]} | W_{1:M}, Q_{1:N}^{[1]}\right) + o(L) \\
& && (6.38)\\
&\quad \overset{(6.23)}{\leq} \sum_{n=1}^{N} \sum_{\eta_n=1}^{t_n} \left[H\left(Y_{n,\eta_n}^{[1]}\right) - H\left(Y_{n,\eta_n}^{[1]} | X_{n,\eta_n}^{[1]}\right)\right] - L + o(L) && (6.39)\\
&\quad = \sum_{n=1}^{N} \sum_{\eta_n=1}^{t_n} I\left(X_{n,\eta_n}^{[1]}; Y_{n,\eta_n}^{[1]}\right) - L + o(L) && (6.40)\\
&\quad \leq \sum_{n=1}^{N} t_n C_n - L + o(L) && (6.41)
\end{aligned}
$$

where (6.31) follows from the independence of the messages, (6.33), (6.37) follow from the decodability of W_1 given $(Q_{1:N}^{[1]}, \tilde{A}_{1:N}^{[1]})$, (6.34) follows from the independence of $(W_{2:M}, Q_{1:N}^{[1]})$, (6.39) follows from the independence of $(W_1, W_{2:M}, Q_{1:N}^{[1]})$, Lemma 6.1, and the fact that conditioning cannot increase entropy, (6.41) follows

from the fact that $I\left(X_{n,\eta_n}^{[m]}; Y_{n,\eta_n}^{[m]}\right) \leq C_n$ by the definition of the nth channel capacity. ■

Finally, in order to capture the recursive structure of the problem in terms of the messages and to express the potential asymmetry of the optimal scheme, we will need the following lemma, which inductively lower bounds the mutual information term in Lemma 6.2. The lemma implies that n_{m-1} databases can apply a symmetric scheme when the retrieval problem is reduced to retrieving message W_{m-1} from the set of $W_{m-1:M}$ messages. For the remaining answer strings, we directly bound them by their corresponding length of the unobserved portion $\sum_{n=n_{m-1}+1}^{N} t_n C_n$.

Lemma 6.3 (Noisy induction lemma) *For all $m \in \{2, \ldots, M\}$ and for an arbitrary $n_{m-1} \in \{1, \cdots, N\}$, the mutual information term in Lemma 6.2 can be inductively lower bounded as,*

$$I\left(W_{m:M}; Q_{1:N}^{[m-1]}, \tilde{A}_{1:N}^{[m-1]} | W_{1:m-1}\right) \geq \frac{1}{n_{m-1}} I\left(W_{m+1:M}; Q_{1:N}^{[m]}, \tilde{A}_{1:N}^{[m]} | W_{1:m}\right) + \frac{1}{n_{m-1}}\left(L - \sum_{n=n_{m-1}+1}^{N} t_n C_n\right) - \frac{o(L)}{n_{m-1}} \tag{6.42}$$

Proof: We start with the left hand side of (6.42) after multiplying by n_{m-1},

$$\begin{aligned} & n_{m-1}\, I\left(W_{m:M}; Q_{1:N}^{[m-1]}, \tilde{A}_{1:N}^{[m-1]} | W_{1:m-1}\right) \\ & \quad \geq n_{m-1}\, I\left(W_{m:M}; Q_{1:n_{m-1}}^{[m-1]}, \tilde{A}_{1:n_{m-1}}^{[m-1]} | W_{1:m-1}\right) && (6.43) \\ & \quad \geq \sum_{n=1}^{n_{m-1}} I\left(W_{m:M}; Q_{n}^{[m-1]}, \tilde{A}_{n}^{[m-1]} | W_{1:m-1}\right) && (6.44) \end{aligned}$$

$$\overset{(6.8)}{=} \sum_{n=1}^{n_{m-1}} I\left(W_{m:M}; Q_n^{[m]}, \tilde{A}_n^{[m]} | W_{1:m-1}\right) \tag{6.45}$$

$$\overset{(6.3)}{=} \sum_{n=1}^{n_{m-1}} I\left(W_{m:M}; \tilde{A}_n^{[m]} | Q_n^{[m]}, W_{1:m-1}\right) \tag{6.46}$$

$$= \sum_{n=1}^{n_{m-1}} H\left(\tilde{A}_n^{[m]} | Q_n^{[m]}, W_{1:m-1}\right) - H\left(\tilde{A}_n^{[m]} | Q_n^{[m]}, W_{1:M}\right) \tag{6.47}$$

$$\geq \sum_{n=1}^{n_{m-1}} H\left(\tilde{A}_n^{[m]} | \tilde{A}_{1:n-1}^{[m]}, Q_{1:n_{m-1}}^{[m]}, W_{1:m-1}\right) - H\left(\tilde{A}_n^{[m]} | \tilde{A}_{1:n-1}^{[m]}, Q_{1:n_{m-1}}^{[m]}, W_{1:M}\right)$$

$$\tag{6.48}$$

$$= \sum_{n=1}^{n_{m-1}} I\left(W_{m:M}; \tilde{A}_n^{[m]} | \tilde{A}_{1:n-1}^{[m]}, Q_{1:n_{m-1}}^{[m]}, W_{1:m-1}\right) \tag{6.49}$$

$$= I\left(W_{m:M}; \tilde{A}_{1:n_{m-1}}^{[m]} | Q_{1:n_{m-1}}^{[m]}, W_{1:m-1}\right) \tag{6.50}$$

$$\overset{(6.3)}{=} I\left(W_{m:M}; Q_{1:n_{m-1}}^{[m]}, \tilde{A}_{1:n_{m-1}}^{[m]} | W_{1:m-1}\right) \tag{6.51}$$

$$\begin{aligned}\overset{(6.3),(6.4)}{=}\ & I\left(W_{m:M}; Q_{1:N}^{[m]}, \tilde{A}_{1:N}^{[m]} | W_{1:m-1}\right) \\ & - I\left(W_{m:M}; \tilde{A}_{n_{m-1}+1:N}^{[m]} | Q_{1:N}^{[m]}, \tilde{A}_{1:n_{m-1}}^{[m]}, W_{1:m-1}\right)\end{aligned} \tag{6.52}$$

$$\begin{aligned}=\ & I\left(W_{m:M}; Q_{1:N}^{[m]}, \tilde{A}_{1:N}^{[m]} | W_{1:m-1}\right) - H\left(\tilde{A}_{n_{m-1}+1:N}^{[m]} | Q_{1:N}^{[m]}, \tilde{A}_{1:n_{m-1}}^{[m]}, W_{1:m-1}\right) \\ & + H\left(\tilde{A}_{n_{m-1}+1:N}^{[m]} | Q_{1:N}^{[m]}, \tilde{A}_{1:n_{m-1}}^{[m]}, W_{1:M}\right)\end{aligned} \tag{6.53}$$

$$\begin{aligned}\overset{(6.24)}{\geq}\ & I\left(W_{m:M}; Q_{1:N}^{[m]}, \tilde{A}_{1:N}^{[m]} | W_{1:m-1}\right) \\ & - \sum_{n=n_{m-1}+1}^{N} \sum_{\eta_n=1}^{t_n} \left[H\left(Y_{n,\eta_n}^{[m]}\right) - H\left(Y_{n,\eta_n}^{[m]} | X_{n,\eta_n}^{[m]}\right)\right]\end{aligned} \tag{6.54}$$

$$\overset{(6.9)}{\geq} I\left(W_{m:M}; W_m, Q_{1:N}^{[m]}, \tilde{A}_{1:N}^{[m]} | W_{1:m-1}\right) - \sum_{n=n_{m-1}+1}^{N} \sum_{\eta_n=1}^{t_n} I\left(X_{n,\eta_n}^{[m]}; Y_{n,\eta_n}^{[m]}\right) - o(L)$$

$$\tag{6.55}$$

$$\begin{aligned}=\ & I\left(W_{m:M}; W_m | W_{1:m-1}\right) + I\left(W_{m:M}; Q_{1:N}^{[m]}, \tilde{A}_{1:N}^{[m]} | W_{1:m}\right) \\ & - \sum_{n=n_{m-1}+1}^{N} \sum_{\eta_n=1}^{t_n} I\left(X_{n,\eta_n}^{[m]}; Y_{n,\eta_n}^{[m]}\right) - o(L)\end{aligned} \tag{6.56}$$

$$= I\left(W_{m+1:M}; Q_{1:N}^{[m]}, \tilde{A}_{1:N}^{[m]} | W_{1:m}\right) + \left(L - \sum_{n=n_{m-1}+1}^{N} \sum_{\eta_n=1}^{t_n} I\left(X_{n,\eta_n}^{[m]}; Y_{n,\eta_n}^{[m]}\right)\right) - o(L) \tag{6.57}$$

$$\geq I\left(W_{m+1:M}; Q_{1:N}^{[m]}, \tilde{A}_{1:N}^{[m]} | W_{1:m}\right) + \left(L - \sum_{n=n_{m-1}+1}^{N} t_n C_n\right) - o(L) \tag{6.58}$$

where (6.43), (6.44) follow from the non-negativity of mutual information, (6.45) follows from the privacy constraint, (6.46) follows from the independence of $\left(W_{m:M}, Q_n^{[m]}\right)$, (6.48) follows from the fact that conditioning cannot increase entropy and from the fact that $(W_{1:M}, Q_{1:n_{m-1}}^{[m]}, \tilde{A}_{1:n-1}^{[m]}) \to (W_{1:M}, Q_n^{[m]}) \to \tilde{A}_n^{[m]}$ forms a Markov chain, (6.51) follows from the independence of the messages and the queries, (6.52) follows from the chain rule, the independence of the queries and the messages, and the fact that $Q_{1:N}^{[m]} \to Q_{1:n_{m-1}}^{[m]} \to \tilde{A}_{1:n_{m-1}}^{[m]}$ forms a Markov chain by (6.4), (6.54) follows from the fact that conditioning reduces entropy and Lemma 6.1, (6.55) follows from the reliability constraint, (6.58) follows from the definition of the channel capacity. Finally, dividing both sides by n_{m-1} leads to (6.42). ■

Now, we are ready to derive an explicit upper bound for the retrieval rate from noisy channels. Fixing the length of the nth answer string to t_n and applying Lemma 6.2 and Lemma 6.3 successively for an arbitrary sequence $\{n_i\}_{i=1}^{M-1} \subset \{1, \cdots, N\}^{M-1}$, we have the following,

$$\begin{aligned} &\sum_{n=1}^{N} t_n C_n - L + \tilde{o}(L) \\ &\overset{(6.30)}{\geq} I\left(W_{2:M}; Q_{1:N}^{[1]}, \tilde{A}_{1:N}^{[1]} | W_1\right) \end{aligned} \tag{6.59}$$

$$\overset{(6.42)}{\geq} \frac{1}{n_1}\left(L-\sum_{n=n_1+1}^{N} t_n C_n\right) + \frac{1}{n_1} I\left(W_{3:M}; Q_{1:N}^{[2]}, \tilde{A}_{1:N}^{[2]} | W_{1:2}\right) \tag{6.60}$$

$$\overset{(6.42)}{\geq} \frac{1}{n_1}\left(L-\sum_{n=n_1+1}^{N} t_n C_n\right) + \frac{1}{n_1 n_2}\left(L-\sum_{n=n_2+1}^{N} t_n C_n\right) + \frac{1}{n_2} I\left(W_{4:M}; Q_{1:N}^{[3]}, \tilde{A}_{1:N}^{[3]} | W_{1:3}\right) \tag{6.61}$$

$$\overset{(6.42)}{\geq} \cdots$$

$$\overset{(6.42)}{\geq} \frac{1}{n_1}\left(L-\sum_{n=n_1+1}^{N} t_n C_n\right) + \frac{1}{n_1 n_2}\left(L-\sum_{n=n_2+1}^{N} t_n C_n\right) + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}\left(L-\sum_{n=n_{M-1}+1}^{N} t_n C_n\right) \tag{6.62}$$

where $\tilde{o}(L) = \left(1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}\right) o(L)$, (6.59) follows from Lemma 6.2, and the remaining bounding steps follow from successive application of Lemma 6.3.

Ordering terms, we have,

$$\left(1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}\right) L \leq \left(\theta(0) + \frac{\theta(n_1)}{n_1} + \cdots + \frac{\theta(n_{M-1})}{\prod_{i=1}^{M-1} n_i}\right) \sum_{n=1}^{N} t_n + \tilde{o}(L) \tag{6.63}$$

where $\theta(\ell) = \sum_{n=\ell+1}^{N} \tau_n C_n$

We conclude the proof by taking $L \to \infty$. Thus, for an arbitrary sequence $\{n_i\}_{i=1}^{M-1}$, we have

$$R(\boldsymbol{\tau}, \mathbf{C}) = \frac{L}{\sum_{n=1}^{N} t_n} \leq \frac{\theta(0) + \frac{\theta(n_1)}{n_1} + \frac{\theta(n_2)}{n_1 n_2} + \cdots + \frac{\theta(n_{M-1})}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{6.64}$$

Finally, we get the tightest bound by minimizing over the sequence $\{n_i\}_{i=1}^{M-1}$ over

the set $\{1, \cdots, N\}$, as

$$R(\boldsymbol{\tau}, \mathbf{C}) \leq \min_{n_i \in \{1,\cdot,N\}} \frac{\theta(0) + \frac{\theta(n_1)}{n_1} + \frac{\theta(n_2)}{n_1 n_2} + \cdots + \frac{\theta(n_{M-1})}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{6.65}$$

$$= \min_{n_i \in \{1,\cdot,N\}} \frac{\sum_{n=1}^{N} \tau_n C_n + \frac{\sum_{n=n_1+1}^{N} \tau_n C_n}{n_1} + \frac{\sum_{n=n_2+1}^{N} \tau_n C_n}{n_1 n_2} + \cdots + \frac{\sum_{n=n_{M-1}+1}^{N} \tau_n C_n}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{6.66}$$

The user and the databases can agree on a traffic ratio vector $\boldsymbol{\tau} \in \mathbb{T} = \{(\tau_1, \cdots, \tau_N) : \tau_n \geq 0 \, \forall n, \sum_{n=1}^{N} \tau_n = 1\}$ that maximizes $R(\boldsymbol{\tau}, \mathbf{C})$, hence the retrieval rate $R(\mathbf{C})$ is upper bounded by,

$$R(\mathbf{C}) \leq \max_{\boldsymbol{\tau} \in \mathbb{T}} \quad R(\boldsymbol{\tau}, \mathbf{C}) \tag{6.67}$$

$$= \max_{\boldsymbol{\tau} \in \mathbb{T}} \min_{n_i \in \{1,\cdots,N\}} \frac{\sum_{n=1}^{N} \tau_n C_n + \frac{\sum_{n=n_1+1}^{N} \tau_n C_n}{n_1} + \frac{\sum_{n=n_2+1}^{N} \tau_n C_n}{n_1 n_2} + \cdots + \frac{\sum_{n=n_{M-1}+1}^{N} \tau_n C_n}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{6.68}$$

6.5 Achievability Proof for NPIR

In this section, we present the achievability proof for the NPIR problem. We show that by means of the random coding argument, each database can independently encode its response such that the probability of error can be made vanishingly small. The databases use the uncoded responses as an indexing mechanism for

choosing codewords from a randomly generated codebook. The uncoded responses, which are the truthful responses to the user queries, vary in length to maximize the retrieval rate. The query structure builds on the achievability proofs for PIR under asymmetric traffic constraints [125].

6.5.1 Motivating Example: $M = 3$, $N = 2$, via BSC(p_1), BSC(p_2)

We illustrate the retrieval scheme for $N = 2$ databases, $M = 3$ messages when the answer strings pass through BSC(p_1) and BSC(p_2). We show that the channel coding (using linear block codes) is *almost separable* from the retrieval scheme (which hinges on the result of [125]). We begin with the case when $(p_1, p_2) = (0.1, 0.2)$, then we extend this technique for all (p_1, p_2) pairs. We will need the following lemma, which shows the achievability of Shannon's channel coding theorem for BSC using linear block codes [122, Theorem 4.17, Corollary 4.18].

Lemma 6.4 (Shannon's coding theorem for BSC [122]) *For BSC(p) with crossover probability $p \in (0, \frac{1}{2})$. Let n, k be integers such that $R = \frac{k}{n} < 1 - H(p)$, and let $\mathbb{E}_{\mathcal{C}}[P_e(\mathcal{C})]$ denote the expected probability of error $P_e(\mathcal{C})$ calculated over all linear $[n, k]$ codes $\mathcal{C}$, assuming a nearest-codeword decoder. Then,*

$$\mathbb{E}_{\mathcal{C}}[P_e(\mathcal{C})] < 2 \cdot 2^{-n\Delta(p,R)} \tag{6.69}$$

for some $\Delta(p, R) > 0$. Moreover, for all $\rho \in (0, 1]$, all but less than ρ of the linear

$[n, k]$ *codes satisfy,*

$$P_e(\mathcal{C}) < \frac{2}{\rho} \cdot 2^{-n\Delta(p,R)} \tag{6.70}$$

The result implies that as long as the rate of the linear $[n, k]$ code is strictly less than the capacity, then there exists a linear $[n, k]$ code with exponentially decreasing probability of error in n with high probability.

6.5.1.1 Achievable Scheme for BSC(0.1), BSC(0.2)

Now, we focus on the case when $(p_1, p_2) = (0.1, 0.2)$. Using the explicit upper bound in (6.22), we infer that $R \leq \frac{4}{\frac{4}{1-H(p_1)}+\frac{3}{1-H(p_2)}}$ which is 0.2183 for $p_1 = 0.1$, $p_2 = 0.2$. To operate at $\tau_2 = \tau_2^{(3)} = \frac{3(1-H(p_1))}{4(1-H(p_2))+3(1-H(p_1))}$, we enforce the ratio between the uncoded traffic, i.e., before channel coding, to be $4 : 3$. This results in coded traffic ratio of $\frac{4}{1-H(p_1)} : \frac{3}{1-H(p_2)}$, which appears in the denominator of the upper bound. Concurrently, this results in retrieving 4 desired bits per scheme repetition, which appears in the numerator.

To that end, the user repeats the following retrieval scheme for ν times. Each repetition of the scheme operates over blocks of $L^* = 4$ bits from all messages $W_{1:3}$. The user permutes the indices of the bits of each message independently and uniformly. Let $a_i(j)$, $b_i(j)$, $c_i(j)$ denote the ith bit of block j from the permuted message W_1, W_2, W_3, respectively. Assume without loss of generality that the desired file is W_1. In block j, the user requests to download a single bit from each message from database 1, i.e., the user requests to download $a_1(j)$, $b_1(j)$, and $c_1(j)$

from database 1. From database 2, the user exploits the side information generated from database 1 by requesting to download the sums $a_2(j) + b_1(j)$, $a_3(j) + c_1(j)$, and $b_2(j) + c_2(j)$. Finally, the user exploits the side information generated from database 1 by downloading $a_4(j) + b_2(j) + c_2(j)$ from database 2. The query table for the jth block is summarized in Table 6.1. Denote the number of uncoded bits requested from the nth database by D_n, then $D_1 = 4$, $D_2 = 3$. This guarantees that the ratio between the uncoded traffic is $4 : 3$ (for any number of repetitions ν). This query structure is private, as all combinations of the sums are included in the queries and the indices of the message bits are uniformly and independently permuted for each block of messages (which operate on different set of bits), the privacy constraint is satisfied.

Table 6.1: The query table for the jth block of $M = 3$, $N = 2$, $p_1 = 0.1$, $p_2 = 0.2$.

Database 1	Database 2
$a_1(j)$	$a_2(j) + b_1(j)$
$b_1(j)$	$a_3(j) + c_1(j)$
$c_1(j)$	$b_2(j) + c_2(j)$
$a_4(j) + b_2(j) + c_2(j)$	

After receiving the queries of the user, the nth database concatenates the uncoded binary answer strings into a vector $U_n^{[1]}$ of length νD_n, i.e.,

$$\begin{aligned} U_1^{[1]} =&[a_1(1) \quad b_1(1) \quad c_1(1) \quad a_4(1) + b_2(1) + c_2(1) \\ &\cdots \quad a_1(\nu) \quad b_1(\nu) \quad c_1(\nu) \quad a_4(\nu) + b_2(\nu) + c_2(\nu)]^T \end{aligned} \tag{6.71}$$

$$\begin{aligned} U_2^{[1]} =&[a_2(1) + b_1(1) \quad a_3(1) + c_1(1) \quad b_2(1) + c_2(1) \\ &\cdots \quad a_2(\nu) + b_1(\nu) \quad a_3(\nu) + c_1(\nu) \quad b_2(\nu) + c_2(\nu)]^T \end{aligned} \tag{6.72}$$

The nth database encodes the vector $U_n^{[1]}$ to a coded answer string $A_n^{[1]}$ of length t_n using a $(t_n, \nu D_n)$ linear block code (which belongs to the set of good codes that satisfy (6.70)) such that:

$$t_n = \left\lceil \frac{\nu D_n}{1 - H(p_n)} \right\rceil \tag{6.73}$$

This ensures that $\frac{\nu D_n}{t_n} < 1 - H(p_n)$. The nth database responds with $A_n^{[1]}$ via the noisy channel BSC(p_n). The user receives the noisy answer string $\tilde{A}_n^{[1]}$ from the nth database.

To perform the decoding, the user employs the nearest-codeword decoder to find an estimate of $A_n^{[1]}$ based on the observation $\tilde{A}_n^{[1]}$. Since $\frac{\nu D_n}{t_n} < 1 - H(p_n)$, using Lemma 6.4 and the union bound, the probability of error in decoding is upper bounded by:

$$P_e(L) \leq P_e(\mathcal{C}_1) + P_e(\mathcal{C}_2) \tag{6.74}$$

$$\leq \frac{2}{\rho} \left[2^{-t_1 \Delta\left(p_1, \frac{\nu D_1}{t_1}\right)} + 2^{-t_2 \Delta\left(p_2, \frac{\nu D_2}{t_2}\right)} \right] \tag{6.75}$$

As $\nu \to \infty$, $L \to \infty$ and $t_n \to \infty$, we have $P_e(L) \to 0$. This ensures the decodability of $U_n^{[1]}$ with high probability. Since the vectors $U_1^{[1]}$, $U_2^{[2]}$ are designed to exploit the side information, the user can cancel the effect of the undesired messages and be left only with the correct W_1 with probability of error $P_e(L)$. This satisfies the reliability constraint.

Finally, we calculate the achievable retrieval rate. The retrieval scheme de-

codes $L = \nu L^* = 4\nu$ bits from the desired messages. The retrieval scheme downloads $t_n = \left\lceil \frac{\nu D_n}{1-H(p_n)} \right\rceil$ bits from the nth database, hence as $\nu \to \infty$, we have

$$R = \frac{L}{t_1 + t_2} \tag{6.76}$$

$$= \frac{\nu L^*}{\frac{\nu D_1}{1-H(p_1)} + \frac{\nu D_2}{1-H(p_2)}} \tag{6.77}$$

$$= \frac{4}{\frac{4}{1-H(p_1)} + \frac{3}{1-H(p_2)}} = 0.2183 \tag{6.78}$$

which matches the upper bound.

6.5.1.2 Achieving the Upper Bound for Arbitrary (p_1, p_2)

Now, we show that the upper bound in (6.22) is achievable for any (p_1, p_2). The idea is to design the uncoded response vectors $U_1^{[1]}$, $U_2^{[2]}$ such that the ratio of their traffic matches one of the corner points of the PIR problem under asymmetric traffic constraints as in Chapter 5.

For $R = \frac{1-H(p_1)}{3}$: For this rate, the user requests to download from database 1 only and does not access database 2. Thus, the user downloads all the contents of database 1 to satisfy the privacy constraint. Specifically, the user downloads $a_1(j), b_1(j), c_1(j)$ at the jth block of the retrieval process. Database 1 encodes the responses $U_1^{[1]}$ into t_1-length answer string using $(t_1, \nu D_1)$, where $D_1 = 3$, and $t_1 = \left\lceil \frac{\nu D_1}{1-H(p_1)} \right\rceil$. The user decodes ν desired symbols from ν repetitions with vanishingly small probability of error. Consequently, $R = \frac{1-H(p_1)}{3}$.

For $R = \frac{2}{\frac{3}{1-H(p_1)}+\frac{1}{1-H(p_2)}}$: For this rate, the user designs the queries such that the traffic ratio between the uncoded responses is $3:1$. Thus, in the jth block, the user requests to download one bit from each message, i.e., the user requests to download $a_1(j), b_1(j), c_1(j)$ from database 1. The user mixes the undesired information obtained from database 1 into one combined symbol $b_1(j)+c_1(j)$ and uses this symbol as a side information in database 2 by requesting to download $a_2(j)+b_1(j)+c_1(j)$. The query table for the jth block of the scheme is depicted in Table 6.2.

Table 6.2: The query table for the jth block of $M = 3$, $N = 2$ to achieve $R = \frac{2}{\frac{3}{1-H(p_1)}+\frac{1}{1-H(p_2)}}$

Database 1	Database 2
$a_1(j), b_1(j), c_1(j)$	$a_2(j)+b_1(j)+c_1(j)$

After repeating the retrieval process ν times, database 1 encodes the responses using a linear $(t_1, \nu D_1) = \left(\left\lceil \frac{3\nu}{1-H(p_1)} \right\rceil, 3\nu\right)$ code, while database 2 encodes its responses using a linear $(t_2, \nu D_2) = \left(\left\lceil \frac{\nu}{1-H(p_2)} \right\rceil, \nu\right)$ code. Using Lemma 4, the user can decode the correct W_1 with vanishingly small probability of error. The user decodes $L = 2\nu$ bits from W_1, hence, as $\nu \to \infty$

$$R = \frac{L}{t_1+t_2} = \frac{2}{\frac{3}{1-H(p_1)}+\frac{1}{1-H(p_2)}} \tag{6.79}$$

For $R = \frac{4}{\frac{4}{1-H(p_1)}+\frac{3}{1-H(p_2)}}$: An instance for this scheme is the $(p_1, p_2) = (0.1, 0.2)$ example. Please refer to Section 6.5.1.1 for the details.

Therefore, the capacity of the PIR problem from BSC(p_1), BSC(p_2) is given

by:

$$C_{\text{PIR}}(p_1, p_2) = \max \left\{ \frac{1 - H(p_1)}{3}, \frac{2}{\frac{3}{1-H(p_1)} + \frac{1}{1-H(p_2)}}, \frac{4}{\frac{4}{1-H(p_1)} + \frac{3}{1-H(p_2)}} \right\} \quad (6.80)$$

6.5.2 General Achievable Scheme

In this section, we present a general achievable scheme for the NPIR problem. The main idea of the scheme is to use the uncoded response from the nth database to user's query as an *index* for choosing the transmitted codeword from a codebook generated according to the optimal probability distribution. The query structure maps to one of the corner points of PIR under asymmetric traffic constraints [125] in order to maximize the retrieval rate.

Following the notations in [125], we denote the number of side information symbols that are used simultaneously in the initial round of downloads at the nth database by $s_n \in \{0, 1, \cdots, M-1\}$, e.g., if $s_n = 1$, then the user requests to download a sum of 1 desired symbol and 1 undesired symbol as a side information in the form of $a + b$, $a + c$, ... etc., while $s_n = 2$ implies that the user mixes every two undesired symbols to form one side information symbol, i.e., the user requests to download $a + b + c$, $a + c + d$, ... etc. For a given non-decreasing sequence $\{n_i\}_{i=0}^{M-1} \subset \{1, \cdots, N\}^M$, the databases are divided into groups, such that group 0 contains database 1 through database n_0, group 1 contains $n_1 - n_0$ databases starting from database $n_0 + 1$, and so on.

Hence, let $s_n = i$ for all $n_{i-1} + 1 \leq n \leq n_i$ with $n_{-1} = 0$ by convention.

Denote $\mathcal{S} = \{i : s_n = i \text{ for some } n \in \{1, \cdots, N\}\}$. We follow the round and stage definitions in [123]. The kth round is the download queries that admit a sum of k different messages (k-sum in [12]). A stage of the kth round is a query block of the kth round that exhausts all $\binom{M}{k}$ combinations of the k-sum. Denote $y_\ell[k]$ to be the number of stages in round k downloaded from the nth database, such that $n_{\ell-1} + 1 \leq n \leq n_\ell$. Our scheme is repeated for ν repetitions. Each repetition has the same query structure and operates over a block of message symbols of length L^*. Denote the total requested symbols from the nth database in one repetition of the scheme by $D_n(\mathbf{n})$. The details of the achievable scheme are as follows:

1. *Codebook construction:* According to the optimal probability distribution $p^*(x_n)$ (that maximizes the mutual information $I(X_n; Y_n)$), the nth database constructs a $\left(2^{\nu D_n(\mathbf{n})}, t_n(\mathbf{n})\right)$ codebook $\mathcal{C}_n$ at random, i.e., $p(x_{n,1}, \cdots, x_{n,t_n(\mathbf{n})}) = \prod_{\eta_n=1}^{t_n(\mathbf{n})} p^*(x_{n,\eta_n})$. Specifically, the codebook $\mathcal{C}_n$ can be written as:

$$\mathcal{C}_n = \begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_{t_n(\mathbf{n})}(1) \\ x_1(2) & x_2(2) & \cdots & x_{t_n(\mathbf{n})}(2) \\ \vdots & \vdots & \vdots & \vdots \\ x_1(2^{\nu D_n(\mathbf{n})}) & x_2(2^{\nu D_n(\mathbf{n})}) & \cdots & x_{t_n(\mathbf{n})}(2^{\nu D_n(\mathbf{n})}) \end{bmatrix}_{2^{\nu D_n(\mathbf{n})} \times t_n(\mathbf{n})} \tag{6.81}$$

where

$$t_n(\mathbf{n}) = \left\lceil \frac{\nu D_n(\mathbf{n})}{C_n} \right\rceil \tag{6.82}$$

This ensures that the rate of $\mathcal{C}_n$, $\frac{\nu D_n(\mathbf{n})}{t_n(\mathbf{n})} < C_n$ to ensure reliable transmission over the noisy channel. The nth database reveals the codebook $\mathcal{C}_n$ to the user.

2. *Initialization at the user side:* The user permutes each message independently and uniformly using a random interleaver, i.e.,

$$\omega_m(i) = W_m(\pi_m(i)), \quad i \in \{1, \cdots, L\} \tag{6.83}$$

where $\omega_m(i)$ is the ith symbol of the permuted W_m, $\pi_m(\cdot)$ is a random interleaver for the mth message that is chosen independently, uniformly, and privately at the user's side.

3. *Initial download:* From the nth database where $1 \leq n \leq n_0$, the user requests to download $\prod_{s\in\mathcal{S}} \binom{M-2}{s-1}$ symbols from the desired message. The user sets the round index $k = 1$. I.e., the user requests the desired symbols from $y_0[1] = \prod_{s\in\mathcal{S}} \binom{M-2}{s-1}$ different stages.

4. *Message symmetry:* To satisfy the privacy constraint, for each stage initiated in the previous step, the user completes the stage by requesting the remaining $\binom{M-1}{k-1}$ k-sum combinations that do not include the desired symbols, in particular, if $k = 1$, the user requests $\prod_{s\in\mathcal{S}} \binom{M-2}{s-1}$ individual symbols from each undesired message.

5. *Database symmetry:* We divide the databases into groups. Group $\ell \in \mathcal{S}$ corresponds to databases $n_{\ell-1}+1$ to n_ℓ. Database symmetry is applied within each group only. Consequently, the user repeats step 2 over each group of

databases, in particular, if $k = 1$, the user downloads $\prod_{s\in\mathcal{S}} \binom{M-2}{s-1}$ individual symbols from each message from the first n_0 databases (group 1).

6. *Exploitation of side information:* The undesired symbols downloaded within the kth round (the k-sums that do not include the desired message) are used as side information in the $(k+1)$th round. This exploitation of side information is performed by requesting to download $(k+1)$-sum consisting of 1 desired symbol and a k-sum of undesired symbols only that were generated in the kth round. Note that for the nth database, if $s_n > k$, then this database does not exploit the side information generated in the kth round. Consequently, the nth database belonging to the ℓth group exploits the side information generated in the kth round from all databases except itself if $s_n \leq k$. Moreover, for $s_n = k$, extra side information can be used in the nth database. This is due to the fact that the user can form $n_0 \prod_{s\in\mathcal{S}\setminus\{s_n\}} \binom{M-2}{s-1}$ extra stages of side information by constructing k-sums of the undesired symbols in round 1 from the databases in group 0.

7. *Repeat* steps 4, 5, 6 after setting $k = k+1$ until $k = M$.

8. *Repetition of the scheme:* Repeat steps $3, \cdots, 7$ for a total of ν repetitions.

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

10. *Encoding the responses to the user's queries:* The nth database responds to

the user queries truthfully. The nth database concatenates all the responses to the user's queries in a vector $U_n^{[i]}$ of length $\nu D_n(\mathbf{n})$. The nth database uses $U_n^{[i]}$ as an index for choosing a codeword from $\mathcal{C}_n$, i.e., the index of the codeword and $U_n^{[i]}$ should be in bijection (e.g., by transforming $U_n^{[i]}$ into a decimal value). Consequently, the nth database responds with,

$$A_n^{[i]} = [x_1(U_n^{[i]}) \quad x_1(U_n^{[i]}) \quad \cdots \quad x_{t_n(\mathbf{n})}(U_n^{[i]})]^T \tag{6.84}$$

6.5.3 Privacy, Reliability, and Achievable Rate

Privacy: The privacy of the scheme follows from the privacy of the inherent PIR scheme under asymmetric traffic constraints. Specifically, for every stage of the kth round initiated, all $\binom{M}{k}$ combinations of the k-sum are included at each round. Thus, the structure of the queries is the same for any desired message at any repetition of the achievable scheme. Due to the random and independent permutation of each message and the random shuffling of the order of the queries, all queries are equally likely independent of the desired message index, and thus the privacy constraint in (6.8) is guaranteed.

Reliability: The user employs *joint typicality decoder* for every noisy answer string $\tilde{A}_n^{[i]}$ to decode the codeword index. From the channel coding theorem [47, Theorem 7.7.1], for every rate $\frac{\nu D_n(\mathbf{n})}{t_n(\mathbf{n})} < C_n$, there exists a sequence of $(2^{\nu D_n(\mathbf{n})}, t_n(\mathbf{n}))$ with maximum probability of error $P_e(\mathcal{C}_n) \to 0$ as $t_n(\mathbf{n}) \to \infty$. By letting $\nu \to \infty$, we have $t_n(\mathbf{n}) \to \infty$, $\frac{\nu D_n(\mathbf{n})}{t_n(\mathbf{n})} < C_n$ and hence we ensure the existence of a good code such

that $P_e(\mathcal{C}_n) \to 0$. By union bound, the probability of error in decoding the indices of the codewords from every database is upper bounded by $P_e \leq \sum_{n=1}^{N} P_e(\mathcal{C}_n) \to 0$.

Since the index of the codeword is bijective to $U_n^{[i]}$, the probability of error in decoding $U_n^{[i]}$ for $n = 1, \cdots, N$ is vanishingly small. Now, by construction of the queries as in [125], all side information symbols used in the $(k+1)$th round are decodable in the kth round or from round 1, the user cancels out these side information and is left with symbols from the desired message. Consequently, there is no error in the decoding given that $U_n^{[i]}$ is correct for every n.

Achievable Rate: The structure of one repetition of our scheme is exactly as [125]. The recursive structure is described using the following system of difference equations that relate the number of stages in the databases belonging to a specific group as shown in [125, Theorem 2]:

$$\begin{aligned}
y_0[k] &= (n_0 - 1)y_0[k-1] + \sum_{j \in \mathcal{S}\setminus\{0\}} (n_j - n_{j-1})y_j[k-1] \\
y_1[k] &= (n_1 - n_0 - 1)y_1[k-1] + \sum_{j \in \mathcal{S}\setminus\{1\}} (n_j - n_{j-1})y_j[k-1] \\
y_\ell[k] &= n_0\xi_\ell\delta[k-\ell-1] + (n_\ell - n_{\ell-1} - 1)y_\ell[k-1] + \sum_{j \in \mathcal{S}\setminus\{\ell\}} (n_j - n_{j-1})y_j[k-1], \quad \ell \geq 2
\end{aligned} \tag{6.85}$$

where $y_\ell[k]$ is the number of stages in the kth round in a database belonging to the ℓth group, i.e., for the nth database, such that $n_{\ell-1} + 1 \leq n \leq n_\ell$.

To calculate $D_n(\mathbf{n})$ where $n_{\ell-1} \leq n \leq n_\ell$, we note that for any stage in the kth

round, the user downloads $\binom{M-1}{k-1}$ desired symbols from a total of $\binom{M}{k}$ downloads. Therefore,

$$D_n(\mathbf{n}) = \sum_{k=1}^{M} \binom{M}{k} y_\ell[k], \quad n_{\ell-1} \leq n \leq n_\ell \tag{6.86}$$

Thus, the total download $\sum_{n=1}^{N} t_n(\mathbf{n})$ from all databases from all repetitions is calculated by observing (6.82) and ignoring the ceiling operator as $\nu \to \infty$,

$$\sum_{n=1}^{N} t_n(\mathbf{n}) = \sum_{n=1}^{N} \frac{\nu D_n(\mathbf{n})}{C_n} \tag{6.87}$$

$$= \nu \left[\sum_{n=1}^{n_0} \frac{\sum_{k=1}^{M} \binom{M}{k} y_0[k]}{C_n} + \sum_{n=n_0+1}^{n_1} \frac{\sum_{k=1}^{M} \binom{M}{k} y_1[k]}{C_n} + \cdots \right] \tag{6.88}$$

$$= \nu \sum_{\ell \in \mathcal{S}} \sum_{n=n_{\ell-1}+1}^{n_\ell} \frac{\sum_{k=1}^{M} \binom{M}{k} y_\ell[k]}{C_n} \tag{6.89}$$

Furthermore, the total desired symbols from all databases from all repetitions is given by,

$$L(\mathbf{n}) = \nu \sum_{\ell \in \mathcal{S}} \sum_{k=1}^{M} \binom{M-1}{k-1} y_\ell[k](n_\ell - n_{\ell-1}) \tag{6.90}$$

Consequently, the following rate is achievable corresponding to the sequence $\mathbf{n}$,

$$R(\mathbf{n}, \mathbf{C}) = \frac{\sum_{\ell \in \mathcal{S}} \sum_{k=1}^{M} \binom{M-1}{k-1} y_\ell[k](n_\ell - n_{\ell-1})}{\sum_{\ell \in \mathcal{S}} \sum_{n=n_{\ell-1}+1}^{n_\ell} \frac{\sum_{k=1}^{M} \binom{M}{k} y_\ell[k]}{C_n}} \tag{6.91}$$

Since this scheme is achievable for every monotone non-decreasing sequence

$\mathbf{n} = \{n_i\}_{i=0}^{M-1}$, the following rate is achievable,

$$R(\mathbf{C}) = \max_{n_0 \leq \cdots \leq n_{M-1} \in \{1,\cdots,N\}} \frac{\sum_{\ell \in \mathcal{S}} \sum_{k=1}^{M} \binom{M-1}{k-1} y_\ell[k](n_\ell - n_{\ell-1})}{\sum_{\ell \in \mathcal{S}} \sum_{n=n_{\ell-1}+1}^{n_\ell} \frac{\sum_{k=1}^{M} \binom{M}{k} y_\ell[k]}{C_n}} \tag{6.92}$$

6.6 PIR from Multiple Access Channel

In this section, we consider the MAC-PIR problem. This problem is an extension of the NPIR model presented in Section 6.2 which consists of N non-colluding and replicated databases storing M messages. In MAC-PIR (see Fig. 6.5), the user sends a query $Q_n^{[i]}$ for the nth database to retrieve W_i privately and correctly. The nth database responds with an answer string $A_n^{[i]} = (X_{n,1}^{[i]}, \cdots, X_{n,t}^{[i]})$. The user receives a noisy observation $\tilde{A}_n^{[i]} = (Y_1^{[i]}, \cdots, Y_t^{[i]})$, where the responses of the databases $(A_1^{[i]}, A_2^{[i]}, \cdots, A_N^{[i]})$ pass through a discrete memoryless channel with a transition probability distribution $p(y|x_1, \cdots, x_N)$, i.e.,

$$P\left(\tilde{A}^{[i]} | A_1^{[i]}, A_2^{[i]}, \cdots, A_N^{[i]}\right) = \prod_{\eta=1}^{t} p\left(y_\eta^{[i]} | x_{1,\eta}^{[i]}, x_{2,\eta}^{[i]}, \cdots, x_{N,\eta}^{[i]}\right) \tag{6.93}$$

In this sense, the retrieval is performed via a *cooperative multiple access channel*, as the databases cooperate to convey the message W_i to a common receiver (the user). The full cooperation is realized via the user queries. Furthermore, in MAC-PIR, the database responses are mixed together to have the noisy observation $\tilde{A}^{[i]}$ in contrast to the noisy PIR problem with orthogonal links presented in Section 6.2.

In MAC-PIR, the user should be able to reconstruct W_i with vanishingly small

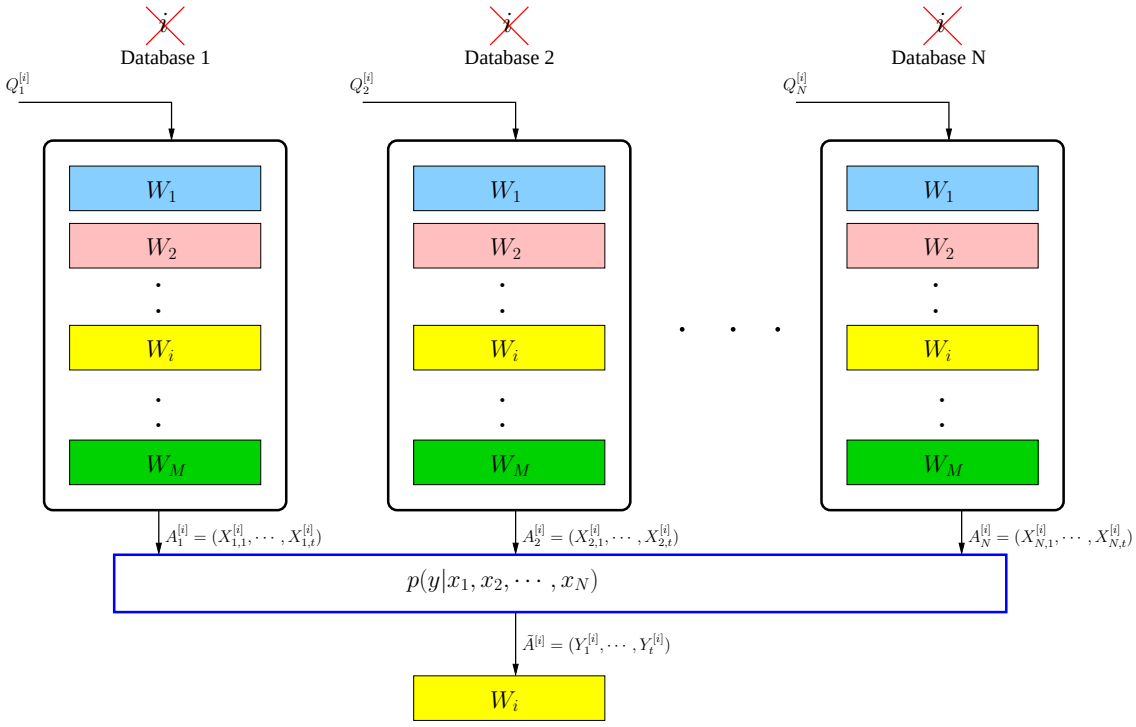


Figure 6.5: The MAC-PIR problem.

probability of error by observing the noisy and mixed output $\tilde{A}^{[i]}$, i.e., the reliability constraint is written as:

$$H(W_i|Q_{1:N}^{[i]}, \tilde{A}^{[i]}) \leq o(L) \tag{6.94}$$

and the privacy constraint is written as:

$$(Q_n^{[i]}, A_n^{[i]}, W_{1:M}) \sim (Q_n^{[j]}, A_n^{[j]}, W_{1:M}), \quad \forall i, j \in \{1, \cdots, M\} \tag{6.95}$$

We observe that the main difference between (6.95) and (6.8) is that we cannot claim that $\tilde{A}^{[i]} \sim \tilde{A}^{[j]}$ in the MAC-PIR problem. This is due to the fact that the user cannot statistically differentiate between the responses corresponding to each

message and hence the user cannot decode the desired message. This is in contrast to the NPIR problem with orthogonal links, where $\tilde{A}^{[i]} \sim \tilde{A}^{[j]}$ due to the Markov chain $(W_{1:M}, Q_n^{[i]}) \to A_n^{[i]} \to \tilde{A}_n^{[i]}$.

The retrieval rate for the MAC-PIR is given by:

$$R = \frac{L}{t} \tag{6.96}$$

and the MAC-PIR capacity is $C_{\text{PIR}} = \sup\ R$ over all retrieval schemes. We note that, without loss of generality, we can assume that all responses from the databases have the same length t in contrast to the NPIR problem with orthogonal links. The reason is that the retrieval rate depends only on the output of the channel and not on the individual responses of the databases. Hence, even if the database responses are different in lengths, we can choose $t = \max_{n\in[N]}\ t_n$ by appending the remaining responses by dummy symbols.

In the sequel, we discuss the issue of separability of channel coding and the information retrieval in MAC-PIR via some examples. Interestingly, we show that the optimal PIR scheme for the additive MAC and logic conjunction/disjunction MAC, the channel coding and the retrieval scheme are dependent on the channel transition probability, and hence channel coding and retrieval procedure are inseparable.

6.6.1 Additive MAC

In the first special case, we consider the additive MAC. In the additive MAC, at each time instant η, the responses of the databases are added together (in modulo-

2) in addition to a random variable $Z_\eta \sim$ Bernoulli(p), which is independent of $(W_{1:M}, Q_{1:N}^{[i]})$ and corresponds to a random additive noise, i.e.,

$$Y_\eta = \sum_{n=1}^{N} X_{n,\eta} + Z_\eta \tag{6.97}$$

The following theorem characterizes the capacity of the MAC-PIR problem if the channel is restricted to additive MACs.

Theorem 6.3 *The additive MAC-PIR capacity is,*

$$C_{PIR} = 1 - H(p) \tag{6.98}$$

where $p \in [0, 0.5)$ is the flipping probability of the additive noise.

We have the following remarks.

Remark 6.6 *For noiseless additive MAC, i.e., $p = 0$ and $Y_\eta = \sum_{n=1}^{N} X_{n,\eta}$, the MAC-PIR capacity is $C_{PIR} = 1$. This implies that there is no penalty due to the privacy constraint, i.e., the user can have privacy for free. Interestingly, this is the first instance where the PIR capacity is independent of the number of databases N and the number of messages M.*

Remark 6.7 *For noiseless additive MAC, i.e., $p = 0$, separation between channel coding and retrieval process is not optimal unlike the NPIR problem with orthogonal links. In fact, the retrieval scheme is dependent on the structure of the channel. To see this, the user generates a random binary vector $\mathbf{h} = [h_1\, h_2\, \cdots\, h_M] \in \{0,1\}^M$.*

The user sends $\mathbf{h}$ *to database 1, flips the* i*th position of* $\mathbf{h}$ *and sends it to database 2, and does not send anything to the remaining databases. Thus, the responses of the databases are,*

$$A_1^{[i]} = \sum_{m=1}^{M} h_m W_m \tag{6.99}$$

$$A_2^{[i]} = \sum_{m=1}^{M} h_m W_m + W_i \tag{6.100}$$

This is exactly the retrieval scheme in [1]. Since the channel is additive and noiseless, $\tilde{A}^{[i]} = A_1^{[i]} + A_2^{[i]} = W_i$*. Hence, the user downloads 1 bit from the channel in order to get 1 bit from the desired file and* $R = 1$*. Here, we note that, the channel performs the processing at the user for free. This implies that by careful design of queries, the user can exploit the channel in its favor to maximize the retrieval rate.*

Proof: We prove the converse and achievability.

The converse proof: To show the converse, we assume that W_1 is the desired message without loss of generality. Then, we have the following implications,

$$\begin{aligned}
L &= H(W_1) && (6.101)\\
&\overset{(6.2),(6.3)}{=} H(W_1|W_{2:M}, Q_{1:N}^{[1]}) && (6.102)\\
&\overset{(6.94)}{\leq} H(W_1|W_{2:M}, Q_{1:N}^{[1]}) - H(W_1|W_{2:M}, Q_{1:N}^{[1]}, \tilde{A}^{[1]}) + o(L) && (6.103)\\
&= I(W_1; \tilde{A}^{[1]}|Q_{1:N}^{[1]}, W_{2:M}) + o(L) && (6.104)\\
&= H(\tilde{A}^{[1]}|Q_{1:N}^{[1]}, W_{2:M}) - H(\tilde{A}^{[1]}|Q_{1:N}^{[1]}, W_{1:M}) + o(L) && (6.105)
\end{aligned}$$

$$\overset{(6.4)}{\leq} H(\tilde{A}^{[1]}) - H(\tilde{A}^{[1]}|Q_{1:N}^{[1]}, W_{1:M}, A_{1:N}^{[1]}) + o(L) \tag{6.106}$$

$$= t - H(\tilde{A}^{[1]}|A_{1:N}^{[1]}) + o(L) \tag{6.107}$$

$$= t - \sum_{\eta=1}^{t} H(Y_{\eta}^{[1]}|X_{1,\eta}^{[1]}, X_{2,\eta}^{[1]}, \cdots, X_{N,\eta}^{[1]}) + o(L) \tag{6.108}$$

$$= t - \sum_{\eta=1}^{t} H\left(\sum_{n=1}^{N} X_{n,\eta}^{[1]} + Z_{\eta}|X_{1,\eta}^{[1]}, X_{2,\eta}^{[1]}, \cdots, X_{N,\eta}^{[1]}\right) + o(L) \tag{6.109}$$

$$= t - \sum_{\eta=1}^{t} H(Z_{\eta}|X_{1,\eta}^{[1]}, X_{2,\eta}^{[1]}, \cdots, X_{N,\eta}^{[1]}) + o(L) \tag{6.110}$$

$$= t(1 - H(p)) + o(L) \tag{6.111}$$

where (6.102) follows from the independence of the messages and the queries, (6.103) follows from the reliability constraint, (6.106) follows from the fact that the answer string $A_n^{[1]}$ is a deterministic function of the messages and the queries, (6.107) follows from the fact that $(W_{1:M}, Q_{1:N}^{[1]}) \to A_{1:N}^{[1]} \to \tilde{A}^{[1]}$ is a Markov chain, (6.108) follows from the fact that the channel is memoryless, and (6.111) follows from the independence of Z_η and $(X_{1,\eta}^{[1]}, X_{2,\eta}^{[1]}, \cdots, X_{N,\eta}^{[1]})$ as a consequence of the independence of $(Z_\eta, W_{1:M}, Q_{1:N}^{[1]})$.

Hence, by reordering terms and taking $L \to \infty$, we have $R = \frac{L}{t} \leq 1 - H(p)$. Note that we can interpret the upper bound as the cooperative MAC bound, i.e., $R \leq I(Y; X_1, X_2, \cdots, X_N) = 1 - H(p)$.

The achievability proof: To show the general achievability, the user submits queries to database 1 and database 2 only and ignores the remaining databases. We note that the additive MAC in this case boils down to $Y_\eta = X_{1,\eta} + X_{2,\eta} + Z_\eta$, which means

that the channel $p(y|x_1, x_2)$ is BSC(p). Consequently, we use again Shannon's coding theorem for BSC in Lemma 6.4.

To that end, let the mth message be a vector $W_m = [W_m(1) \quad W_m(2) \quad \cdots \quad W_m(L)]$ of length L. The user repeats the following scheme L times. For the jth repetition of the scheme, the user generates a random binary vector $\mathbf{h}(j) = [h_1(j) \quad h_2(j) \quad \cdots \quad h_M(j)] \in \{0,1\}^M$. The user sends the following queries to the databases:

$$Q_1^{[i]}(j) = \mathbf{h}(j) \tag{6.112}$$

$$Q_2^{[i]}(j) = \mathbf{h}(j) + \mathbf{e}_i \tag{6.113}$$

where $\mathbf{e}_i$ is the unit vector containing 1 only at the ith position. The queries are private since $Q_n^{[i]}$ is a vector picked uniformly from $\{0,1\}^M$ for any message i.

For the jth repetition of the scheme, the database uses the received query vector as a combining vector for the jth element of all messages. The nth database concatenates all responses in a vector $U_n^{[i]}$ of length L, hence

$$U_1^{[i]} = \left[\sum_{m=1}^{M} h_m(1)W_m(1) \quad \sum_{m=1}^{M} h_m(2)W_m(2) \quad \cdots \quad \sum_{m=1}^{M} h_m(L)W_m(L)\right] \tag{6.114}$$

$$U_2^{[i]} = \left[\sum_{m=1}^{M} h_m(1)W_m(1) + W_i(1) \quad \sum_{m=1}^{M} h_m(2)W_m(2) + W_i(2) \right. \\ \left. \cdots \quad \sum_{m=1}^{M} h_m(L)W_m(L) + W_i(L)\right] \tag{6.115}$$

From Lemma 6.4, for $p \in (0, 0.5)$, all but ρ linear $[t, L]$ block codes $\mathcal{C}$, where

$\frac{L}{t} = R < 1 - H(p)$ that have $P_e(\mathcal{C}) < \frac{2}{\rho} \cdot 2^{-t\Delta(p,R)}$. Then, the databases agree on the same $[t, L]$ code from the family of good codes, where $t = \frac{L}{\lfloor 1-H(p) \rfloor}$. The nth database encodes $U_n^{[i]}$ independently by the same $[t, L]$ linear block code to output $A_n^{[i]}$.

After passing through the noisy channel, the noisy observation is given by:

$$\tilde{A}^{[i]} = A_1^{[i]} + A_2^{[i]} + Z_{1:t} \tag{6.116}$$

$$= \hat{A}^{[i]} + Z_{1:t} \tag{6.117}$$

Since the two databases employ the same linear block code, the sum of the two codewords $\hat{A}^{[i]} = A_1^{[i]} + A_2^{[i]}$ is also a valid codeword corresponding to the sum $U_1^{[i]} + U_2^{[i]}$.

Consequently, as $L \to \infty$, $t \to \infty$, the probability of error in decoding the sum $U_1^{[i]} + U_2^{[i]}$ is $P_e(L) \to 0$. By observing that $U_1^{[i]} + U_2^{[i]} = W_i$, the reliability proof follows. ■

Remark 6.8 *In the achievability proof, the PIR scheme relies on the additivity of the channel. In particular, the scheme uses a linear block code to exploit the fact that the sum of two codewords from a linear block code is also a valid codeword. Consequently, the retrieval process depends on the channel transition probability explicitly as opposed to the NPIR problem with orthogonal links.*

6.6.2 Logic Conjunction/Disjunction MACs

In this section, we show that we can achieve privacy for free for MACs other than the additive MACs. We illustrate this result by considering the MAC-PIR problem through channels that output the logical conjunctions (logic AND)/disjunctions (logic OR) of the inputs. Let $\wedge$ denote the logical conjunction operator, $\vee$ denote the logical disjunction operator, and $\neg$ denote the logical negation operator. The input-output relation of the discrete memoryless logical conjunction channel is given as:

$$Y_\eta = \bigwedge_{n=1}^{N} X_{n,\eta} \tag{6.118}$$

For the logical conjunction channel, we have the following capacity result.

Theorem 6.4 *In the logical conjunction MAC-PIR problem, if $N \geq 2^{M-1}$, then the MAC-PIR capacity is $C_{PIR} = 1$, where M is the number of messages.*

We have the following observations:

Remark 6.9 *Similar to the additive MAC, there is no loss due to the privacy constraint for the conjunction MAC. In this case, the capacity depends on the number of messages M, and the number of databases N unlike the additive MAC. Interestingly, the result shows the first instance of a threshold for the number of databases at which the full unconstrained capacity can be achieved $N = 2^{M-1}$, which is dependent on the number of messages M.*

Remark 6.10 *We note that the minimum number of databases N that results in $C_{PIR} = 1$ is still an open problem. In fact, the capacity for $N < 2^{M-1}$ is also an interesting open problem.*

Proof: It suffices to show only the achievability for this problem as the retrieval rate is trivially upper bounded by 1. To that end, the user submits queries to 2^{M-1} databases and submits nothing to the remaining databases. The user generates the random variables $(Z_1, \cdots, Z_M)$ independently, privately, and uniformly from $\{0, 1\}$. The random variable $Z_m \sim \text{Bernoulli}(\frac{1}{2})$ is a Bernoulli random variable that represents the negation state of the mth message literal in the first query $Q_1^{[i]}$, i.e., if $Z_m = 1$, this means that the user requests W_m in $Q_1^{[i]}$, while $Z_m = 0$ means that the user requests $\neg W_m$ in $Q_1^{[i]}$. Let $\tilde{W}_m$ be the requested literal from the mth message in $Q_1^{[i]}$, hence,

$$\tilde{W}_m = \begin{cases} W_m, & Z_m = 1 \\ \neg W_m, & Z_m = 0 \end{cases} \tag{6.119}$$

Now, without loss of generality, assume that W_1 is the desired message. From database 1, the user requests to download the disjunction $X_1 = \bigvee_{m=1}^{M} \tilde{W}_m$. From every other database, the user requests the same literal $\tilde{W}_1$ with a new disjunction of the remaining messages with different negation pattern than what is requested from database 1. I.e., from database 2, the user requests the disjunction $X_2 = \tilde{W}_1 \vee \neg\tilde{W}_2 \vee \bigvee_{m\in[M]\setminus\{1,2\}} \tilde{W}_m$. From database 3, the user requests the disjunction $X_3 = \tilde{W}_1 \vee \neg\tilde{W}_3 \vee \bigvee_{m\in[M]\setminus\{1,3\}} \tilde{W}_m$, $\cdots$ etc. Denote the disjunction of messages

$W_{2:M}$ requested from the nth database by F_n, where $n \in \{1, \cdots, 2^{M-1}\}$, then the received observation at the user is

$$Y = \left(\bigvee_{m=1}^{M} \tilde{W}_m \right) \wedge \left(\tilde{W}_1 \vee \neg \tilde{W}_2 \vee \bigvee_{m \in [M] \setminus \{1,2\}} \tilde{W}_m \right) \wedge \left(\tilde{W}_1 \vee \neg \tilde{W}_3 \vee \bigvee_{m \in [M] \setminus \{1,3\}} \tilde{W}_m \right) \wedge \cdots \tag{6.120}$$

$$= \tilde{W}_1 \vee \bigwedge_{i=1}^{2^{M-1}} F_i \tag{6.121}$$

$$= \tilde{W}_1 \tag{6.122}$$

where (6.121) follows from successively applying the Boolean relation $(\tilde{W}_1 \vee G_1) \wedge (\tilde{W}_1 \vee G_2) = \tilde{W}_1 \vee (G_1 \wedge G_2)$ for any logical expressions G_1, G_2. (6.122) follows from the fact that there exist 2^{M-1} different negation states for the literals from $W_{2:M}$, each negation state is requested from one database in the form of logical expression F_i, hence the conjunction of all these logical expressions $\bigwedge_{i=1}^{2^{M-1}} F_i = 0$ as all possible product of sums of $W_{2:M}$ exist in the conjunction. This satisfies the reliability constraint. Another way to see this result is that the queries are designed such that they cover *exactly half* the M-dimensional Karnaugh map, which can be reduced to either W_1 or $\neg W_1$.

Furthermore, since the negation state for every message is chosen uniformly, independently, and uniformly for each message, the probability of receiving specific query from the user is $\frac{1}{2^M}$ irrespective to the desired message, which guarantees the privacy. ■

Illustrative example: $M = 3$ messages, $N = 4$ databases with conjunction channel: As an explicit example, let $M = 3$, $N = 2^{M-1} = 4$, then the user requests the following:

$$X_1 = \tilde{W}_1 \vee \tilde{W}_2 \vee \tilde{W}_3 \tag{6.123}$$

$$X_2 = \tilde{W}_1 \vee \neg\tilde{W}_2 \vee \tilde{W}_3 \tag{6.124}$$

$$X_3 = \tilde{W}_1 \vee \tilde{W}_2 \vee \neg\tilde{W}_3 \tag{6.125}$$

$$X_4 = \tilde{W}_1 \vee \neg\tilde{W}_2 \vee \neg\tilde{W}_3 \tag{6.126}$$

Hence, the output of the channel is,

$$Y = X_1 \wedge X_2 \wedge X_3 \wedge X_4 \tag{6.127}$$

$$= (\tilde{W}_1 \vee \tilde{W}_2 \vee \tilde{W}_3) \wedge (\tilde{W}_1 \vee \neg\tilde{W}_2 \vee \tilde{W}_3) \wedge (\tilde{W}_1 \vee \tilde{W}_2 \vee \neg\tilde{W}_3) \wedge (\tilde{W}_1 \vee \neg\tilde{W}_2 \vee \neg\tilde{W}_3) \tag{6.128}$$

$$= (\tilde{W}_1 \vee (\tilde{W}_2 \vee \tilde{W}_3) \wedge (\neg\tilde{W}_2 \vee \tilde{W}_3)) \wedge (\tilde{W}_1 \vee (\tilde{W}_2 \vee \neg\tilde{W}_3) \wedge (\neg\tilde{W}_2 \vee \neg\tilde{W}_3)) \tag{6.129}$$

$$= (\tilde{W}_1 \vee W_3) \wedge (\tilde{W}_1 \vee \neg\tilde{W}_3) \tag{6.130}$$

$$= \tilde{W}_1 \tag{6.131}$$

Thus, the user can decode W_1 from Y as the user knows the correct negation pattern for $\tilde{W}_1$ privately. The scheme is private as all queries are equally likely with probability $\frac{1}{8}$ irrespective to the desired message. Since the user downloads 1 bit to

retrieve 1 bit from the desired message, the retrieval rate $R = 1$.

Remark 6.11 *We note that the result is still valid if the channel is replaced by a disjunction channel, i.e.,*

$$Y_\eta = \bigvee_{n=1}^{N} X_{n,\eta} \tag{6.132}$$

In this case, the user submits the same queries for the databases with replacing every disjunction operator with a conjunction operator. The proof of reliability follows from the duality of the product-of-sum and the sum-of-product.

Remark 6.12 *The achievable scheme for the conjunction channel is a non-linear retrieval scheme that depends on the non-linear characteristics of the channel in contrast to the linear retrieval scheme used for the additive channel. This confirms the non-separability between the retrieval scheme and the channel coding needed for reliable communication through the channel.*

6.6.3 Selection Channel

In this example, we illustrate the fact that the *privacy for free* phenomenon may not be always feasible for any arbitrary channel in the MAC-PIR problem. To illustrate this, we consider the selection channel. In this channel, the user selects to connect to one database only at random and sticks to it throughout the transmission, i.e.,

$$Y_\eta = X_{n,\eta}, \quad n \sim \text{uniform } \{1, \cdots, N\} \tag{6.133}$$

In this channel, the user is connected to the same database at every channel use. This implies that the user faces a single-database ($N = 1$) PIR problem at every channel use. The optimal PIR strategy for $N = 1$ is to download all the messages (M messages) from the connected database. Thus, the PIR capacity is given by $C_{\text{PIR}} = \frac{1}{M}$.

It is worth noting that there is another slight variant of the selection channel, in which the user selects to connect to one database at random at every channel use, i.e.,

$$Y_\eta = X_{n(\eta),\eta}, \quad n(\eta) \sim \text{uniform}\,\{1, \cdots, N\} \tag{6.134}$$

where $n(\eta)$ corresponds to the database index at channel use η. Then, $C_{PIR} \leq C = (1 + \frac{1}{N} + \cdots + \frac{1}{N^{M-1}})^{-1}$ trivially as the capacity of the classical PIR C [12], in which all the databases are connected to the user, is an upper bound for this problem, as the user can choose to ignore all the responses except the ones in the classical PIR problem. For the achievability, the user can repeat the achievable scheme in [12] ν times, which results in using the selection channel $t = \nu\frac{L}{C} = \nu\frac{N(N^M-1)}{N-1}$. At channel use η, the user chooses a new query element from $Q^{[i]}_{n(\eta)}$ and submits it to database $n(\eta)$. As $\nu \to \infty$, by strong law of large numbers, each database will be visited t_n times, where $t_n \to \frac{t}{N}$ in the limit for every n. Hence, all bits are decodable by the decodability of the scheme in [12] and $C_{PIR} = C = (1 + \frac{1}{N} + \cdots + \frac{1}{N^{M-1}})^{-1} < 1$ as well.

6.7 Conclusions

In this chapter, we introduced noisy PIR with orthogonal links (NPIR), and PIR from multiple access channels (MAC-PIR). We focused on the issue of the separability of the channel coding and the retrieval scheme. For the NPIR problem, we proved that the channel coding and the retrieval scheme are *almost separable* in the sense that every database implements its own channel coding independently from other databases. The problem is coupled only through agreeing on a suitable traffic ratio vector to maximize the retrieval rate. On the other hand, these conclusions are not valid for the MAC-PIR problem. We showed two examples, namely: PIR from additive MAC and PIR from logical conjunction/disjunction MAC. In these examples, we showed that the channel coding and retrieval schemes are indeed *inseparable* unlike in the NPIR problem. In both cases, we showed that by careful design of joint retrieval and coding schemes, we can attain the full capacity $C_{PIR} = 1 - H(p)$ and $C_{PIR} = 1$, respectively, with no loss due to the privacy constraint.

CHAPTER 7

Private Information Retrieval Through Wiretap Channel II

7.1 Introduction

In this chapter, we consider the problem of private information retrieval through wiretap channel II (PIR-WTC-II). In PIR-WTC-II, a user wants to retrieve a single message (file) privately out of M messages, which are stored in N replicated and non-communicating databases. An external eavesdropper observes a fraction μ_n (of its choice) of the traffic exchanged between the nth database and the user. In addition to the privacy constraint, the databases should encode the returned answer strings such that the eavesdropper learns absolutely nothing about the *contents* of the databases. We aim at characterizing the capacity of the PIR-WTC-II under the combined privacy and security constraints. We obtain a general upper bound for the problem in the form of a max-min optimization problem, which extends the converse proof of the PIR problem under asymmetric traffic constraints. We propose an achievability scheme that satisfies the security constraint by encoding a secret key, which is generated securely at each database, into an artificial noise vector using an MDS code. The user and the databases operate at one of the corner points of

the achievable scheme for the PIR under asymmetric traffic constraints such that the retrieval rate is maximized under the imposed security constraint. The upper bound and the lower bound match for the case of $M = 2$ and $M = 3$ messages, for any N, and any $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)$.

7.2 System Model

Consider a classical PIR model, in which there are N non-colluding and replicated databases, each storing the same content of M messages (or files). The message W_m is represented as a vector of length L, whose elements are picked from a finite field $\mathbb{F}_q^L$ with a sufficiently large alphabet. The messages $W_{1:M} = \{W_1, \cdots, W_M\}$ are independent and identically distributed, hence,

$$H(W_m) = L, \quad m \in \{1, \cdots, M\} \tag{7.1}$$

$$H(W_{1:M}) = ML, \quad (q\text{-ary bits}) \tag{7.2}$$

We assume that the messages are uncoded and fixed, i.e., we assume that the contents of the databases cannot be coded to satisfy the security constraint during the storage phase.

In classical PIR, a user wants to retrieve a message W_i from the N databases without revealing the identity of the message i to any individual database. The user prepares N queries, one for each database. The user sends $Q_n^{[i]}$ to the nth database. Since the user has no knowledge about the realization of $W_{1:M}$, the queries and the

messages are statistically independent, i.e.,

$$I(Q_{1:N}^{[i]}; W_{1:M}) = 0, \quad i \in \{1, \cdots, M\} \tag{7.3}$$

where $Q_{1:N}^{[i]} = \{Q_1^{[i]}, \cdots, Q_N^{[i]}\}$. Furthermore, to ensure the privacy of W_i, the user should constrain the query intended to retrieve W_i to be indistinguishable from the query intended to retrieve any other message W_j at any individual database. Thus, the privacy constraint is formalized as,

$$(Q_n^{[i]}, A_n^{[i]}, W_{1:M}) \sim (Q_n^{[j]}, A_n^{[j]}, W_{1:M}), \quad \forall j \in \{1, \cdots, M\} \tag{7.4}$$

where $\sim$ denotes statistical equivalence.

The nth database, after receiving the query $Q_n^{[i]}$, responds with a t_n-length answering string $A_n^{[i]}$. Note that we allow the user and the databases to choose arbitrary lengths for the answer strings such that they maximize the retrieval rate. The answer string is generally a *stochastic* mapping of the messages $W_{1:M}$ and the received query $Q_n^{[i]}$, hence,

$$H(A_n^{[i]} | Q_n^{[i]}, W_{1:M}, \mathcal{G}_n) = 0, \quad n \in \{1, \cdots, N\} \tag{7.5}$$

where $\mathcal{G}_n$ is a random variable independent of all other random variables, whose realization is known at the nth database only and not shared with any other database or the user a priori of the transmission. We denote the traffic ratio vector by

$\boldsymbol{\tau} = (\tau_1, \cdots, \tau_N)$. The traffic ratio at the nth database τ_n is given by,

$$\tau_n = \frac{t_n}{\sum_{i=1}^{N} t_i} \tag{7.6}$$

We assume that the answer strings are transmitted through a WTC-II (see Fig. 7.1). In this case, an external eavesdropper (wiretapper) wishes to learn about the contents of the databases by observing the queries and answer strings exchanged by the user and the databases. In PIR-WTC-II, the user observes the t_n-length answer string $A_n^{[i]}$ from the nth database through a noiseless channel. On the other hand, the eavesdropper can observe a fraction μ_n from the nth answer string. More specifically, the eavesdropper arbitrarily chooses any set of positions $\mathcal{S}_n \subset \{1, \cdots, t_n\}$ to observe from the nth answer string, such that $|\mathcal{S}_n| = \mu_n t_n$, i.e., the output of the eavesdropper channel is given by,

$$Z_n^{[i]} = A_n^{[i]}(\mathcal{S}_n), \quad n \in \{1, \cdots, N\} \tag{7.7}$$

We denote the unobserved portion of the answer string by $Y_n^{[i]} = A_n^{[i]}(\bar{\mathcal{S}}_n)$, where $\bar{\mathcal{S}}_n = \{1, \cdots, N\} \backslash \mathcal{S}_n$, thus, $A_n^{[i]} = (Y_n^{[i]}, Z_n^{[i]})$. We write the eavesdropping ratios as a vector $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)$. Without loss of generality, we assume that the databases are arranged ascendingly in μ_n, i.e., $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_N$, i.e., the first database is the least threatened (most secure) and the Nth database is the most threatened (least secure).

Upon preparing the answer string, the databases should encode the answer

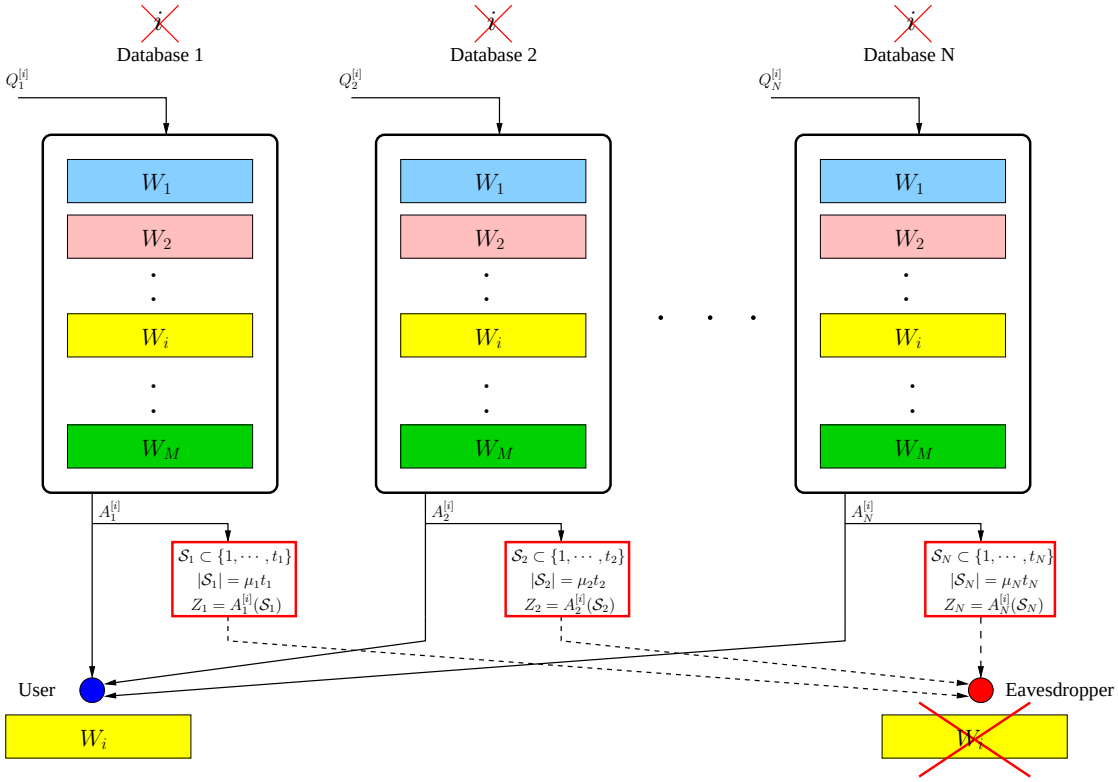


Figure 7.1: Secure PIR problem through wiretap channel II.

strings such that the eavesdropper learns nothing from observing any μ_n fraction from the traffic from the nth database even with observing the queries submitted by the user. Consequently, we write the security constraint as,

$$I(W_{1:M}; Z_{1:N}^{[i]}, Q_{1:N}^{[i]}) = 0 \tag{7.8}$$

Additionally, the user should be able to reconstruct the desired message W_i from the collected answer strings with arbitrarily small probability of error. Using Fano's inequality, we write the reliability constraint as,

$$H(W_i | Q_{1:N}^{[i]}, A_{1:N}^{[i]}) = o(L) \tag{7.9}$$

where $\frac{o(L)}{L} \to 0$ as $L \to \infty$.

For a fixed N, M, traffic ratio vector $\boldsymbol{\tau}$, and eavesdropping ratio vector $\boldsymbol{\mu}$, a retrieval rate $R(\boldsymbol{\tau}, \boldsymbol{\mu})$ is achievable if there exists a PIR scheme which satisfies the privacy constraint (7.4), security constraint (7.8), and the reliability constraint (7.9) for some message length $L(\boldsymbol{\tau}, \boldsymbol{\mu})$ and answer strings of lengths $\{t_n(\boldsymbol{\tau}, \boldsymbol{\mu})\}_{n=1}^{N}$ such that $\tau_n = \frac{t_n(\boldsymbol{\tau}, \boldsymbol{\mu})}{\sum_{i=1}^{N} t_i(\boldsymbol{\tau}, \boldsymbol{\mu})}$, where the retrieval rate is therefore given by,

$$R(\boldsymbol{\tau}, \boldsymbol{\mu}) = \frac{L(\boldsymbol{\tau}, \boldsymbol{\mu})}{\sum_{n=1}^{N} t_n(\boldsymbol{\tau}, \boldsymbol{\mu})} \tag{7.10}$$

We note that in this problem, the user and the databases can agree on a traffic ratio vector $\boldsymbol{\tau}$ to maximize the retrieval rate, thus, we can express the secure retrieval rate under eavesdropping capabilities $\boldsymbol{\mu}$, $R(\boldsymbol{\mu})$, as,

$$R(\boldsymbol{\mu}) = \max_{\boldsymbol{\tau}} \; R(\boldsymbol{\tau}, \boldsymbol{\mu}) \tag{7.11}$$

Note that the message lengths can grow arbitrarily large to conform with standard information-theoretic arguments. The capacity of the PIR-WTC-II problem $C(\boldsymbol{\mu})$ is defined as the supremum of all achievable retrieval rates over all achievable schemes, i.e., $C(\boldsymbol{\mu}) = \sup \; R(\boldsymbol{\mu})$.

7.3 Main Results and Discussions

In this section, we present the main results of this chapter. Our first result characterizes a general upper bound for the PIR-WTC-II problem for fixed M, N, and an

arbitrary $\boldsymbol{\mu}$.

Theorem 7.1 (Upper bound) *For the PIR-WTC-II problem under eavesdropping capabilities $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)$, the capacity is upper bounded by,*

$$C(\boldsymbol{\mu}) \leq \bar{C}(\boldsymbol{\mu}) = \max_{\boldsymbol{\tau} \in \mathbb{T}} \min_{n_i \in \{1,\cdot,N\}} \frac{\sum_{n=1}^{N}(1-\mu_n)\tau_n + \frac{\sum_{n=n_1+1}^{N}(1-\mu_n)\tau_n}{n_1} + \cdots + \frac{\sum_{n=n_{M-1}+1}^{N}(1-\mu_n)\tau_n}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{7.12}$$

where $\mathbb{T} = \left\{ \boldsymbol{\tau} : \tau_n \geq 0 \quad \forall n \in [1:N], \quad \sum_{n=1}^{N} \tau_n = 1 \right\}$.

The proof of this upper bound is given in Section 7.4. We have the following remarks.

Remark 7.1 *When $\boldsymbol{\mu} = (0, \cdots, 0)$, i.e., without any security constraints, the upper bound reduces to:*

$$\bar{C}(\boldsymbol{\mu}) = \max_{\boldsymbol{\tau} \in \mathbb{T}} \min_{n_i \in \{1,\cdots,N\}} \frac{\sum_{n=1}^{N} \tau_n + \frac{\sum_{n=n_1+1}^{N} \tau_n}{n_1} + \cdots + \frac{\sum_{n=n_{M-1}+1}^{N} \tau_n}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{7.13}$$

$$= \max_{\boldsymbol{\tau} \in \mathbb{T}} \min_{n_i \in \{1,\cdots,N\}} \frac{1 + \frac{\sum_{n=n_1+1}^{N} \tau_n}{n_1} + \cdots + \frac{\sum_{n=n_{M-1}+1}^{N} \tau_n}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{7.14}$$

$$= \max_{\boldsymbol{\tau}} \quad \tilde{C}(\boldsymbol{\tau}) \tag{7.15}$$

$$= \frac{1}{1 + \frac{1}{N} + \cdots + \frac{1}{N^{M-1}}} \tag{7.16}$$

where the inner problem in (7.14) *is precisely the upper bound of the PIR problem under asymmetric traffic $\boldsymbol{\tau}$ [125]. From [125], we know that $\tilde{C}(\boldsymbol{\tau})$ is maximized by adopting symmetric schemes, i.e., $\tau_n = \frac{1}{N}$, which achieves the PIR capacity C*

in [12].

Remark 7.2 *If the PIR-WTC-II problem is further constrained by the asymmetric traffic constraints $\boldsymbol{\tau}$, the corresponding upper bound $\bar{C}(\boldsymbol{\mu}, \boldsymbol{\tau})$ is given by the inner problem of* (7.12)*, i.e.,*

$$\bar{C}(\boldsymbol{\mu}, \boldsymbol{\tau}) = \min_{n_i \in \{1,\cdots,N\}} \frac{\sum_{n=1}^{N}(1-\mu_n)\tau_n + \frac{\sum_{n=n_1+1}^{N}(1-\mu_n)\tau_n}{n_1} + \cdots + \frac{\sum_{n=n_{M-1}+1}^{N}(1-\mu_n)\tau_n}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{7.17}$$

Hence, without the asymmetric traffic constraints, the user and the databases can agree on $\boldsymbol{\tau}$ that maximizes the retrieval rate, which results in the outer maximization over $\boldsymbol{\tau}$. This is reminiscent of the classical converse proof for the channel coding theorem, where a converse argument is constructed for an arbitrary input distribution of the transmission codebook, and then the converse proof is concluded with a maximization step over all the input distributions.

Remark 7.3 *The upper bound $\bar{C}(\boldsymbol{\mu})$ in Theorem 7.1 can be written as the following linear programming problem:*

$$\begin{aligned} \bar{C}(\boldsymbol{\mu}) = \max_{\boldsymbol{\tau}, R} \quad & R \\ s.t. \quad & R \leq \frac{\sum_{n=1}^{N}(1-\mu_n)\tau_n + \frac{\sum_{n=n_1+1}^{N}(1-\mu_n)\tau_n}{n_1} + \cdots + \frac{\sum_{n=n_{M-1}+1}^{N}(1-\mu_n)\tau_n}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}}, \forall \mathbf{n} \\ & \tau_n \geq 0, \quad n = 1, \cdots, N \\ & \sum_{n=1}^{N} \tau_n = 1 \end{aligned} \tag{7.18}$$

where $\mathbf{n} = (n_1, \cdots, n_{M-1}) \subset \{1, \cdots, N\}^{M-1}$*, i.e., the number of constraints are finite (at most* $N^{M-1}+2$ *constraints). Hence, the optimal solution of this optimization problem is attained at one of the corner points of the feasible set.*

Next, we present a general lower bound on $C(\boldsymbol{\mu})$ for fixed M, N.

Theorem 7.2 (Lower bound) *For PIR-WTC-II, for a monotone non-decreasing sequence* $\mathbf{n} = \{n_i\}_{i=0}^{M-1} \subset \{1, \cdots, N\}^M$*, let* $n_{-1} = 0$*, and* $\mathcal{S} = \{i \geq 0 : n_i - n_{i-1} > 0\}$*. Denote* $y_\ell[k]$ *to be the number of stages of the achievable scheme that downloads* k*-sums from the* n*th database in one repetition of the scheme, such that* $n_{\ell-1} \leq n \leq n_\ell$*, and* $\ell \in \mathcal{S}$*. Let* $\xi_\ell = \prod_{s \in \mathcal{S}\setminus\{\ell\}} \binom{M-2}{s-1}$*. The number of stages* $y_\ell[k]$ *is characterized by the following system of difference equations:*

$$
\begin{aligned}
y_0[k] &= (n_0 - 1) y_0[k-1] + \sum_{j \in \mathcal{S}\setminus\{0\}} (n_j - n_{j-1}) y_j[k-1] \\
y_1[k] &= (n_1 - n_0 - 1) y_1[k-1] + \sum_{j \in \mathcal{S}\setminus\{1\}} (n_j - n_{j-1}) y_j[k-1] \\
y_\ell[k] &= n_0 \xi_\ell \delta[k-\ell-1] + (n_\ell - n_{\ell-1} - 1) y_\ell[k-1] + \sum_{j \in \mathcal{S}\setminus\{\ell\}} (n_j - n_{j-1}) y_j[k-1], \quad \ell \geq 2
\end{aligned}
\tag{7.19}
$$

where $\delta[\cdot]$ *denotes the Kronecker delta function. The initial conditions of* (7.19) *are* $y_0[1] = \prod_{s \in \mathcal{S}} \binom{M-2}{s-1}$*, and* $y_j[k] = 0$ *for* $k \leq j$*. Consequently, the traffic ratio vector* $\boldsymbol{\tau}(\mathbf{n}) = (\tau_1(\mathbf{n}), \cdots, \tau_N(\mathbf{n}))$ *corresponding to the sequence* $\mathbf{n} = \{n_i\}_{i=0}^{M-1}$ *is given by:*

$$
\tau_n(\mathbf{n}) = \frac{\sum_{k=1}^{M} \binom{M}{k} y_j[k]}{\sum_{\ell \in \mathcal{S}} \sum_{k=1}^{M} \binom{M}{k} y_\ell[k] (n_\ell - n_{\ell-1})}, \quad n_{j-1} + 1 \leq n \leq n_j \tag{7.20}
$$

Then, the achievable rate corresponding to $\mathbf{n}$ *is given by:*

$$R(\mathbf{n}, \boldsymbol{\mu}) = \frac{\sum_{\ell \in \mathcal{S}} \sum_{k=1}^{M} \binom{M-1}{k-1} y_\ell[k](n_\ell - n_{\ell-1})}{\sum_{\ell \in \mathcal{S}} \sum_{n=n_{\ell-1}+1}^{n_\ell} \frac{\sum_{k=1}^{M} \binom{M}{k} y_\ell[k]}{1-\mu_n}} \tag{7.21}$$

Consequently, the capacity $C(\boldsymbol{\mu})$ *is lower bounded by:*

$$C(\boldsymbol{\mu}) \geq R(\boldsymbol{\mu}) = \max_{n_0 \leq \cdots \leq n_{M-1} \in \{1, \cdots, N\}} R(\mathbf{n}, \boldsymbol{\mu}) \tag{7.22}$$

$$= \max_{n_0 \leq \cdots \leq n_{M-1} \in \{1, \cdots, N\}} \frac{\sum_{\ell \in \mathcal{S}} \sum_{k=1}^{M} \binom{M-1}{k-1} y_\ell[k](n_\ell - n_{\ell-1})}{\sum_{\ell \in \mathcal{S}} \sum_{n=n_{\ell-1}+1}^{n_\ell} \frac{\sum_{k=1}^{M} \binom{M}{k} y_\ell[k]}{1-\mu_n}} \tag{7.23}$$

The proof of Theorem 7.2 can be found in Section 7.5. We have the following remarks.

Remark 7.4 *For fixed* M, N, *the number of the achievable rates* $R(\mathbf{n}, \boldsymbol{\mu})$ *in Theorem 7.2 corresponds to the number of monotone non-decreasing sequences* $\mathbf{n} = \{n_i\}_{i=0}^{M-1}$, *which is equal to* $\binom{M+N-1}{M}$.

Remark 7.5 *After achieving the corner points in Theorem 7.2, which achieve* $R(\mathbf{n}, \boldsymbol{\mu})$, *one can perform time-sharing between the corner points to obtain an achievable* $R(\boldsymbol{\tau}, \boldsymbol{\mu})$ *for any* $\boldsymbol{\tau}$. *The highest possible achievable rate can be obtained by maximizing over* $\boldsymbol{\tau}$. *However, this is not needed as time-sharing results in a piece-wise affine function in* $\boldsymbol{\tau}$. *Hence, maximizing over* $\boldsymbol{\tau}$ *would result in operating directly at one of the corner points.*

Remark 7.6 *We note that the core of the achievability scheme is the PIR scheme under asymmetric traffic constraints in Chapter 5. Hence, the recursive structure*

described by (7.19) *is directly inherited from [125]. Nevertheless, two main differences appear in the final rate expression. First, the answer string length from every database belonging to the same group is different in contrast to [125]. This is due to the fact that every database experiences a different eavesdropping capability* μ_n *in general, hence the* n*th database encrypts its responses with a key, whose length depends on* μ_n*, thus the key lengths are different in general. Second, there is no need for time-sharing over the corner points as shown in Remark 7.5.*

In the following corollary, we settle the capacity $C(\boldsymbol{\mu})$ for $M = 2$, $M = 3$, and arbitrary N.

Corollary 7.1 (Exact capacity for $M = 2$ and $M = 3$ messages) *For PIR-WTC-II, the capacity* $C(\boldsymbol{\mu})$ *for* $M = 2, 3$*, and an arbitrary* N *is given by:*

$$C(\boldsymbol{\mu}) = \begin{cases} \max_{n_0,n_1\in\{1,\cdots,N\}} \frac{n_0 n_1}{\sum_{n=1}^{n_0} \frac{n_0+1}{1-\mu_n} + \sum_{n=n_0+1}^{n_1} \frac{n_0}{1-\mu_n}}, & M = 2 \\ \max_{n_0,n_1,n_2\in\{1,\cdots,N\}} \frac{n_0 n_1 n_2}{\sum_{n=1}^{n_0} \frac{n_0 n_1+n_0+1}{1-\mu_n} + \sum_{n=n_0+1}^{n_1} \frac{n_0 n_1+n_0}{1-\mu_n} + \sum_{n=n_1+1}^{n_2} \frac{n_0 n_1}{1-\mu_n}}, & M = 3 \end{cases} \tag{7.24}$$

The proof of Corollary 7.1 can be found in Section 7.5.4.

Remark 7.7 *The explicit capacity expressions in Corollary 7.1 can be interpreted using basic circuit theory. To see that for* $M = 2$ *for a given* (n_0, n_1)*, consider the circuit in Fig. 7.2. The circuit has a current source of* $n_0 n_1$ *units. The circuit consists of* $n_0 + n_1$ *parallel resistors. The* n*th resistor has the value of* $R_n = \frac{1-\mu_n}{n_0+1}$ *if* $1 \leq n \leq n_0$*, and* $R_n = \frac{1-\mu_n}{n_0}$ *if* $n_0 + 1 \leq n \leq n_1$*. Hence, the capacity* $C(\boldsymbol{\mu})$ *is*

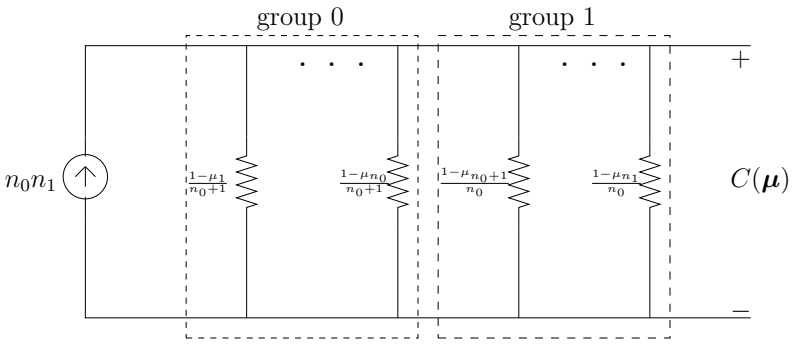


Figure 7.2: Circuit interpretation of $C(\boldsymbol{\mu})$ for $M = 2$.

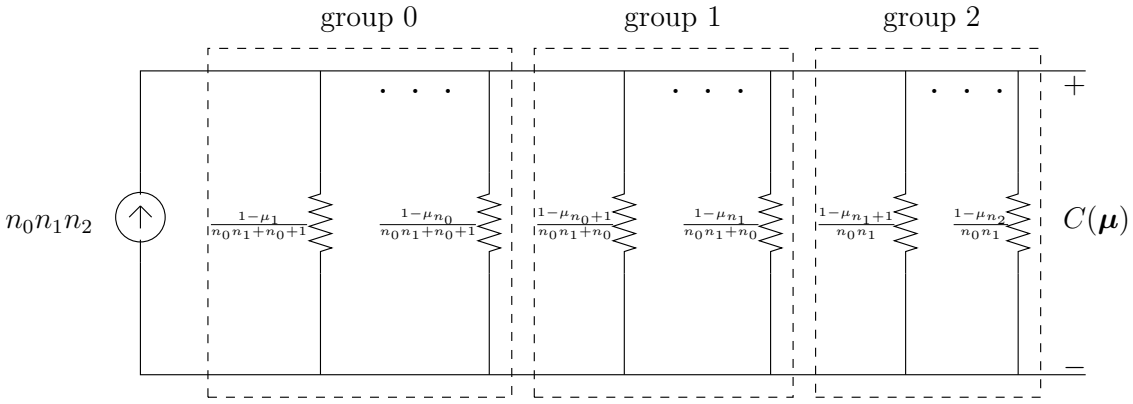


Figure 7.3: Circuit interpretation of $C(\boldsymbol{\mu})$ for $M = 3$.

the voltage across the current source. A similar interpretation can be inferred from Fig. 7.3 for the case of $M = 3$. Interestingly, this interpretation implies that in order to maximize the retrieval rate (the voltage across the equivalent resistance of the circuit), one should pick n_0, n_1, n_2 such that the resistance of each parallel branch is as symmetric as possible. This is due to the fact that the equivalent resistance of parallel resistors is less than the resistance of the least resistor.

Finally, in the next corollary, we present an explicit achievable rate for $R(\boldsymbol{\mu})$ when $N = 2$, and an arbitrary M. The proof of the corollary can be found in Section 7.5.5

Corollary 7.2 (Achievable retrieval rate for $N = 2$) *For PIR-WTC-II with*

$N = 2$ and an arbitrary M, let $s_2 = \{1, \cdots, M-1\}$, then the secure PIR capacity $C(\boldsymbol{\mu})$ is lower bounded by:

$$\max \left\{ \max_{s_2 \in \{0,\cdots,M-1\}} \frac{\binom{M-2}{s_2-1} + \sum_{k=0}^{M-s_2-1} \binom{M-1}{s_2+k}}{\frac{1}{1-\mu_1}\left[M\binom{M-2}{s_2-1} + \sum_{k=1}^{\lfloor \frac{M-s_2}{2} \rfloor} \binom{M}{s_2+2k}\right] + \frac{1}{1-\mu_2}\left[\sum_{k=0}^{\lfloor \frac{M-s_2-1}{2} \rfloor} \binom{M}{s_2+2k+1}\right]}, \frac{1-\mu_1}{M} \right\} \tag{7.25}$$

Remark 7.8 *We note the strong connection between the PIR-WTC-II problem and the PIR problem under asymmetric traffic constraints in Chapter 5. In PIR-WTC-II problem, the nth database uses a secret key of length $\mu_n t_n$ to span the entire space of the eavesdropper. This in turn leaves $(1-\mu_n)t_n$ symbols for meaningful queries. Since the eavesdropping vulnerabilities of the databases are different in general (different μ_n), the meaningful queries are naturally constrained, e.g., we expect the first database (the most secure) to support more meaningful queries than the remaining databases. However, the main difference between the two problems is that in the PIR problem under asymmetric traffic constraints [125], the traffic ratio vector $\boldsymbol{\tau}$ is fixed (by the problem formulation) in contrast to the PIR-WTC-II problem, where the user and the databases can agree on a traffic ratio vector $\boldsymbol{\tau}$ to maximize the retrieval rate under the fixed eavesdropping capabilities $\boldsymbol{\mu}$.*

Remark 7.9 *We now compare our model with the PIR model in [21,41]. In [21,41], there is an eavesdropper, which observes all communication of E out of N databases, whose identities are unknown to the user. We restrict the comparison to the case*

$T = 1$ *(i.e., no collusion between the databases). In this case, the capacity of the secure PIR problem in [41] (abbreviated as T-EPIR problem) is* $1 - \frac{E}{N}$*. This requires a common randomness, which is shared between the databases and unknown to the user, of length* $\frac{E}{N-E}$ *[41, Theorem 1]. We note that the capacity expression is independent of the number of messages in [41]. For the symmetric version of the problem in [21], the capacity expression is also* $1 - \frac{E}{N}$*. Interestingly, in the symmetric version of the problem, the common randomness among the databases is used to satisfy both the database privacy and the security constraints simultaneously.*

On the other hand, in our model, the eavesdropper wiretaps all N *databases according to the given* $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)$*. The user knows the ratio of the traffic which is observed by the eavesdropper from each database, i.e.,* $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)$*, but does not know which positions are being observed. Surprisingly, our model does not need any shared randomness among the databases or with the user, i.e., here we are able to achieve nontrivial PIR rates with zero shared randomness rates.*

As a concrete example, let $M = 3$*, and for a fair comparison, let* $\mu_n = \frac{E}{N}$ *for all* $n \in \{1, \cdots, N\}$ *in our model. The rationale for this choice of* μ_n *is that in [41], the eavesdropper has access to a total of* $E \cdot t$ *observations, where* t *is the length of the answer string from any database in [41]. Now, for symmetric* $\mu_n = \frac{E}{N}$ *in our model, all answer string lengths need to be symmetric, i.e.,* $t_n = t$ *for all* n*, and therefore, the eavesdropper accesses a total of* $\frac{E}{N} \cdot N \cdot t = E \cdot t$ *observations here as it does in [41]. The capacity for this case in our model, from Corollary 7.1, is* $\frac{1-\frac{E}{N}}{1+\frac{1}{N}+\frac{1}{N^2}}$*, which is attained with* $n_0 = n_1 = n_2 = N$ *in the corollary. This rate is strictly less than the rate in [41], which is* $1 - \frac{E}{N}$*, however, [41] requires a shared*

randomness between the databases at a rate of at least $\frac{E}{N-E}$*, while in our case no shared randomness is required.*

7.4 Converse Proof

In this section, we derive a general upper bound for the retrieval rate under the privacy and security constraints (7.4), (7.8) for the PIR-WTC-II problem. Our converse proof extends the techniques of [12] to incorporate the security constraint. In addition, since the eavesdropper observes a different fraction of the traffic from each database, we do not expect that the answer strings (and consequently the traffic ratios) from each database to be symmetric in length. Thus, we modify the converse proof in [12] to account for this prospected traffic asymmetry along the lines of Chapter 5. However, different from [125], traffic ratios are not given, and must be chosen; the eavesdropping ratios $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)$ are given here. Our converse proof extends the proof in [125] to account for the imposed security constraint.

In the next lemma, we discuss some consequences of the security constraint in (7.8). The security constraint introduces some interesting conditional independence properties which simplify the converse proof.

Lemma 7.1 (Security consequences) *In the PIR-WTC-II problem, the following implications are true due to the security constraint* (7.8)*:*

1. *Messages are conditionally independent given the observed part of the answer*

strings at the eavesdropper $Z_{1:N}^{[i]}$, i.e.,

$$I(W_m; W_{[1:M]\setminus\{m\}} | Z_{1:N}^{[i]}) = 0, \quad i,\, m \in \{1, \cdots, M\} \tag{7.26}$$

2. *There is no leakage of W_m from all the queries $Q_{1:N}^{[i]}$, the eavesdropper observations $Z_{1:N}^{[i]}$, and any subset of messages $W_{\mathcal{S}} = \{W_i : i \in \mathcal{S}\}$ such that $m \notin \mathcal{S}$,*

$$I(W_m; W_{\mathcal{S}}, Z_{1:N}^{[i]}, Q_{1:N}^{[i]}) = 0, \quad i,\, m \in \{1, \cdots, M\} \tag{7.27}$$

In particular,

$$I(W_m; W_{m:M} | W_{1:m-1}, Z_{1:N}^{[i]}) = L, \quad i,\, m \in \{1, \cdots, M\} \tag{7.28}$$

3. *The eavesdropper's observations $Z_{1:N}^{[i]}$ and the messages are conditionally independent given the queries $Q_{1:N}^{[i]}$, i.e., for sets $\mathcal{S}_1$, $\mathcal{S}_2$, such that $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$,*

$$I(W_{\mathcal{S}_1}; Z_{1:N}^{[i]} | Q_{1:N}^{[i]}, W_{\mathcal{S}_2}) = 0, \quad i \in \{1, \cdots, M\} \tag{7.29}$$

In particular,

$$I(W_{m:M}; Z_{1:N}^{[m-1]} | W_{1:m-1}) = 0, \quad m \in \{2, \cdots, M\} \tag{7.30}$$

4. *The messages and the queries are conditionally independent given the eaves-*

dropper's observations, i.e., for sets $\mathcal{S}_1$, $\mathcal{S}_2$, such that $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$,

$$I(W_{\mathcal{S}_1}; Q_{1:N}^{[i]} | W_{\mathcal{S}_2}, Z_{1:N}^{[i]}) = 0, \quad i \in \{1, \cdots, M\} \tag{7.31}$$

5. *The messages $W_{m:M}$ and the queries $Q_{k+1:N}^{[m]}$ for any $k \in \{1, \cdots, N\}$ are conditionally independent given $\left(W_{1:m-1}, Z_{1:N}^{[m]}, Q_{1:k}^{[m]}, Y_{1:k}^{[m]}\right)$, i.e.,*

$$I(W_{m:M}; Q_{k+1:N}^{[m]} | W_{1:m-1}, Z_{1:N}^{[m]}, Q_{1:k}^{[m]}, Y_{1:k}^{[m]}) = 0 \tag{7.32}$$

Proof:

1. From the security constraint (7.8), we have $I(W_{1:M}; Z_{1:N}^{[i]}, Q_{1:N}^{[i]}) = 0$, which further implies that $I(W_{1:M}; Z_{1:N}^{[i]}) = 0$. This can be expanded as:

$$\begin{aligned} 0 &= I(W_m, W_{[1:M]\setminus\{m\}}; Z_{1:N}^{[i]}) && (7.33)\\ &= I(W_m; Z_{1:N}^{[i]}) + I(W_{[1:M]\setminus\{m\}}; Z_{1:N}^{[i]} | W_m) && (7.34)\\ &= I(W_{[1:M]\setminus\{m\}}; Z_{1:N}^{[i]}) + I(W_m; Z_{1:N}^{[i]} | W_{[1:M]\setminus\{m\}}) && (7.35) \end{aligned}$$

which implies that all four terms in (7.34), (7.35) are zero. Then, consider

$$\begin{aligned} I(W_m; W_{[1:M]\setminus\{m\}}, Z_{1:n}^{[i]}) &= I(W_m; Z_{1:N}^{[i]}) + I(W_m; W_{[1:M]\setminus\{m\}} | Z_{1:N}^{[i]}) && (7.36)\\ &= I(W_m; W_{[1:M]\setminus\{m\}}) + I(W_m; Z_{1:N}^{[i]} | W_{[1:M]\setminus\{m\}}) \\ & && (7.37) \end{aligned}$$

which together with (7.34), (7.35) and the independence of the messages imply (7.26).

2. From the security constraint (7.8), we have $I(W_m, W_{\mathcal{S}}; Q_{1:N}^{[i]}, Z_{1:N}^{[i]}) = 0$ by the non-negativity of mutual information. This can be further expanded as

$$0 = I(W_m, W_{\mathcal{S}}; Q_{1:N}^{[i]}, Z_{1:N}^{[i]}) = I(W_{\mathcal{S}}; Q_{1:N}^{[i]}, Z_{1:N}^{[i]}) + I(W_m; Q_{1:N}^{[i]}, Z_{1:N}^{[i]} | W_{\mathcal{S}}) \tag{7.38}$$

From the second term on the right hand side, we have $I(W_m; Q_{1:N}^{[i]}, Z_{1:N}^{[i]} | W_{\mathcal{S}}) = 0$, which implies (7.27) by the independence of the messages, as $I(W_m; W_{\mathcal{S}}, Z_{1:N}^{[i]}, Q_{1:N}^{[i]}) = I(W_m; W_{\mathcal{S}}) + I(W_m; Z_{1:N}^{[i]}, Q_{1:N}^{[i]} | W_{\mathcal{S}})$.

For (7.28), we note that (7.27) implies that $I(W_m; W_{1:m-1}, Z_{1:N}^{[i]}) = 0$ by the non-negativity of mutual information, which further implies that $I(W_m; Z_{1:N}^{[i]} | W_{1:m-1}) = 0$. Now,

$$\begin{aligned} I(W_m; W_{m:M} | W_{1:m-1}, Z_{1:N}^{[i]}) =& H(W_m | W_{1:m-1}, Z_{1:N}^{[i]}) && (7.39) \\ =& H(W_m | W_{1:m-1}) - I(W_m; Z_{1:N}^{[i]} | W_{1:m-1}) && (7.40) \\ =& L && (7.41) \end{aligned}$$

where the last equality follows from the independence of the messages.

3. From the security constraint (7.8) and the non-negativity of mutual information, we have $I(W_{\mathcal{S}_1}, W_{\mathcal{S}_2}; Z_{1:N}^{[i]}, Q_{1:N}^{[i]}) = 0$, which can be expanded as $I(W_{\mathcal{S}_2}; Z_{1:N}^{[i]}, Q_{1:N}^{[i]}) + I(W_{\mathcal{S}_1}; Z_{1:N}^{[i]}, Q_{1:N}^{[i]} | W_{\mathcal{S}_2}) = 0$, which implies that

$I(W_{\mathcal{S}_1}; Z_{1:N}^{[i]}, Q_{1:N}^{[i]}|W_{\mathcal{S}_2}) = 0$. We futher expand it as:

$$0 = I(W_{\mathcal{S}_1}; Q_{1:N}^{[i]}|W_{\mathcal{S}_2}) + I(W_{\mathcal{S}_1}; Z_{1:N}^{[i]}|Q_{1:N}^{[i]}, W_{\mathcal{S}_2}) \tag{7.42}$$

which leads to (7.29) by the non-negativity of mutual information.

For (7.30), we note from (7.29) that $I(W_{m:M}; Z_{1:N}^{[m-1]}|Q_{1:N}^{[m-1]}, W_{1:m-1}) = 0$, hence

$$0 = I(W_{m:M}; Z_{1:N}^{[m-1]}, Q_{1:N}^{[m-1]}|W_{1:m-1}) - I(W_{m:M}; Q_{1:N}^{[m-1]}|W_{1:m-1}) \tag{7.43}$$

Now, $I(W_{m:M}; Q_{1:N}^{[m-1]}|W_{1:m-1}) = 0$ by the independence of the messages and the queries in (7.3), and this implies (7.30) by the non-negativity of mutual information.

4. Using the same argument as in item 3 above and reversing the order of the chain rule in (7.42) leads to (7.31).

5. We have

$$\begin{aligned}
&I(W_{m:M}; Q_{k+1:N}^{[m]}|W_{1:m-1}, Z_{1:N}^{[m]}, Q_{1:k}^{[m]}, Y_{1:k}^{[m]}) \\
=&I(W_{m:M}; Q_{k+1:N}^{[m]}, Y_{1:k}^{[m]}|W_{1:m-1}, Z_{1:N}^{[m]}, Q_{1:k}^{[m]}) \\
&- I(W_{m:M}; Y_{1:k}^{[m]}|W_{1:m-1}, Z_{1:N}^{[m]}, Q_{1:k}^{[m]}) \qquad (7.44) \\
=&I(W_{m:M}; Q_{k+1:N}^{[m]}|W_{1:m-1}, Z_{1:N}^{[m]}, Q_{1:k}^{[m]}) + I(W_{m:M}; Y_{1:k}^{[m]}|W_{1:m-1}, Z_{1:N}^{[m]}, Q_{1:N}^{[m]}) \\
&- I(W_{m:M}; Y_{1:k}^{[m]}|W_{1:m-1}, Z_{1:N}^{[m]}, Q_{1:k}^{[m]}) \qquad (7.45)
\end{aligned}$$

$$=0 \tag{7.46}$$

where $I(W_{m:M}; Q_{k+1:N}^{[m]} | W_{1:m-1}, Z_{1:N}^{[m]}, Q_{1:k}^{[m]}) = 0$ from (7.31) and the non-negativity of mutual information, and since $Q_{1:N}^{[m]} \rightarrow Q_{1:k}^{[m]} \rightarrow Y_{1:k}^{[m]}$ is a Markov chain, we have $I(W_{m:M}; Y_{1:k}^{[m]} | W_{1:m-1}, Z_{1:N}^{[m]}, Q_{1:N}^{[m]}) = I(W_{m:M}; Y_{1:k}^{[m]} | W_{1:m-1}, Z_{1:N}^{[m]}, Q_{1:k}^{[m]})$.

■

We will need the following lemma, which characterizes a lower bound on the interference from the undesired messages within the portion of answers that is unobserved by the eavesdropper (and hence secure). Since the user must download at least L symbols to retrieve the desired message, the difference $\sum_{n=1}^{N}(1-\mu_n)t_n - L$ denotes the interference terms within the unobserved (by the eavesdropper) portion of the answers.

Lemma 7.2 (Interference lower bound) *For the PIR-WTC-II problem, the interference from undesired messages within the unobserved portion of the answer strings by the eavesdropper $\sum_{n=1}^{N}(1-\mu_n)t_n - L$ is lower bounded by,*

$$\sum_{n=1}^{N}(1-\mu_n)t_n - L + o(L) \geq I\left(W_{2:M}; Q_{1:N}^{[1]}, Y_{1:N}^{[1]} | W_1, Z_{1:N}^{[1]}\right) \tag{7.47}$$

We note that Lemma 7.2 is a generalization of [12, Lemma 5] to the problem of PIR-WTC-II. If $\mu_n = 0$ for all $n \in [1:N]$, then Lemma 7.2 reduces to [12, Lemma 5] as $Z_{1:N}^{[1]}$ (the eavesdropper observations) is absent and $Y_{1:N}^{[1]} = A_{1:N}^{[i]}$ in that case.

Proof: We start with the right hand side of (7.47),

$$
\begin{aligned}
& I(W_{2:M}; Q_{1:N}^{[1]}, Y_{1:N}^{[1]} | W_1, Z_{1:N}^{[1]}) \\
\overset{(7.26)}{=} & I\left(W_{2:M}; W_1, Q_{1:N}^{[1]}, Y_{1:N}^{[1]} | Z_{1:N}^{[1]}\right) && (7.48) \\
= & I\left(W_{2:M}; Q_{1:N}^{[1]}, Y_{1:N}^{[1]} | Z_{1:N}^{[1]}\right) + I\left(W_{2:M}; W_1 | A_{1:N}^{[1]}, Q_{1:N}^{[1]}\right) && (7.49) \\
\overset{(7.9)}{=} & I\left(W_{2:M}; Q_{1:N}^{[1]}, Y_{1:N}^{[1]} | Z_{1:N}^{[1]}\right) + o(L) && (7.50) \\
\overset{(7.31)}{=} & I\left(W_{2:M}; Y_{1:N}^{[1]} | Q_{1:N}^{[1]}, Z_{1:N}^{[1]}\right) + o(L) && (7.51) \\
= & H\left(Y_{1:N}^{[1]} | Q_{1:N}^{[1]}, Z_{1:N}^{[1]}\right) - H\left(Y_{1:N}^{[1]} | Q_{1:N}^{[1]}, Z_{1:N}^{[1]}, W_{2:M}\right) + o(L) && (7.52) \\
\leq & \sum_{n=1}^{N} (1-\mu_n) t_n - H\left(W_1, Y_{1:N}^{[1]} | Q_{1:N}^{[1]}, Z_{1:N}^{[1]}, W_{2:M}\right) + H\left(W_1 | A_{1:N}^{[1]}, Q_{1:N}^{[1]}, W_{2:M}\right) + o(L) && (7.53) \\
\overset{(7.9)}{=} & \sum_{n=1}^{N} (1-\mu_n) t_n - H\left(W_1, Y_{1:N}^{[1]} | Q_{1:N}^{[1]}, Z_{1:N}^{[1]}, W_{2:M}\right) + o(L) && (7.54) \\
= & \sum_{n=1}^{N} (1-\mu_n) t_n - H\left(W_1 | Q_{1:N}^{[1]}, Z_{1:N}^{[1]}, W_{2:M}\right) - H\left(Y_{1:N}^{[1]} | Q_{1:N}^{[1]}, Z_{1:N}^{[1]}, W_{1:M}\right) + o(L) && (7.55) \\
\leq & \sum_{n=1}^{N} (1-\mu_n) t_n - H\left(W_1 | Q_{1:N}^{[1]}, Z_{1:N}^{[1]}, W_{2:M}\right) + o(L) && (7.56) \\
\overset{(7.27)}{=} & \sum_{n=1}^{N} (1-\mu_n) t_n - L + o(L) && (7.57)
\end{aligned}
$$

where (7.48) follows from the conditional independence of messages in Lemma 7.1, (7.50), (7.54) follow from the decodability of W_1 given $(Q_{1:N}^{[1]}, A_{1:N}^{[1]})$, (7.51) follows from the conditional independence of the messages and the queries in Lemma 7.1, (7.53) follows from conditioning reduces entropy and the fact that $H(Y_{1:N}^{[1]}) \leq$

$\sum_{n=1}^{N}(1-\mu_n)t_n$ from the WTC-II model, (7.56) follows from the non-negativity of the entropy function, and (7.57) follows from zero leakage property of W_1 from (7.27) which implies $H(W_1|Q_{1:N}^{[1]}, Z_{1:N}^{[1]}, W_{2:M}) = H(W_1) = L$. ■

In the following lemma, we derive an induction relation for the right hand side of the expression in (7.47). This lemma extends [12, Lemma 6] in two major ways. First, we incorporate the security constraint in the proof by observing that $(W_{1:M}, Z_{1:N}^{[m]})$ are independent. Second, and more significantly, the main difference between this lemma and [12, Lemma 6] is the fact that not all databases can use a symmetric scheme due to the asymmetry of the fraction that the eavesdropper can observe. Consequently, we denote n_{m-1} to be the number of databases that can apply a symmetric scheme when the retrieval problem is reduced to retrieving message W_{m-1} from the set of $W_{m-1:M}$ messages. For the remaining answer strings, we directly bound them by their corresponding length of the unobserved portion $\sum_{n=n_{m-1}+1}^{N}(1-\mu_n)t_n$.

Lemma 7.3 (Induction lemma) *For all $m \in \{2,\dots,M\}$ and for an arbitrary $n_{m-1} \in \{1,\cdots,N\}$, the mutual information term in Lemma 7.2 can be inductively lower bounded as,*

$$\begin{aligned} &I\left(W_{m:M}; Q_{1:N}^{[m-1]}, Y_{1:N}^{[m-1]} | W_{1:m-1}, Z_{1:N}^{[m-1]}\right) \\ &\quad \geq \frac{1}{n_{m-1}}\left[I\left(W_{m+1:M}; Q_{1:N}^{[m]}, Y_{1:N}^{[m]} | W_{1:m}, Z_{1:N}^{[m]}\right) + \left(L - \sum_{n=n_{m-1}+1}^{N}(1-\mu_n)t_n\right) - o(L)\right] \end{aligned} \tag{7.58}$$

Proof: We start with the left hand side of (7.58), after multiplying by n_{m-1},

$$
\begin{aligned}
& n_{m-1}\, I\left(W_{m:M}; Q_{1:N}^{[m-1]}, Y_{1:N}^{[m-1]} | W_{1:m-1}, Z_{1:N}^{[m-1]}\right) \\
& \quad \overset{(7.30)}{=} n_{m-1}\, I\left(W_{m:M}; Q_{1:N}^{[m-1]}, A_{1:N}^{[m-1]} | W_{1:m-1}\right) && (7.59) \\
& \quad \geq n_{m-1}\, I\left(W_{m:M}; Q_{1:n_{m-1}}^{[m-1]}, A_{1:n_{m-1}}^{[m-1]} | W_{1:m-1}\right) && (7.60) \\
& \quad \geq \sum_{n=1}^{n_{m-1}} I\left(W_{m:M}; Q_{n}^{[m-1]}, A_{n}^{[m-1]} | W_{1:m-1}\right) && (7.61) \\
& \quad \overset{(7.4)}{=} \sum_{n=1}^{n_{m-1}} I\left(W_{m:M}; Q_{n}^{[m]}, A_{n}^{[m]} | W_{1:m-1}\right) && (7.62) \\
& \quad \overset{(7.3)}{=} \sum_{n=1}^{n_{m-1}} I\left(W_{m:M}; A_{n}^{[m]} | Q_{n}^{[m]}, W_{1:m-1}\right) && (7.63) \\
& \quad \overset{(7.29)}{=} \sum_{n=1}^{n_{m-1}} I\left(W_{m:M}; Y_{n}^{[m]} | Q_{n}^{[m]}, W_{1:m-1}, Z_{n}^{[m]}\right) && (7.64) \\
& \quad = \sum_{n=1}^{n_{m-1}} H\left(Y_{n}^{[m]} | Q_{n}^{[m]}, W_{1:m-1}, Z_{n}^{[m]}\right) - H\left(Y_{n}^{[m]} | Q_{n}^{[m]}, W_{1:M}, Z_{n}^{[m]}\right) && (7.65) \\
& \quad \geq \sum_{n=1}^{n_{m-1}} H\left(Y_{n}^{[m]} | Y_{1:n-1}^{[m]}, Q_{1:n_{m-1}}^{[m]}, W_{1:m-1}, Z_{1:N}^{[m]}\right) \\
& \qquad - H\left(Y_{n}^{[m]} | Y_{1:n-1}^{[m]}, Q_{1:n_{m-1}}^{[m]}, W_{1:M}, Z_{1:N}^{[m]}\right) && (7.66) \\
& \quad = \sum_{n=1}^{n_{m-1}} I\left(W_{m:M}; Y_{n}^{[m]} | Y_{1:n-1}^{[m]}, Q_{1:n_{m-1}}^{[m]}, W_{1:m-1}, Z_{1:N}^{[m]}\right) && (7.67) \\
& \quad = I\left(W_{m:M}; Y_{1:n_{m-1}}^{[m]} | Q_{1:n_{m-1}}^{[m]}, W_{1:m-1}, Z_{1:N}^{[m]}\right) && (7.68) \\
& \quad \overset{(7.31)}{=} I\left(W_{m:M}; Q_{1:n_{m-1}}^{[m]}, Y_{1:n_{m-1}}^{[m]} | W_{1:m-1}, Z_{1:N}^{[m]}\right) && (7.69) \\
& \quad \overset{(7.32)}{=} I\left(W_{m:M}; Q_{1:N}^{[m]}, Y_{1:N}^{[m]} | W_{1:m-1}, Z_{1:N}^{[m]}\right) \\
& \qquad - I\left(W_{m:M}; Y_{n_{m-1}+1:N}^{[m]} | Q_{1:N}^{[m]}, Y_{1:n_{m-1}}^{[m]}, W_{1:m-1}, Z_{1:N}^{[m]}\right) && (7.70) \\
& \quad \geq I\left(W_{m:M}; Q_{1:N}^{[m]}, Y_{1:N}^{[m]} | W_{1:m-1}, Z_{1:N}^{[m]}\right) - H\left(Y_{n_{m-1}+1:N}^{[m]}\right) && (7.71) \\
& \quad \geq I\left(W_{m:M}; Q_{1:N}^{[m]}, Y_{1:N}^{[m]} | W_{1:m-1}, Z_{1:N}^{[m]}\right) - \sum_{n=n_{m-1}+1}^{N} (1-\mu_n) t_n && (7.72)
\end{aligned}
$$

$$= I\left(W_{m:M}; W_m, Q_{1:N}^{[m]}, Y_{1:N}^{[m]} | W_{1:m-1}, Z_{1:N}^{[m]}\right) - I\left(W_{m:M}; W_m | W_{1:m-1}, Q_{1:N}^{[m]}, A_{1:N}^{[m]}\right) - \sum_{n=n_{m-1}+1}^{N} (1-\mu_n) t_n \tag{7.73}$$

$$\overset{(7.9)}{=} I\left(W_{m:M}; W_m, Q_{1:N}^{[m]}, Y_{1:N}^{[m]} | W_{1:m-1}, Z_{1:N}^{[m]}\right) - \sum_{n=n_{m-1}+1}^{N} (1-\mu_n) t_n - o(L) \tag{7.74}$$

$$= I\left(W_{m:M}; W_m | W_{1:m-1}, Z_{1:N}^{[m]}\right) + I\left(W_{m:M}; Q_{1:N}^{[m]}, Y_{1:N}^{[m]} | W_{1:m}, Z_{1:N}^{[m]}\right) - \sum_{n=n_{m-1}+1}^{N} (1-\mu_n) t_n - o(L) \tag{7.75}$$

$$\overset{(7.28)}{=} I\left(W_{m+1:M}; Q_{1:N}^{[m]}, Y_{1:N}^{[m]} | W_{1:m}, Z_{1:N}^{[m]}\right) + \left(L - \sum_{n=n_{m-1}+1}^{N} (1-\mu_n) t_n\right) - o(L) \tag{7.76}$$

where (7.59) follows from the conditional independence of the messages and $Z_{1:N}^{[m-1]}$ in (7.30) as a consequence of the security constraint, (7.60), (7.61) follow from the non-negativity of mutual information, (7.62) follows from the privacy constraint, (7.63) follows from the independence of the queries and the messages, (7.64) follows from the conditional independence of the messages and $Z_n^{[m]}$ in (7.29) and the non-negativity of mutual information, (7.66) follows from conditioning reduces entropy and $\left(Q_{1:n_{m-1}}^{[m]}, Z_{1:N}^{[m]}, W_{1:M}, Y_{1:n-1}^{[m]}\right) \to \left(Q_n^{[m]}, W_{1:M}, Z_n^{[n]}\right) \to Y_n^{[m]}$, (7.69) follows from (7.31) and the non-negativity of mutual information, (7.70) follows from the chain rule and (7.32), (7.71) follows from the fact that $I\left(W_{m:M}; Y_{n_{m-1}+1:N}^{[m]} | Q_{1:N}^{[m]}, Y_{1:n_{m-1}}^{[m]}, W_{1:m-1}, Z_{1:N}^{[m]}\right) \leq H\left(Y_{1:n_{m-1}}^{[m]}\right)$, (7.72) follows from the fact that conditioning reduces entropy and $H(Y_{n_{m-1}+1:N}^{[m]}) \leq \sum_{n=n_{m-1}+1}^{N} (1-\mu_n) t_n$ in the WTC-II model, (7.74) follows from the reliability constraint, (7.76) follows from the no leakage property of W_m from (7.28) as a consequence of the security

constraint. Finally, dividing both sides by n_{m-1} leads to (7.58). ■

Now, we are ready to prove an explicit upper bound for the retrieval rate in the PIR-WTC-II problem $R(\boldsymbol{\mu})$ by applying Lemma 7.2 and Lemma 7.3 successively. For a pre-specified answer string lengths $\{t_n\}_{n=1}^N$, and an arbitrary sequence $\{n_i\}_{i=1}^{M-1}$, we can write

$$
\begin{aligned}
&\sum_{n=1}^{N}(1-\mu_n)t_n - L + \tilde{o}(L) \\
&\overset{(7.47)}{\geq} I\left(W_{2:M}; Q_{1:N}^{[1]}, Y_{1:N}^{[1]} | W_1, Z_{1:N}^{[1]}\right) && (7.77)\\
&\overset{(7.58)}{\geq} \frac{1}{n_1}\left(L - \sum_{n=n_1+1}^{N}(1-\mu_n)t_n\right) + \frac{1}{n_1} I\left(W_{3:M}; Q_{1:N}^{[2]}, Y_{1:N}^{[2]} | W_{1:2}, Z_{1:N}^{[2]}\right) && (7.78)\\
&\overset{(7.58)}{\geq} \frac{1}{n_1}\left(L - \sum_{n=n_1+1}^{N}(1-\mu_n)t_n\right) + \frac{1}{n_1 n_2}\left(L - \sum_{n=n_2+1}^{N}(1-\mu_n)t_n\right) \\
&\quad + \frac{1}{n_2} I\left(W_{4:M}; Q_{1:N}^{[3]}, Y_{1:N}^{[3]} | W_{1:3}, Z_{1:N}^{[3]}\right) && (7.79)\\
&\overset{(7.58)}{\geq} \cdots \\
&\overset{(7.58)}{\geq} \frac{1}{n_1}\left(L - \sum_{n=n_1+1}^{N}(1-\mu_n)t_n\right) + \frac{1}{n_1 n_2}\left(L - \sum_{n=n_2+1}^{N}(1-\mu_n)t_n\right) + \cdots \\
&\quad + \frac{1}{\prod_{i=1}^{M-1} n_i}\left(L - \sum_{n=n_{M-1}+1}^{N}(1-\mu_n)t_n\right) && (7.80)
\end{aligned}
$$

where $\tilde{o}(L) = \left(1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}\right) o(L)$, (7.77) follows from Lemma 7.2, and the remaining bounding steps follow from successive application of Lemma 7.3.

Ordering terms and letting $\tau_n = \frac{t_n}{\sum_{i=1}^{N} t_i}$, we have,

$$\left(1+\frac{1}{n_1}+\frac{1}{n_1 n_2}+\cdots+\frac{1}{\prod_{i=1}^{M-1} n_i}\right)L \leq \left(\phi(0)+\frac{\phi(n_1)}{n_1}+\cdots+\frac{\phi(n_{M-1})}{\prod_{i=1}^{M-1} n_i}\right)\sum_{n=1}^{N} t_n+\tilde{o}(L) \tag{7.81}$$

where $\phi(\ell) = \sum_{n=\ell+1}^{N}(1-\mu_n)\tau_n$ corresponds to the sum of the unobserved traffic ratios by the eavesdropper from databases $[\ell+1:N]$.

We conclude the proof by taking $L \to \infty$. Thus, for an arbitrary sequence $\{n_i\}_{i=1}^{M-1}$ we have

$$R(\boldsymbol{\tau},\boldsymbol{\mu}) = \frac{L}{\sum_{n=1}^{N} t_n} \leq \frac{\phi(0)+\frac{\phi(n_1)}{n_1}+\frac{\phi(n_2)}{n_1 n_2}+\cdots+\frac{\phi(n_{M-1})}{\prod_{i=1}^{M-1} n_i}}{1+\frac{1}{n_1}+\frac{1}{n_1 n_2}+\cdots+\frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{7.82}$$

The bound in (7.82) for $R(\boldsymbol{\tau},\boldsymbol{\mu})$ is valid for any arbitrary sequence $\{n_i\}_{i=1}^{M-1}$. Hence, we obtain the tightest upper bound for $R(\boldsymbol{\tau},\boldsymbol{\mu})$ by minimizing over the sequence $\{n_i\}_{i=1}^{M-1}$ over the set $\{1,\cdots,N\}$ to get

$$R(\boldsymbol{\tau},\boldsymbol{\mu}) \leq \min_{n_1,\cdots,n_{M-1}\in\{1,\cdots,N\}} \frac{\phi(0)+\frac{\phi(n_1)}{n_1}+\frac{\phi(n_2)}{n_1 n_2}+\cdots+\frac{\phi(n_{M-1})}{\prod_{i=1}^{M-1} n_i}}{1+\frac{1}{n_1}+\frac{1}{n_1 n_2}+\cdots+\frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{7.83}$$

Finally, since the user and the databases can choose any suitable traffic ratio vector $\boldsymbol{\tau}$ in the set $\mathbb{T}$ such that:

$$\mathbb{T} = \left\{\boldsymbol{\tau} : \tau_n \geq 0 \quad \forall n \in [1:N], \quad \sum_{n=1}^{N} \tau_n = 1\right\} \tag{7.84}$$

by maximizing over $\boldsymbol{\tau} = (\tau_1, \tau_2, \cdots, \tau_N)$ in the set $\mathbb{T}$, we obtain the following upper bound for $R(\boldsymbol{\mu})$,

$$R(\boldsymbol{\mu}) \leq \max_{\boldsymbol{\tau}\in\mathbb{T}} \min_{n_i\in\{1,\cdots,N\}} \frac{\phi(0) + \frac{\phi(n_1)}{n_1} + \frac{\phi(n_2)}{n_1 n_2} + \cdots + \frac{\phi(n_{M-1})}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{7.85}$$

$$= \max_{\boldsymbol{\tau}\in\mathbb{T}} \min_{n_i\in\{1,\cdots,N\}} \frac{\sum_{n=1}^{N}(1-\mu_n)\tau_n + \frac{\sum_{n=n_1+1}^{N}(1-\mu_n)\tau_n}{n_1} + \cdots + \frac{\sum_{n=n_{M-1}+1}^{N}(1-\mu_n)\tau_n}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{7.86}$$

7.5 Achievable Scheme

In this section, we present a general achievable scheme for PIR-WTC-II. The scheme builds on the achievable scheme in Chapter 5. The main idea of the achievable scheme is that since the databases are eavesdropped by varying eavesdropping capabilities $\boldsymbol{\mu}$, then it would be beneficial for the user to query the databases using the PIR scheme under asymmetric traffic constraints. Furthermore, the databases should *encrypt* the answers such that the user can decode the *meaningful* transmission by observing the entire answer string, while the encryption keys span the eavesdropper's entire observation space, ensuring the security of downloaded content. The user and the databases agree on the traffic ratio vector $\boldsymbol{\tau}$ that maximizes the achievable secure PIR rate.

In the following, we illustrate the main ingredients of the achievable scheme by presenting the case of $M = 3$ messages and $N = 2$ databases for an arbitrary $\boldsymbol{\mu}$.

7.5.1 Motivating Example: $M = 3$ Messages, $N = 2$ Databases

In this section, we first show an explicit upper bound for the capacity expression $\bar{C}(\boldsymbol{\mu})$. Then, we show the capacity-achieving scheme for the concrete example of $\boldsymbol{\mu} = (\frac{1}{4}, \frac{1}{2})$. We conclude this section by showing how to extend the achievable scheme for arbitrary $\boldsymbol{\mu}$.

7.5.1.1 Explicit Upper Bound for $M = 3$ Messages, $N = 2$ Databases

From Theorem 7.1, the upper bound of $\bar{C}(\boldsymbol{\mu})$ is given by:

$$\bar{C}(\boldsymbol{\mu}) = \max_{\boldsymbol{\tau} \in \mathbb{T}} \min_{n_i \in \{1,2\}} \frac{\sum_{n=1}^{2}(1-\mu_n)\tau_n + \frac{\sum_{n=n_1+1}^{2}(1-\mu_n)\tau_n}{n_1} + \frac{\sum_{n=n_2+1}^{2}(1-\mu_n)\tau_n}{n_1 n_2}}{1 + \frac{1}{n_1} + \frac{1}{n_1 n_2}} \tag{7.87}$$

By observing that $\tau_1 = 1 - \tau_2$, this can be explicitly written as the following linear program:

$$\begin{aligned}
\max_{\tau_2, R} \quad & R \\
\text{s.t.} \quad & R \leq \frac{1}{3}(1-\mu_1) + \left[(1-\mu_2) - \frac{1}{3}(1-\mu_1)\right]\tau_2 \\
& R \leq \frac{2}{5}(1-\mu_1) + \left[\frac{4}{5}(1-\mu_2) - \frac{2}{5}(1-\mu_1)\right]\tau_2 \\
& R \leq \frac{4}{7}(1-\mu_1) + \left[\frac{4}{7}(1-\mu_2) - \frac{4}{7}(1-\mu_1)\right]\tau_2 \\
& 0 \leq \tau_2 \leq 1
\end{aligned} \tag{7.88}$$

Note that the bound corresponding to $n_1 = 2, n_2 = 1$ is not included in (7.88) as it would be inactive for any $\boldsymbol{\mu}$. Since (7.88) is a linear program, the optimal solution

exists among the corner points of the feasible region. The first corner point, is $\tau_2^{(1)} = 0$, which leads to the bound $\bar{C}(\boldsymbol{\mu}) \leq \frac{1-\mu_1}{3}$. The second corner point occurs at the intersection of the first two constraints, i.e., $\tau_2^{(2)}$ satisfies:

$$\frac{1}{3}(1-\mu_1) + \left[(1-\mu_2) - \frac{1}{3}(1-\mu_1)\right]\tau_2^{(2)} = \frac{2}{5}(1-\mu_1) + \left[\frac{4}{5}(1-\mu_2) - \frac{2}{5}(1-\mu_1)\right]\tau_2^{(2)} \tag{7.89}$$

which leads to,

$$\tau_2^{(2)} = \frac{(1-\mu_1)}{3(1-\mu_2)+(1-\mu_1)} \tag{7.90}$$

with a corresponding bound of $\bar{C}(\boldsymbol{\mu}) \leq \frac{2(1-\mu_1)(1-\mu_2)}{3(1-\mu_2)+(1-\mu_1)}$. Similarly, the third corner point $\tau_2^{(3)}$ occurs at the intersection of the second and third constraints, hence $\tau_2^{(3)} = \frac{3(1-\mu_1)}{4(1-\mu_2)+3(1-\mu_1)}$ with the corresponding bound of $\bar{C}(\boldsymbol{\mu}) \leq \frac{4(1-\mu_1)(1-\mu_2)}{4(1-\mu_2)+3(1-\mu_1)}$. Finally, at $\tau_2 = 1$, we have the bound $\bar{C}(\boldsymbol{\mu}) \leq \frac{4(1-\mu_2)}{7}$ which is no larger than $\frac{4(1-\mu_1)(1-\mu_2)}{4(1-\mu_2)+3(1-\mu_1)}$ by the monotonicity of $\boldsymbol{\mu}$, hence it can be ignored.

Consequently, the explicit upper bound for $M = 3$, $N = 2$ is given by

$$\bar{C}(\boldsymbol{\mu}) = \max\left\{\frac{1-\mu_1}{3},\ \frac{2(1-\mu_1)(1-\mu_2)}{3(1-\mu_2)+(1-\mu_1)},\ \frac{4(1-\mu_1)(1-\mu_2)}{4(1-\mu_2)+3(1-\mu_1)}\right\} \tag{7.91}$$

7.5.1.2 Concrete Example: $\mu_1 = \frac{1}{4}$, $\mu_2 = \frac{1}{2}$

Before the retrieval process, the user permutes the indices of the symbols of W_1, W_2, W_3 independently, uniformly, and privately. Assume without loss of generality

that W_1 is the desired message. Let a_i, b_i, c_i be the permuted symbols from W_1, W_2, W_3, respectively. In the case of $\mu_1 = \frac{1}{4}$, $\mu_2 = \frac{1}{2}$, the explicit upper bound in (7.91) is $\bar{C}(\boldsymbol{\mu}) = \frac{4(1-\mu_1)(1-\mu_2)}{4(1-\mu_2)+3(1-\mu_1)} = \frac{6}{17}$. To achieve this bound, we focus first on the *meaningful* queries, i.e., the queries without the randomness that is added to satisfy the security constraint. From the first database, the user asks for an individual symbol from every message, i.e., asks for a_1, b_1, c_1. From database 2, the user does not ask for new individual symbols but rather exploits the side information that is generated from database 1 to query for 2-sums from database 2, i.e., the user asks for $a_2 + b_1$, $a_3 + c_1$, $b_2 + c_2$ from database 2. Then, the user exploits $b_2 + c_2$ as side information to ask for $a_4 + b_2 + c_2$ from database 1. To get an integer number of downloads for the meaningful queries, which covers $(1 - \mu_n)t_n$ from the downloaded symbols from the nth database, the scheme is repeated ν times. Since this scheme gets 4 symbols from database 1 and 3 symbols from database 2, we choose the repetition factor of the scheme ν such that:

$$(1 - \mu_1)t_1 = 4\nu \;\Rightarrow\; t_1 = \frac{16\nu}{3} \tag{7.92}$$

$$(1 - \mu_2)t_2 = 3\nu \;\Rightarrow\; t_2 = 6\nu \tag{7.93}$$

Then, the minimal ν is $\nu = 3$. Database 1 generates the independent keys $K_1 = \left(k_1^{(1)}, \cdots, k_4^{(1)}\right)$, such that K_1 is picked uniformly from $\mathbb{F}_q^4$. Database 1 encodes these random keys using a $(16, 4)$ MDS code, to get $u_{[1:16]}$, i.e.,

$$u_{[1:16]} = \mathbf{MDS}_{16\times 4} K_1 \tag{7.94}$$

Similarly, database 2 generates $K_2 = \left(k_1^{(2)}, \cdots, k_9^{(2)}\right)$ uniformly from $\mathbb{F}_q^9$. Database 2 encodes the keys using an $(18, 9)$ MDS code, to get $v_{[1:16]}$, i.e.,

$$v_{[1:18]} = \mathbf{MDS}_{18\times 9} K_2 \tag{7.95}$$

Now, all the meaningful downloads are *encrypted* by the coded keys. Furthermore, the user downloads $u_{[13:16]}$ separately from database 1, and $v_{[10:18]}$ from database 2. The query table is shown in Table. 7.1.

Table 7.1: The query table for $M = 3$, $N = 2$, $\mu_1 = \frac{1}{4}$, $\mu_2 = \frac{1}{2}$.

Database 1	Database 2
$a_1 + u_1$ $b_1 + u_2$ $c_1 + u_3$	$a_2 + b_1 + v_1$ $a_3 + c_1 + v_2$ $b_2 + c_2 + v_3$
$a_4 + b_2 + c_2 + u_4$	
$a_5 + u_5$ $b_3 + u_6$ $c_3 + u_7$	$a_6 + b_3 + v_4$ $a_7 + c_3 + v_5$ $b_4 + c_4 + v_6$
$a_8 + b_4 + c_4 + u_8$	
$a_9 + u_9$ $b_5 + u_{10}$ $c_5 + u_{11}$	$a_{10} + b_5 + v_7$ $a_{11} + c_5 + v_8$ $b_6 + c_6 + v_9$
$a_{12} + b_6 + c_6 + u_{12}$	
$u_{13}, u_{14}, u_{15}, u_{16}$	$v_{10}, u_{11}, u_{12}, u_{13}, v_{14}$ $v_{15}, u_{16}, u_{17}, u_{18}$

For the decodability, since database 1 encodes its keys K_1 using a $(16, 4)$ MDS code, by the MDS property, any 4 symbols suffice to reconstruct $u_{[1:16]}$. The user downloads $u_{[13:16]}$ separately, hence $u_{[1:12]}$ can be reconstructed and canceled from the downloads to get the meaningful information only. Similarly, database 2 encodes the keys K_2 using an $(18, 9)$ MDS code, hence $v_{[10:18]}$ suffice to reconstruct $v_{[1:9]}$

and can be canceled from the meaningful downloads. Furthermore, since the side information at any database is obtained from the undesired symbols downloaded from the second database, all undesired symbols can be canceled and the user is left only with $a_{[1:12]}$, which are the desired symbols.

For the security, since $\mu_1 = \frac{1}{4}$ and $\mu_2 = \frac{1}{2}$, the eavesdropper can obtain any 4 symbols out of total 16 downloaded symbols from database 1, and any 9 symbols out of total 18 downloaded symbols from database 2. Since K_1, K_2 are generated uniformly and independently from $\mathbb{F}_q^4$, $\mathbb{F}_q^9$, respectively, any 4 symbols $(u_{i_1}, \cdots, u_{i_4})$ from $u_{[1:16]}$ are independent and uniformly distributed over $\mathbb{F}_q$, and similarly for any 9 symbols $(v_{j_1}, \cdots, v_{j_9})$ from $v_{[1:18]}$. Consequently, the leakage at the eavesdropper is upper bounded by:

$$I(W_{1:3}; Z_{1:2}^{[1]}) = H(Z_{1:2}) - H(Z_{1:2}|W_{1:3}) \tag{7.96}$$

$$\leq \log_q 13 - H\left(\begin{bmatrix} u_{i_1} \\ \vdots \\ u_{i_4} \\ v_{j_1} \\ \vdots \\ v_{j_9} \end{bmatrix}\right) = 0 \tag{7.97}$$

For the privacy, as all combinations of the sums are included in the queries and the indices of the message symbols are uniformly and independently permuted, the privacy constraint is satisfied. Hence, the user downloads $t_1 = 16$ symbols from

database 1, and $t_2 = 18$ symbols from database 2. From these downloads, the user can decode $L = 12$ symbols from W_1. Hence, $R = \frac{12}{34} = \frac{6}{17}$, which matches the upper bound.

7.5.1.3 Achieving the Upper Bound for Arbitrary $\boldsymbol{\mu}$

Now, we show how to achieve the upper bound in (7.91) for general $\boldsymbol{\mu}$. As shown in the example of $\mu_1 = \frac{1}{4}$, $\mu_2 = \frac{1}{2}$, the user downloads $\mu_1 t_1$ as individual symbols from the coded keys from database 1, and $\mu_2 t_2$ as individual symbols from the coded keys from database 2. This leaves $(1-\mu_1)t_1$, $(1-\mu_2)t_2$, respectively for meaningful symbols. Furthermore, each scheme should be repeated ν times to ensure that t_1, $t_2 \in \mathbb{N}$. In the following, we focus on the meaningful symbols without the coded keys. We show only one repetition of the scheme.

For $R(\boldsymbol{\mu}) = \frac{1-\mu_1}{3}$: To achieve this rate, the user applies the trivial retrieval scheme [1], and downloads all messages from database 1, i.e., the user downloads a_1, b_1, c_1 from database 1. Hence, $t_2 = 0$ and

$$(1-\mu_1)t_1 = 3\nu \;\Rightarrow\; t_1 = \frac{3\nu}{1-\mu_1} \tag{7.98}$$

where ν is chosen such that $t_1 \in \mathbb{N}$. From every repetition, the user gets 1 symbol from W_1. Hence, $L = \nu$. The user asks for $\mu_1 t_1 = \frac{3\mu_1\nu}{1-\mu_1}$ individual coded symbols from the keys, and the database encrypts the downloads with coded keys constructed from a $(\frac{3\nu}{1-\mu_1}, \frac{3\mu_1\nu}{1-\mu_1})$ MDS code. This ensures the security. The achievable rate in

this case is

$$R = \frac{L}{t_1 + t_2} = \frac{\nu}{\frac{3\nu}{1-\mu_1}} = \frac{1-\mu_1}{3} \tag{7.99}$$

For $R(\boldsymbol{\mu}) = \frac{2(1-\mu_1)(1-\mu_2)}{3(1-\mu_2)+(1-\mu_1)}$: To achieve this rate, the user downloads individual symbols from all messages from database 1, i.e., the user downloads a_1, b_1, c_1 from database 1. The user combines the two undesired symbols b_1, c_1 into a 2-sum $b_1 + c_1$ and uses it as a side information in database 2. The query table for one repetition of the scheme for the meaningful symbols (without showing the keys) is shown in Table 7.2.

Table 7.2: The meaningful symbols for $M = 3$, $N = 2$ to achieve $\frac{2(1-\mu_1)(1-\mu_2)}{3(1-\mu_2)+(1-\mu_1)}$.

Database 1	Database 2
a_1, b_1, c_1	$a_2 + b_1 + c_1$

In this case, the scheme is repeated ν times such that t_1, $t_2 \in \mathbb{N}$,

$$(1-\mu_1)t_1 = 3\nu \;\Rightarrow\; t_1 = \frac{3\nu}{1-\mu_1} \tag{7.100}$$

$$(1-\mu_2)t_2 = 1\nu \;\Rightarrow\; t_2 = \frac{\nu}{1-\mu_2} \tag{7.101}$$

Database 1 encodes $\mu_1 t_1 = \frac{3\nu\mu_1}{1-\mu_1}$ independent and uniformly distributed keys using a $(\frac{3\nu}{1-\mu_1}, \frac{3\nu\mu_1}{1-\mu_1})$ MDS code to obtain the coded keys that are added to each download. Similarly, database 2 encodes $\mu_2 t_2 = \frac{\nu\mu_2}{1-\mu_2}$ keys using a $(\frac{\nu}{1-\mu_2}, \frac{\nu\mu_2}{1-\mu_2})$ MDS code to obtain the coded symbols. Using this scheme, the user decodes $L = 2\nu$ from

the desired messages. Consequently,

$$R = \frac{L}{t_1 + t_2} = \frac{2\nu}{\frac{3\nu}{1-\mu_1} + \frac{\nu}{1-\mu_2}} = \frac{2(1-\mu_1)(1-\mu_2)}{3(1-\mu_2) + (1-\mu_1)} \tag{7.102}$$

For $R(\boldsymbol{\mu}) = \frac{4(1-\mu_1)(1-\mu_2)}{4(1-\mu_2)+3(1-\mu_1)}$: An instance for this scheme is the $\mu_1 = \frac{1}{4}$, $\mu_2 = \frac{1}{2}$ example. To avoid repetition, we give only the general rate. As shown in the example, $t_1 = \frac{4\nu}{1-\mu_1}$, and $t_2 = \frac{3\nu}{1-\mu_2}$. From every repetition, the user can decode 4 symbols, hence $L = 4\nu$. Thus,

$$R = \frac{L}{t_1 + t_2} = \frac{4\nu}{\frac{4\nu}{1-\mu_1} + \frac{3\nu}{1-\mu_2}} = \frac{4(1-\mu_1)(1-\mu_2)}{4(1-\mu_2) + 3(1-\mu_1)} \tag{7.103}$$

This completes the description of the capacity-achieving scheme for PIR-WTC-II for $M = 3$, $N = 2$, and arbitrary $\boldsymbol{\mu}$. The capacity region $C(\boldsymbol{\mu})$ is shown in Fig. 7.4. In Fig. 7.5, we illustrate the partitioning of the $\boldsymbol{\mu}$ space in terms of the active capacity expression; note by convention $\mu_2 \geq \mu_1$.

7.5.2 General Achievable Scheme

In this section, we present the general achievable scheme for PIR-WTC-II that achieves the retrieval rate in Theorem 7.2. The core of the achievable scheme is the achievable scheme of the corner points in the PIR problem under asymmetric traffic constraints in Chapter 5. A new ingredient is needed to satisfy the security constraint, namely, encrypting the answer strings by random keys. The nth database uses a random key K_n of length $\mu_n t_n$ that is sufficient to span the space of the

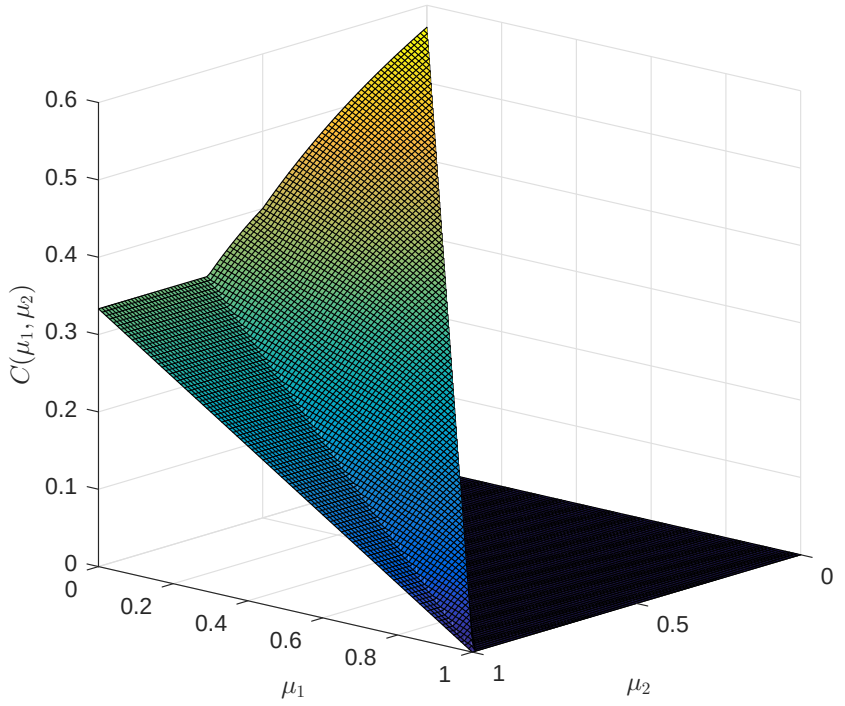


Figure 7.4: Capacity for $M = 3$, $N = 2$ as a function of μ_1 and μ_2.

eavesdropper's observations. The nth database encodes K_n using a $(t_n, \mu_n t_n)$ MDS code and uses the resulting codeword to *encrypt* each downloaded symbol from the meaningful downloads in addition to $\mu_n t_n$ individual symbols of coded key symbols only. For completeness, we include all related details of the scheme in [125] in addition to the new ingredients.

We use the same terminology as in [125]. Let $s_n \in \{0, 1, \cdots, M-1\}$ denote the number of side information symbols that are used simultaneously in the initial round of downloads at the nth database. For a given non-decreasing sequence $\{n_i\}_{i=0}^{M-1} \subset \{1, \cdots, N\}^M$, the databases are divided into groups, such that group 0 contains database 1 through database n_0, group 1 contains $n_1 - n_0$ databases starting from database $n_0 + 1$, and so on.

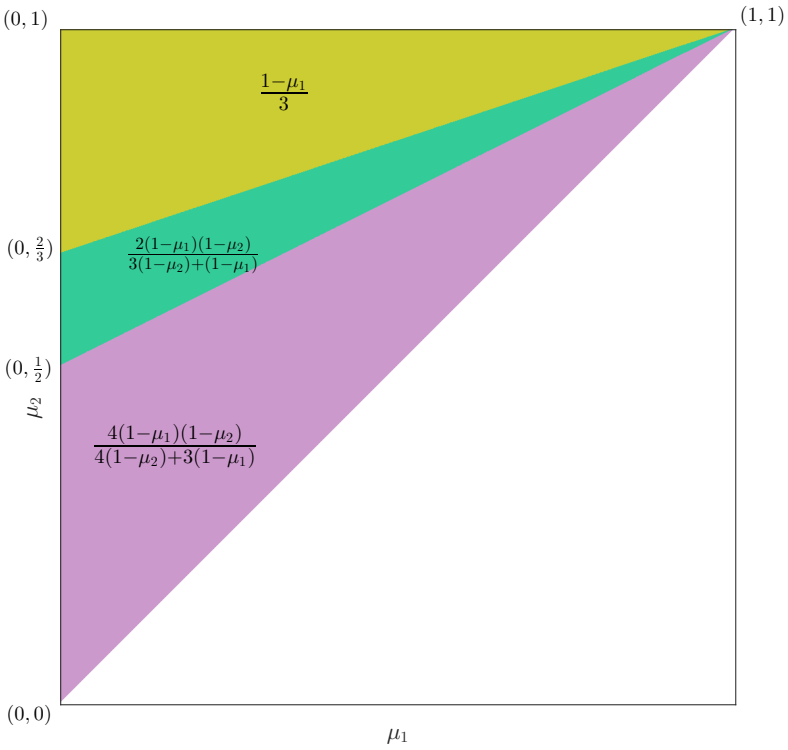


Figure 7.5: Partitions of $\boldsymbol{\mu}$ space according to the active capacity expression for $M = 3$, $N = 2$.

Hence, let $s_n = i$ for all $n_{i-1}+1 \leq n \leq n_i$ with $n_{-1} = 0$ by convention. Denote $\mathcal{S} = \{i : s_n = i \,\text{for some}\, n \in \{1, \cdots, N\}\}$. We follow the round and stage definitions in [123]. The kth round is the download queries that admit a sum of k different messages (k-sum in [12]). A stage of the kth round is a query block of the kth round that exhausts all $\binom{M}{k}$ combinations of the k-sum. Denote $y_\ell[k]$ to be the number of stages in round k downloaded from the nth database, such that $n_{\ell-1} + 1 \leq n \leq n_\ell$. The details of the achievable scheme are as follows:

1. *Calculation of the number of repetitions:* The user and the databases agree on appropriate answer string lengths $t_n(\mathbf{n}, \boldsymbol{\mu})$, $n = 1, \cdots, N$. To that end, the

scheme associated with $\mathbf{n} = \{n_i\}_{i=0}^{M-1}$ is repeated ν times such that:

$$t_n(\mathbf{n}, \boldsymbol{\mu}) = \frac{\nu D_n(\mathbf{n})}{1 - \mu_n} \in \mathbb{N}, \quad \forall n \in \{1, \cdots, N\} \tag{7.104}$$

where $D_n(\mathbf{n})$ is the number of meaningful downloads corresponding to one repetition of the achievable scheme associated with the monotone non-decreasing sequence $\mathbf{n} = \{n_i\}_{i=0}^{M-1}$.

2. *Preparation of the keys:* The nth database generates a random key K_n. The random key K_n is of length $\mu_n t_n$, such that elements of K_n are independent and uniformly distributed over $\mathbb{F}_q$. The nth database encodes K_n to an *artificial noise* vector $u_{[1:t_n]}^{(n)}$ using a $(t_n, \mu_n t_n)$ MDS code, i.e.,

$$u_{[1:t_n]}^{(n)} = \mathbf{MDS}_{t_n \times \mu_n t_n} K_n \tag{7.105}$$

3. *Initialization at the user side:* The user permutes each message independently and uniformly using a random interleaver, i.e.,

$$x_m(i) = W_m(\pi_m(i)), \quad i \in \{1, \cdots, L\} \tag{7.106}$$

where $x_m(i)$ is the ith symbol of the permuted W_m, $\pi_m(\cdot)$ is a random interleaver for the mth message that is chosen independently, uniformly, and privately at the user's side.

4. *Initial download:* From the nth database where $1 \leq n \leq n_0$, the user down-

loads $\prod_{s\in\mathcal{S}}\binom{M-2}{s-1}$ symbols from the desired message. The user sets the round index $k=1$. I.e., the user starts downloading the desired symbols from $y_0[1]=\prod_{s\in\mathcal{S}}\binom{M-2}{s-1}$ different stages.

5. *Message symmetry:* To satisfy the privacy constraint, for each stage initiated in the previous step, the user completes the stage by downloading the remaining $\binom{M-1}{k-1}$ k-sum combinations that do not include the desired symbols, in particular, if $k=1$, the user downloads $\prod_{s\in\mathcal{S}}\binom{M-2}{s-1}$ individual symbols from each undesired message.

6. *Database symmetry:* We divide the databases into groups. Group $\ell\in\mathcal{S}$ corresponds to databases $n_{\ell-1}+1$ to n_ℓ. Database symmetry is applied within each group only. Consequently, the user repeats step 2 over each group of databases, in particular, if $k=1$, the user downloads $\prod_{s\in\mathcal{S}}\binom{M-2}{s-1}$ individual symbols from each message from the first n_0 databases (group 1).

7. *Exploitation of side information:* The initial exploitation of side information is group-dependent as well. Specifically, the undesired symbols downloaded within the kth round (the k-sums that do not include the desired message) are used as side information in the $(k+1)$th round. This exploitation of side information is performed by downloading $(k+1)$-sum consisting of 1 desired symbol and a k-sum of undesired symbols only that were generated in the kth round. However, the main difference from [12] is that, for the nth database, if $s_n>k$, then this database does not exploit the side information generated in the kth round. Consequently, the nth database belonging to the ℓth group

exploits the side information generated in the kth round from all databases except itself if $s_n \leq k$. Moreover, for $s_n = k$, extra side information can be used in the nth database. This is due to the fact that the user can form $n_0 \prod_{s \in \mathcal{S} \setminus \{s_n\}} \binom{M-2}{s-1}$ extra stages of side information by constructing k-sums of the undesired symbols in round 1 from the databases in group 0.

8. *Repeat* steps 5, 6, 7 after setting $k = k + 1$ until $k = M$.

9. *Repetition of the scheme:* Repeat steps $4, \cdots, 8$ for a total of ν repetitions.

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

11. *Encryption of the downloads:* The database encrypts each meaningful download by adding one symbol from $u^{(n)}_{[1:(1-\mu_n)t_n]}$. Furthermore, the user downloads $u^{(n)}_{[(1-\mu_n)t_n+1:t_n]}$ coded key symbols individually. This guarantees the security.

7.5.3 Decodability, Privacy, Security, and Achievable Rate

Decodability: To see the decodability, we note that the user receives $\mu_n t_n$ individual artificial noise symbols $u^{(n)}_{[(1-\mu_n)t_n+1:t_n]}$ from the nth database. From the MDS property of the $(t_n, \mu_n t_n)$ MDS code, any $\mu_n t_n$ coded symbols suffice to reconstruct the entire t_n coded symbols. Hence, the user can reconstruct and cancel $u^{(n)}_{[1:t_n]}$ by the knowledge of $u^{(n)}_{[(1-\mu_n)t_n+1:t_n]}$. Consequently, after canceling the artificial noise symbols, the user is left with only the meaningful symbols in the answer strings.

Now, by construction, in the $(k+1)$th round at the nth database, the user exploits the side information generated in the kth round in the remaining active databases by adding 1 symbol of the desired message with k-sum of undesired messages which was downloaded previously in the kth round. Moreover, for the nth database belonging to the ℓth group at the $(\ell+1)$th round, the user adds every ℓ symbols of the undesired symbols downloaded from group 0 to make one side information symbol. Since the user downloads $\prod_{\ell \in \mathcal{S}} \binom{M-2}{\ell-1}$ from every database in the first n_0 databases (group 0), the user can exploit such side information to initiate $n_0 \prod_{\ell \in \mathcal{S} \setminus \{\ell\}} \binom{M-2}{\ell-1}$ stages in the $(\ell+1)$th round from every database in group ℓ. Since all side information symbols used in the $(k+1)$th round is decodable in the kth round or from round 1, the user cancels out these side information and is left with symbols from the desired message.

Privacy: The privacy of the scheme follows from the privacy of the inherent PIR scheme under asymmetric traffic constraints. Specifically, for every stage of the kth round initiated in the exploitation of the side information step, all $\binom{M}{k}$ combinations of the k-sum are included at each round. Thus, the structure of the queries is the same for any desired message. The privacy constraint in (7.4) is satisfied by the random and independent permutation of each message and the random shuffling of the order of the queries. This ensures that all queries are equally likely independent of the desired message index.

Security: From the nth database key K_n is of length $\mu_n t_n$. The elements of K_n are independent and uniformly distributed in $\mathbb{F}_q$. The nth database encodes K_n into the artificial noise vector $u^{(n)}_{[1:t_n]}$ using a $(t_n, \mu_n t_n)$ MDS code. Since any $\mu_n t_n$ columns of the generator matrix of the MDS code are full rank, the mapping from K_n to any $\mu_n t_n$ symbols from the artificial noise vector $U_n = [u^{(n)}_{i_1}, \cdots, u^{(n)}_{i_{\mu_n t_n}}]$ is a bijection, and consequently, $U_n \sim K_n$, where $\sim$ denotes statistical equivalence. Moreover, since there is no shared randomness between databases, the elements of $(K_1, \cdots, K_N)$, and consequently the elements of $(U_1, \cdots, U_N)$ are independent and uniformly distributed in $\mathbb{F}_q$.

Now, the eavesdropper chooses to observe $\mu_n t_n$ symbols from the nth answer string $A^{[i]}_n$. Denote the eavesdropper observations by $Z^{[i]}_n \in \mathbb{F}_q^{\mu_n t_n}$. Since all downloaded symbols are encrypted using $u^{(n)}_{[1:t_n]}$ (counting the downloads that contain solely the artificial noise). Denote the artificial noise symbols within $Z^{[i]}_n$ by U_n. Hence, the leakage at the eavesdropper can be upper bounded by:

$$I(W_{1:M}; Z^{[i]}_{1:N}) = H(Z^{[i]}_{1:N}) - H(Z^{[i]}_{1:N}|W_{1:M}) \tag{7.107}$$

$$\leq \sum_{n=1}^{N} \mu_n t_n - H\left(\begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{bmatrix}\right) \tag{7.108}$$

$$= \sum_{n=1}^{N} \mu_n t_n - \sum_{n=1}^{N} \mu_n t_n = 0 \tag{7.109}$$

where (7.109) follows from the fact that any $\mu_n t_n$ artificial noise symbols are inde-

pendent. Note that the units of calculation is q-ary symbols.

Achievable Rate: For the calculation of the achievable rate, we focus first on one repetition of the scheme. Without adding the artificial noise symbols, the structure of one repetition of our scheme is exactly as [125]. The recursive structure of the achievable scheme can be described using the following system of difference equations that relate the number of stages in the databases belonging to a specific group as shown in [125, Theorem 2]:

$$\begin{aligned}
y_0[k] &= (n_0-1)y_0[k-1] + \sum_{j\in\mathcal{S}\setminus\{0\}} (n_j-n_{j-1})y_j[k-1] \\
y_1[k] &= (n_1-n_0-1)y_1[k-1] + \sum_{j\in\mathcal{S}\setminus\{1\}} (n_j-n_{j-1})y_j[k-1] \\
y_\ell[k] &= n_0\xi_\ell\delta[k-\ell-1] + (n_\ell-n_{\ell-1}-1)y_\ell[k-1] + \sum_{j\in\mathcal{S}\setminus\{\ell\}} (n_j-n_{j-1})y_j[k-1], \quad \ell\geq 2
\end{aligned} \tag{7.110}$$

where $y_\ell[k]$ is the number of stages in the kth round in a database belonging to the ℓth group, i.e., for the nth database, such that $n_{\ell-1}+1 \leq n \leq n_\ell$.

Hence, to calculate $D_n(\mathbf{n})$ such that $n_{\ell-1} \leq n \leq n_\ell$, which is the number of meaningful downloads from the nth database belonging to the ℓth group, corresponding to one repetition of the achievable scheme associated with the sequence $\mathbf{n} = \{n_i\}_{i=0}^{M-1}$, we note that for any stage in the kth round, the user downloads $\binom{M-1}{k-1}$

desired symbols from a total of $\binom{M}{k}$ downloads. Therefore,

$$D_n(\mathbf{n}) = \sum_{k=1}^{M} \binom{M}{k} y_\ell[k], \quad n_{\ell-1} \leq n \leq n_\ell \tag{7.111}$$

Consequently, the total download $\sum_{n=1}^{N} t_n(\mathbf{n})$ from all databases from all repetitions is calculated by observing (7.104),

$$\sum_{n=1}^{N} t_n(\mathbf{n}, \boldsymbol{\mu}) = \sum_{n=1}^{N} \frac{\nu D_n(\mathbf{n})}{1-\mu_n} \tag{7.112}$$

$$= \nu \left[\sum_{n=1}^{n_0} \frac{\sum_{k=1}^{M} \binom{M}{k} y_0[k]}{1-\mu_n} + \sum_{n=n_0+1}^{n_1} \frac{\sum_{k=1}^{M} \binom{M}{k} y_1[k]}{1-\mu_n} + \cdots \right] \tag{7.113}$$

$$= \nu \sum_{\ell \in \mathcal{S}} \sum_{n=n_{\ell-1}+1}^{n_\ell} \frac{\sum_{k=1}^{M} \binom{M}{k} y_\ell[k]}{1-\mu_n} \tag{7.114}$$

Furthermore, the total desired symbols from all databases from all repetitions is given by,

$$L(\mathbf{n}) = \nu \sum_{\ell \in \mathcal{S}} \sum_{k=1}^{M} \binom{M-1}{k-1} y_\ell[k](n_\ell - n_{\ell-1}) \tag{7.115}$$

Thus, the following rate is achievable corresponding to the sequence $\mathbf{n}$,

$$R(\mathbf{n}, \boldsymbol{\mu}) = \frac{\sum_{\ell \in \mathcal{S}} \sum_{k=1}^{M} \binom{M-1}{k-1} y_\ell[k](n_\ell - n_{\ell-1})}{\sum_{\ell \in \mathcal{S}} \sum_{n=n_{\ell-1}+1}^{n_\ell} \frac{\sum_{k=1}^{M} \binom{M}{k} y_\ell[k]}{1-\mu_n}} \tag{7.116}$$

Since this scheme is achievable for every monotone non-decreasing sequence

$\mathbf{n} = \{n_i\}_{i=0}^{M-1}$, the following rate is achievable,

$$R(\boldsymbol{\mu}) = \max_{n_0 \leq \cdots \leq n_{M-1} \in \{1,\cdots,N\}} \frac{\sum_{\ell \in \mathcal{S}} \sum_{k=1}^{M} \binom{M-1}{k-1} y_\ell[k](n_\ell - n_{\ell-1})}{\sum_{\ell \in \mathcal{S}} \sum_{n=n_{\ell-1}+1}^{n_\ell} \frac{\sum_{k=1}^{M} \binom{M}{k} y_\ell[k]}{1-\mu_n}} \tag{7.117}$$

7.5.4 Optimality for $M = 2$ and $M = 3$ Messages

In this section, we prove the optimality of our scheme for $M = 2$ and $M = 3$. The proof relies on relating the upper bound for the PIR-WTC-II problem with the upper bound for the PIR problem under asymmetric traffic constraints. From the settled optimality of the achievable scheme of the meaningful symbols for $M = 2$, $M = 3$ for the PIR problem under asymmetric traffic constraints, we conclude the optimality of our scheme for PIR-WTC-II.[1]

We return to the upper bound in Theorem 7.1,

$$\bar{C}(\boldsymbol{\mu}) = \max_{\boldsymbol{\tau} \in \mathbb{T}} \min_{n_i \in \{1,\cdots,N\}} \frac{\sum_{n=1}^{N}(1-\mu_n)\tau_n + \frac{\sum_{n=n_1+1}^{N}(1-\mu_n)\tau_n}{n_1} + \cdots + \frac{\sum_{n=n_{M-1}+1}^{N}(1-\mu_n)\tau_n}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{7.118}$$

$$= \max_{\boldsymbol{\tau} \in \mathbb{T}} \sum_{n=1}^{N}(1-\mu_n)\tau_n \cdot \min_{n_i \in \{1,\cdots,N\}} \frac{1 + \frac{\sum_{n=n_1+1}^{N}(1-\mu_n)\tau_n}{n_1 \cdot \sum_{n=1}^{N}(1-\mu_n)\tau_n} + \cdots + \frac{\sum_{n=n_{M-1}+1}^{N}(1-\mu_n)\tau_n}{\prod_{i=1}^{M-1} n_i \cdot \sum_{n=1}^{N}(1-\mu_n)\tau_n}}{1 + \frac{1}{n_1} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{7.119}$$

$$= \max_{\boldsymbol{\tau} \in \mathbb{T}} \sum_{n=1}^{N}(1-\mu_n)\tau_n \cdot \min_{n_i \in \{1,\cdots,N\}} \frac{1 + \frac{1}{n_1}\sum_{n=n_1+1}^{N}\tilde{\tau}_n + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}\sum_{n=n_{M-1}+1}^{N}\tilde{\tau}_n}{1 + \frac{1}{n_1} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{7.120}$$

[1] Alternatively, for a specified N, $\boldsymbol{\mu}$, we can prove the optimality by showing that the KKT conditions of the upper bound optimization problem are satisfied by our achievable scheme.

$$= \max_{\boldsymbol{\tau} \in \mathbb{T}} \sum_{n=1}^{N} (1-\mu_n)\tau_n \cdot \tilde{C}(\tilde{\boldsymbol{\tau}}) \tag{7.121}$$

where $\tilde{\tau}_n$ is obtained by the change of variable $\tilde{\tau}_n = \frac{(1-\mu_n)\tau_n}{\sum_{i=1}^{N}(1-\mu_i)\tau_i}$ and the inner problem $\tilde{C}(\tilde{\boldsymbol{\tau}})$ is defined as:

$$\tilde{C}(\tilde{\boldsymbol{\tau}}) = \min_{n_i \in \{1,\cdots,N\}} \frac{1 + \frac{1}{n_1}\sum_{n=n_1+1}^{N} \tilde{\tau}_n + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i} \sum_{n=n_{M-1}+1}^{N} \tilde{\tau}_n}{1 + \frac{1}{n_1} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{7.122}$$

The inner problem is precisely the upper bound for the PIR problem under asymmetric traffic constraints $\tilde{\boldsymbol{\tau}}$ in [125, Theorem 1].

In the following lemma, we show that the solution of $\bar{C}(\boldsymbol{\mu})$ exists at one of the corner points of $\tilde{C}(\tilde{\boldsymbol{\tau}})$.

Lemma 7.4 *The solution of $\bar{C}(\boldsymbol{\mu})$ exists at one of the corner points of $\tilde{C}(\tilde{\boldsymbol{\tau}})$ after the change of variables $\tau_n = \frac{\sum_{i=1}^{N}(1-\mu_i)\tau_i}{(1-\mu_n)}$.*

Proof: To show this, we note that the upper bound in Theorem 7.1 can be written as the following linear program as discussed in Remark 7.3:

$$\begin{aligned} \max_{\boldsymbol{\tau}, R} \quad & R \\ \text{s.t.} \quad & R \leq \frac{\sum_{n=1}^{N}(1-\mu_n)\tau_n + \frac{\sum_{n=n_1+1}^{N}(1-\mu_n)\tau_n}{n_1} + \cdots + \frac{\sum_{n=n_{M-1}+1}^{N}(1-\mu_n)\tau_n}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}}, \quad \forall \mathbf{n} \\ & \sum_{n=1}^{N} \tau_n = 1, \quad \tau_n \geq 0,\ n = 1, \cdots, N \end{aligned} \tag{7.123}$$

Equivalently, from (7.120), we can write the optimization problem correspond-

ing to the upper bound as:

$$\begin{aligned}
\max_{\boldsymbol{\tau}\in\mathbb{T},\tilde{R},\tilde{\boldsymbol{\tau}}} \quad & \sum_{n=1}^{N}(1-\mu_n)\tau_n\cdot\tilde{R} \\
\text{s.t.} \quad & \tilde{R}\leq\frac{1+\frac{1}{n_1}\sum_{n=n_1+1}^{N}\tilde{\tau}_n+\cdots+\frac{1}{\prod_{i=1}^{M-1}n_i}\sum_{n=n_{M-1}+1}^{N}\tilde{\tau}_n}{1+\frac{1}{n_1}+\cdots+\frac{1}{\prod_{i=1}^{M-1}n_i}}, \quad \forall\mathbf{n} \\
& \sum_{n=1}^{N}\tilde{\tau}_n=1, \quad \tilde{\tau}_n\geq 0,\ n=1,\cdots,N \\
& \tilde{\tau}_n=\frac{(1-\mu_n)\tau_n}{\sum_i(1-\mu_i)\tau_i}, \quad n=1,\cdots,N
\end{aligned} \tag{7.124}$$

We note that the constraints of this equivalent problem is the same as constraints of the upper bounds of the PIR problem under the asymmetric traffic constraints $\tilde{\boldsymbol{\tau}}$.

Since there are a finite number of constraints ($N^{M-1}+2$ constraints), the feasible region is a polyhedron, thus, the solution for $\bar{C}(\boldsymbol{\mu})$ resides at a corner point of this polyhedron.

For any corner point of this optimization problem, $(N+1)$ constraints are active (i.e., met with equality) and linearly independent.

Since these constraints take the form of

$$R=\frac{\sum_{n=1}^{N}(1-\mu_n)\tau_n+\frac{\sum_{n=n_1+1}^{N}(1-\mu_n)\tau_n}{n_1}+\cdots+\frac{\sum_{n=n_{M-1}+1}^{N}(1-\mu_n)\tau_n}{\prod_{i=1}^{M-1}n_i}}{1+\frac{1}{n_1}+\cdots+\frac{1}{\prod_{i=1}^{M-1}n_i}} \tag{7.125}$$

by dividing both sides by $\sum_{i=1}^{N}(1-\mu_i)\tau_i > 0$, the constraint become

$$\tilde{R} = \frac{R}{\sum_{i=1}^{N}(1-\mu_i)\tau_i} = \frac{1 + \frac{1}{n_1}\sum_{n=n_1+1}^{N}\tilde{\tau}_n + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}\sum_{n=n_{M-1}+1}^{N}\tilde{\tau}_n}{1 + \frac{1}{n_1} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{7.126}$$

Hence, the condition of intersection of the active constraints of the $\bar{C}(\boldsymbol{\mu})$ is the same as the condition of the intersection of the bounds of $\tilde{C}(\tilde{\boldsymbol{\tau}})$ after the change of variables. Thus, it suffices to consider the corner points of the inner problem and map the solution using the change of variables $\tau_n = \frac{\sum_{i=1}^{N}(1-\mu_i)\tau_i}{(1-\mu_n)}$. ■

Consequently, for a corner point of the inner problem $(\tilde{\boldsymbol{\tau}}^*, \tilde{C}(\tilde{\boldsymbol{\tau}}^*))$, we have the reverse change of variables

$$\tau_n^* = \tilde{\tau}_n^* \cdot \frac{\sum_{i=1}^{N}(1-\mu_i)\tau_i^*}{1-\mu_n} \tag{7.127}$$

Now, since $\sum_{n=1}^{N}\tau_n^* = 1$, $\sum_{n=1}^{N}\tilde{\tau}_n^* \cdot \frac{\sum_{i=1}^{N}(1-\mu_i)\tau_i^*}{1-\mu_n} = 1$, which leads to

$$\sum_{i=1}^{N}(1-\mu_i)\tau_i = \frac{1}{\sum_{n=1}^{N}\frac{\tilde{\tau}_n}{1-\mu_n}} \tag{7.128}$$

Denote $\bar{C}(\tilde{\boldsymbol{\tau}}^*, \boldsymbol{\mu})$ to be the upper bound of the PIR-WTC-II problem corresponding to the corner point $(\tilde{\boldsymbol{\tau}}^*, \tilde{C}(\tilde{\boldsymbol{\tau}}^*))$ of the inner problem, hence from (7.121), we have

$$\bar{C}(\tilde{\boldsymbol{\tau}}^*, \boldsymbol{\mu}) = \sum_{i=1}^{N}(1-\mu_i)\tau_i \cdot \tilde{C}(\tilde{\boldsymbol{\tau}}^*) \tag{7.129}$$

$$= \frac{\tilde{C}(\tilde{\boldsymbol{\tau}}^*)}{\sum_{n=1}^{N}\frac{\tilde{\tau}_n}{1-\mu_n}} \tag{7.130}$$

Thus, the upper bound can be written in terms of the corner points of the inner

problem $\{\tilde{\boldsymbol{\tau}}^{(i)}\}_{i=1}^{\theta}$, where θ is the total number of corner points as

$$\bar{C}(\boldsymbol{\mu}) = \max_{i\in\{1,\cdots,\theta\}} \frac{\tilde{C}(\tilde{\boldsymbol{\tau}}^{(i)})}{\sum_{n=1}^{N} \frac{\tilde{\boldsymbol{\tau}}^{(i)}}{1-\mu_n}} \tag{7.131}$$

7.5.4.1 $M = 2$ Messages

From [125], we know that for $M = 2$, all the corner points of the inner problem are in fact optimal. For an increasing sequence (n_0, n_1), the corner points are characterized by:

$$\tilde{\tau}_n = \begin{cases} \frac{n_0+1}{n_0(n_1+1)}, & 1 \le n \le n_0 \\ \frac{1}{n_1+1}, & n_0+1 \le n \le n_1 \\ 0, & n_1+1 \le n \le N \end{cases} \quad \Rightarrow \quad \tilde{C}(\tilde{\boldsymbol{\tau}}) = \frac{n_1}{n_1+1} \tag{7.132}$$

Hence, the upper bound for $M = 2$ can be explicitly written as:

$$\bar{C}(\boldsymbol{\mu}) = \max_{n_0,n_1\in\{1,\cdots,N\}} \frac{\frac{n_1}{n_1+1}}{\sum_{n=1}^{n_0} \frac{n_0+1}{n_0(n_1+1)(1-\mu_n)} + \sum_{n=n_0+1}^{n_1} \frac{1}{(n_1+1)(1-\mu_n)}} \tag{7.133}$$

$$= \max_{n_0,n_1\in\{1,\cdots,N\}} \frac{n_0 n_1}{\sum_{n=1}^{n_0} \frac{n_0+1}{1-\mu_n} + \sum_{n=n_0+1}^{n_1} \frac{n_0}{1-\mu_n}} \tag{7.134}$$

From the achievability side, for a sequence (n_0, n_1), the system of difference equations in Theorem 7.2 reduces to

$$y_0[k] = (n_0 - 1)y_0[k-1] \tag{7.135}$$

$$y_1[k] = n_0 y_0[k-1] \tag{7.136}$$

for $k = 1, 2$, where $y_0[1] = 1$, and $y_1[1] = 0$. Hence, $y_0[2] = n_0 - 1$, and $y_1[2] = n_0$. Consequently, the achievable rate in Theorem 7.2 is explicitly evaluated for $M = 2$ as:

$$R(\boldsymbol{\mu}) = \max_{n_0, n_1 \in \{1, \cdots, N\}} \frac{\sum_{\ell \in \mathcal{S}} \sum_{k=1}^{M} \binom{M-1}{k-1} y_\ell[k](n_\ell - n_{\ell-1})}{\sum_{\ell \in \mathcal{S}} \sum_{n=n_{\ell-1}+1}^{n_\ell} \frac{\sum_{k=1}^{M} \binom{M}{k} y_\ell[k]}{1-\mu_n}} \tag{7.137}$$

$$= \max_{n_0, n_1 \in \{1, \cdots, N\}} \frac{n_0 n_1}{\sum_{n=1}^{n_0} \frac{n_0+1}{1-\mu_n} + \sum_{n=n_0+1}^{n_1} \frac{n_0}{1-\mu_n}} \tag{7.138}$$

which matches the upper bound and concludes the optimality for $M = 2$.

7.5.4.2 $M = 3$ Messages

Similarly, from [125], the corner points of the inner problem occur for an increasing sequence (n_0, n_1, n_2). The corner points are characterized by:

$$\tilde{\tau}_n = \begin{cases} \frac{n_0 n_1 + n_0 + 1}{n_0(n_2 n_1 + n_1 + 1)}, & 1 \leq n \leq n_0 \\ \frac{n_1 + 1}{n_2 n_1 + n_1 + 1}, & n_0 + 1 \leq n \leq n_1 \\ \frac{n_1}{n_2 n_1 + n_1 + 1} & n_1 + 1 \leq n \leq n_2 \\ 0, & n_2 + 1 \leq n \leq N \end{cases} \quad \Rightarrow \quad \tilde{C}(\tilde{\boldsymbol{\tau}}) = \frac{n_1 n_2}{n_1 n_2 + n_1 + 1} \tag{7.139}$$

Hence, the upper bound in (7.131) is explicitly written as:

$$\bar{C}(\boldsymbol{\mu}) = \max_{n_0, n_1, n_2 \in \{1, \cdots, N\}} \frac{n_0 n_1 n_2}{\sum_{n=1}^{n_0} \frac{n_0 n_1 + n_0 + 1}{1-\mu_n} + \sum_{n=n_0+1}^{n_1} \frac{n_0 n_1 + n_0}{1-\mu_n} + \sum_{n=n_1+1}^{n_2} \frac{n_0 n_1}{1-\mu_n}} \tag{7.140}$$

From the achievability side, we have the following system of difference equa-

tions for $k = 1, 2, 3$:

$$y_0[k] = (n_0 - 1)y_0[k-1] + (n_1 - n_0)y_1[k-1] + (n_2 - n_1)y_2[k-1] \tag{7.141}$$

$$y_1[k] = n_0 y_0[k-1] + (n_1 - n_0 - 1)y_1[k-1] + (n_2 - n_1)y_2[k-1] \tag{7.142}$$

$$y_2[k] = n_0\delta[k-3] + n_0 y_0[k-1] + (n_1 - n_0)y_1[k-1] + (n_2 - n_1 - 1)y_2[k-1] \tag{7.143}$$

with the initial conditions $y_0[1] = 1$, $y_1[1] = 0$, and $y_2[1] = y_2[2] = 0$. Evaluating $y_\ell[k]$, for $\ell = 0, 1, 2$, and $k = 1, 2, 3$ recursively leads to $y_0[2] = n_0 - 1$, $y_1[2] = n_0$, $y_0[3] = n_1 n_0 - 2n_0 + 1$, $y_1[3] = n_1 n_0 - 2n_0$, and $y_2[3] = n_1 n_0$. Consequently, the achievable rate from Theorem 7.2 is explicitly expressed as:

$$R(\boldsymbol{\mu}) = \max_{n_0, n_1 \in \{1, \cdots, N\}} \frac{\sum_{\ell \in \mathcal{S}} \sum_{k=1}^{M} \binom{M-1}{k-1} y_\ell[k](n_\ell - n_{\ell-1})}{\sum_{\ell \in \mathcal{S}} \sum_{n=n_{\ell-1}+1}^{n_\ell} \frac{\sum_{k=1}^{M} \binom{M}{k} y_\ell[k]}{1-\mu_n}} \tag{7.144}$$

$$= \max_{n_0, n_1, n_2 \in \{1, \cdots, N\}} \frac{n_0 n_1 n_2}{\sum_{n=1}^{n_0} \frac{n_0 n_1 + n_0 + 1}{1-\mu_n} + \sum_{n=n_0+1}^{n_1} \frac{n_0 n_1 + n_0}{1-\mu_n} + \sum_{n=n_1+1}^{n_2} \frac{n_0 n_1}{1-\mu_n}} \tag{7.145}$$

which matches the upper bound and concludes the optimality for $M = 3$.

Remark 7.10 *We note that the meaningful portion of the answer strings follows the combinatorial water-filling shown in [125] for $M = 2$ and $M = 3$. This means that the less threatened (more secure) databases are returning more meaningful symbols than the less secure ones, hence, $\tilde{\tau}_n \geq \tilde{\tau}_k$, if $n < k$. However, the length of the entire answer string including the artificial noise symbols may not follow the same structure, e.g., in the example in Section 7.5.1.2, we see that $t_1 = 16$ and $t_2 = 18$,*

i.e., $\tau_2 > \tau_1$, *while* $\tilde{\tau}_2 < \tilde{\tau}_1$.

7.5.5 Achievable Rate for $N = 2$ and Arbitrary M

Following the analysis of this case in [125], let $s_2 \in \{0, \cdots, M-1\}$ be the number of side information symbols that are used simultaneously in the initial round download in the second database.

Hence, the user starts with downloading $\binom{M-2}{s_2-1}$ stages of individual symbols (i.e., the user downloads $M\binom{M-2}{s_2-1}$ symbols from round 1 from all messages) from the first database to create 1 stage of side information in the (s_2+1)th round. After the initial exploitation of side information, the two databases exchange side information. More specifically, from database 1 in the (s_2+2k)th round, where $k = 1, \cdots, \lfloor \frac{M-s_2}{2} \rfloor$, the user exploits the side information generated in database 2 in the (s_2+2k-1)th round to download $\binom{M-1}{s_2+2k-1}$ desired symbols from total download in the (s_2+2k)th round of $\binom{M}{s_2+2k}$. Similarly from database 2, in the (s_2+2k+1)th round, where $k = 0, \cdots, \lfloor \frac{M-s_2-1}{2} \rfloor$, the user exploits the side information generated in database 1 in the (s_2+2k)th round, and downloads $\binom{M-1}{s_2+2k}$ desired symbols from total of $\binom{M}{s_2+2k+1}$ downloads in the (s_2+2k+1)th round. Thus, using the calculation in [125], we have

$$D_1(s_2) = M\binom{M-2}{s_2-1} + \sum_{k=1}^{\lfloor \frac{M-s_2}{2} \rfloor} \binom{M}{s_2+2k} \tag{7.146}$$

$$D_2(s_2) = \sum_{k=0}^{\lfloor \frac{M-s_2-1}{2} \rfloor} \binom{M}{s_2+2k+1} \tag{7.147}$$

where $D_n(s_2)$ corresponds to the length of the meaningful downloads within the nth database from one repetition of the scheme, therefore, the total download of the scheme is given by:

$$t_1(s_2) + t_2(s_1) = \frac{D_1(s_2)}{1-\mu_1} + \frac{D_2(s_2)}{1-\mu_2} \tag{7.148}$$

$$= \frac{1}{1-\mu_1}\left[M\binom{M-2}{s_2-1} + \sum_{k=1}^{\lfloor \frac{M-s_2}{2} \rfloor}\binom{M}{s_2+2k}\right] + \frac{1}{1-\mu_2}\left[\sum_{k=0}^{\lfloor \frac{M-s_2-1}{2} \rfloor}\binom{M}{s_2+2k+1}\right] \tag{7.149}$$

The message length does not change due to the security constraint, hence, directly from [125], we have

$$L(s_2) = \binom{M-2}{s_2-1} + \sum_{k=0}^{M-s_2-1}\binom{M-1}{s_2+k} \tag{7.150}$$

Consequently, the achievable rate is explicitly given as:

$$R(\boldsymbol{\mu}) = \max_{s_2\in\{0,\cdots,M-1\}} \frac{\binom{M-2}{s_2-1} + \sum_{k=0}^{M-s_2-1}\binom{M-1}{s_2+k}}{\frac{1}{1-\mu_1}\left[M\binom{M-2}{s_2-1}+\sum_{k=1}^{\lfloor \frac{M-s_2}{2} \rfloor}\binom{M}{s_2+2k}\right]+\frac{1}{1-\mu_2}\left[\sum_{k=0}^{\lfloor \frac{M-s_2-1}{2} \rfloor}\binom{M}{s_2+2k+1}\right]} \tag{7.151}$$

including the corner point corresponding to the trivial rate, i.e., when the user deactivates the retrieval process from the second database, leading to (7.25).

7.5.6 Further Examples

In this section, we present further examples to clarify the achievable scheme for additional tractable values of M, N.

7.5.6.1 $M = 4$ Messages, $N = 2$ Databases

In this example, we show the achievable scheme for $M = 4$, $N = 2$, and arbitrary $\boldsymbol{\mu}$. This example helps us to show that our achievable scheme does not achieve the capacity for all $\boldsymbol{\mu}$. For $M = 4$, we have $M + 1 = 5$ possible achievable schemes, corresponding to $s_2 = \{0, 1, \cdots, 3\}$ and one other achievable scheme corresponding to the trivial scheme of downloading the contents of database 1. Let a_i, b_i, c_i, d_i denote the randomly permuted symbols from W_1, W_2, W_3, W_4, respectively. In all achievable schemes, the nth database generates a key K_n with length $\mu_n t_n$ and encodes it to generate an artificial noise vector $u^{(n)}_{[1:t_n]}$ using a $(t_n, \mu_n t_n)$ MDS code. The nth database provides $\mu_n t_n$ individual symbols of artificial noise. In all cases, the scheme is repeated ν times such that:

$$t_n(\mathbf{n}, \boldsymbol{\mu}) = \frac{\nu D_n(\mathbf{n})}{1 - \mu_n} \in \mathbb{N}, \quad \forall n \in \{1, 2\} \tag{7.152}$$

Now, we focus on one repetition of the achievable scheme. We further concentrate on the meaningful queries, i.e., before adding the artificial noise vector.

The trivial scheme corresponding to $\mathbf{n} = (1, 1, 1, 1)$: In one repetition of the scheme, the user downloads a_1, b_1, c_1, d_1 from database 1. Hence, $D_1(\mathbf{n}) = 4$. Consequently,

$t_1(\mathbf{n}, \boldsymbol{\mu}) = \frac{4\nu}{1-\mu_1}$. As the user decodes 1 symbol from W_1 in each repetition, $L_1(\mathbf{n}) = \nu$. Hence, $R(\mathbf{n}, \boldsymbol{\mu}) = \frac{1-\mu_1}{4}$ is achievable.

The scheme corresponding to $\mathbf{n} = (1, 1, 1, 2)$: In this case, $s_2 = 3$, i.e., the user exploits 3 side-information symbols simultaneously in database 2, i.e., focusing on one repetition of the scheme, from database 1, the user downloads a_1, b_1, c_1, d_1. The user combines $b_1 + c_1 + d_1$ and uses this side information to get a_2 from database 2, i.e., the user downloads $a_2+b_1+c_1+d_1$. Hence, $D_1(\mathbf{n}) = 4$, $D_2(\mathbf{n}) = 1$. Consequently, $t_1(\mathbf{n}, \boldsymbol{\mu}) = \frac{4\nu}{1-\mu_1}$, and $t_2(\mathbf{n}, \boldsymbol{\mu}) = \frac{\nu}{1-\mu_2}$. As the user decodes 2 symbols from W_1 in each repetition, $L_1(\mathbf{n}) = 2\nu$. Hence, $R(\mathbf{n}, \boldsymbol{\mu}) = \frac{2}{\frac{4}{1-\mu_1}+\frac{1}{1-\mu_2}}$ is achievable. The query table of the meaningful queries (without the artificial noise) for one repetition of the scheme is shown in Table 7.3.

Table 7.3: Meaningful queries for $M = 4$, $N = 2$, $s_2 = 3$.

Database 1	Database 2
a_1, b_1, c_1, d_1	$a_2 + b_1 + c_1 + d_1$

The scheme corresponding to $\mathbf{n} = (1, 1, 2, 2)$: In this case $s_2 = 2$, hence the user combines every 2 undesired symbols from database 1 to form one side information symbol. To that end, the user downloads $\binom{M-2}{s_2-1} = 2$ stages of individual symbols (1-sum) from database 1, so that the user forms 2-sums that can be used in database 2 as side information to start round 3 directly. More specifically, the user downloads $a_3 + b_1 + c_1$, $a_4 + b_2 + d_1$, $a_5 + c_2 + d_2$ from database 2 taking into considerations that all these undesired symbols are decodable from database 1. The user completes the stage by downloading $b_3 + c_3 + d_3$ that can be further exploited in database 1

by downloading $a_6 + b_3 + c_3 + d_3$. Hence, $D_1(\mathbf{n}) = 9$, $D_2(\mathbf{n}) = 4$. Consequently, $t_1(\mathbf{n}, \boldsymbol{\mu}) = \frac{9\nu}{1-\mu_1}$ and $t_2(\mathbf{n}, \boldsymbol{\mu}) = \frac{4\nu}{1-\mu_2}$. As the user decodes 6 symbols from W_1 in each repetition, $L(\mathbf{n}) = 6\nu$. Hence, $R(\mathbf{n}, \boldsymbol{\mu}) = \frac{6}{\frac{9}{1-\mu_1}+\frac{4}{1-\mu_2}}$ is achievable. The query table of the meaningful queries (without the artificial noise) for one repetition of the scheme is shown in Table 7.4.

Table 7.4: Meaningful queries for $M = 4$, $N = 2$, $s_2 = 2$.

Database 1	Database 2
a_1, b_1, c_1, d_1 a_2, b_2, c_2, d_2	$a_3 + b_1 + c_1$ $a_4 + b_2 + d_1$ $a_5 + c_2 + d_2$ $b_3 + c_3 + d_3$
$a_6 + b_3 + c_3 + d_3$	

The scheme corresponding to $\mathbf{n} = (1, 2, 2, 2)$: In this case $s_2 = 1$, hence the user exploits the individual undesired symbols downloaded from database 1 directly as a side information in database 2. To that end, the user exploits the side information generated in round 1 by downloading a_2+b_1, a_3+c_1, and a_4+d_1. The user completes the stage by downloading undesired symbols consisting of 2-sums that do not include a_i, hence the user downloads $b_2 + c_2$, $b_3 + d_2$, $c_3 + d_3$. The undesired symbols are exploited in database 1, thus the user downloads $a_5 + b_2 + c_2$, $a_6 + b_3 + d_2$, and $a_7 + c_3 + d_3$. The user completes the stage by downloading $b_4 + c_4 + d_4$, which can be exploited in database 2 by downloading $a_8 + b_4 + c_4 + d_4$. Hence, $D_1(\mathbf{n}) = 8$, $D_2(\mathbf{n}) = 7$. Consequently, $t_1(\mathbf{n}, \boldsymbol{\mu}) = \frac{8\nu}{1-\mu_1}$, and $t_2(\mathbf{n}, \boldsymbol{\mu}) = \frac{7\nu}{1-\mu_2}$. As the user decodes 8 symbols from W_1 in each repetition, $L(\mathbf{n}) = 8\nu$. Hence, $R(\mathbf{n}, \boldsymbol{\mu}) = \frac{8}{\frac{8}{1-\mu_1}+\frac{7}{1-\mu_2}}$ is achievable. The query table of the meaningful queries (without the artificial noise)

for one repetition of the scheme is shown in Table 7.5.

Table 7.5: The query table for $M = 4$, $N = 2$, $s_2 = 1$.

Database 1	Database 2
a_1, b_1, c_1, d_1	$a_2 + b_1$ $a_3 + c_1$ $a_4 + d_1$ $b_2 + c_2$ $b_3 + d_2$ $c_3 + d_3$
$a_5 + b_2 + c_2$ $a_6 + b_3 + d_2$ $a_7 + c_3 + d_3$ $b_4 + c_4 + d_4$	$a_8 + b_4 + c_4 + d_4$

As in the case of $M = 3$, under the assumption that $\mu_1 \leq \mu_2$, the symmetric scheme in [12] does not achieve any larger retrieval rates at any $\boldsymbol{\mu}$. Hence, the following rate is achievable,

$$R(\boldsymbol{\mu}) = \max\left\{\frac{1-\mu_1}{4},\ \frac{2}{\frac{4}{1-\mu_1}+\frac{1}{1-\mu_2}},\ \frac{6}{\frac{9}{1-\mu_1}+\frac{4}{1-\mu_2}},\ \frac{8}{\frac{8}{1-\mu_1}+\frac{7}{1-\mu_2}}\right\} \tag{7.153}$$

In Fig. 7.6, we illustrate the partitioning of the $\boldsymbol{\mu}$ space in terms of the active achievable scheme. In Fig. 7.7, we plot the gap versus $\boldsymbol{\mu}$ for $M = 4$, $N = 2$. We note that the gap is upper bounded by 0.0051 and this gap exists only for specific regimes of $\boldsymbol{\mu}$.

7.5.6.2 $M = 2$ Messages, $N = 3$ Databases

In this example, we show the achievable scheme for $M = 2$, $N = 3$, and arbitrary $\boldsymbol{\mu}$. Again we focus on the meaningful queries in our exposition to avoid repetition. The artificial noise incorporation is exactly as in the previous examples. Let a_i, b_i

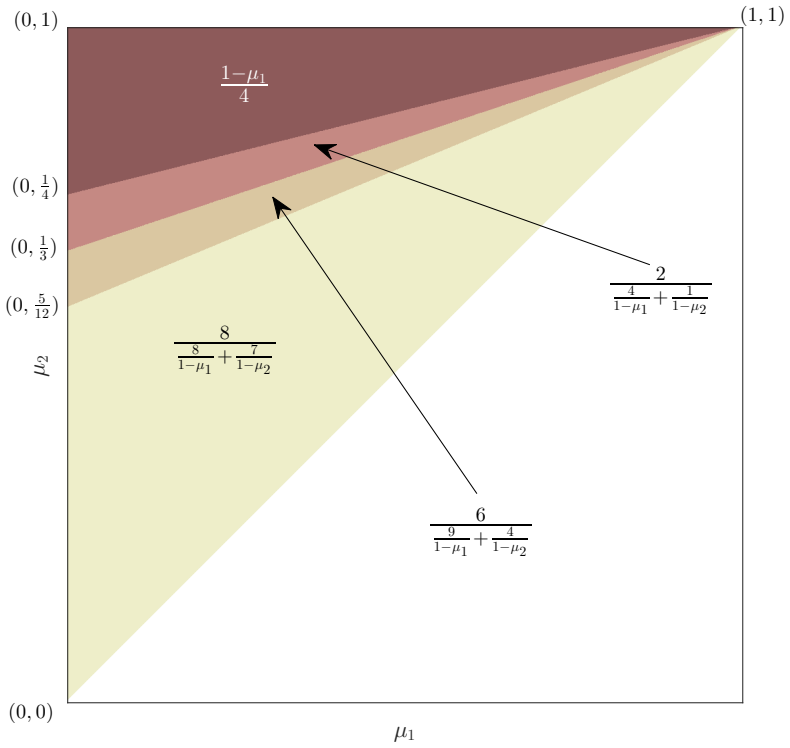


Figure 7.6: Partitions of $\boldsymbol{\mu}$ space according to retrieval rate expression for $M = 4$, $N = 2$.

denote the randomly permuted symbols from W_1, W_2, respectively.

The trivial scheme corresponding to $(n_0, n_1) = (1, 1)$: In this case, the user deactivates the retrieval from database 2. Hence, in one repetition, the user downloads a_1, b_1 from database 1 only. Therefore, $D_1(1, 1) = 2$ which leads to $t_1(1, 1, \boldsymbol{\mu}) = \frac{2\nu}{1-\mu_1}$. From one repetition of the scheme, the user decodes 1 symbol from W_1, hence $L = \nu$ symbols. This gives the rate $R(1, 1, \boldsymbol{\mu}) = \frac{1-\mu_1}{2}$.

The scheme corresponding to $(n_0, n_1) = (1, 2)$: In this case, the user exploits the undesired symbols in database 1 as a side information in database 2 only and deactivates database 3. Hence, in one repetition, the user downloads a_1, b_1 from database

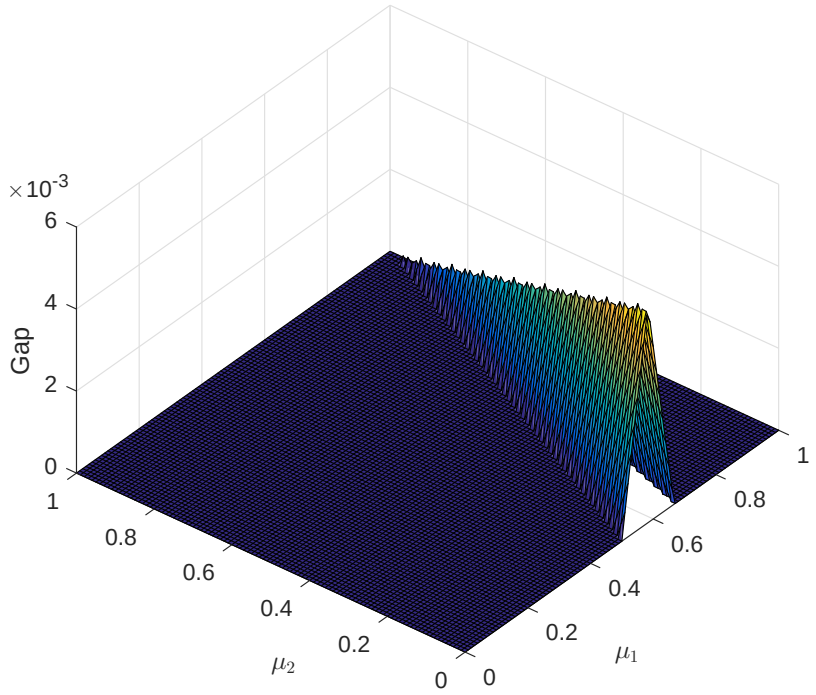


Figure 7.7: Capacity gap for the case of $M = 4$, $N = 2$.

1, and uses b_1 as side information in database 2 by downloading $a_2 + b_1$. Therefore, $D_1(1,2) = 2$, $D_2(1,2) = 1$ which leads to $t_1(1,2,\boldsymbol{\mu}) = \frac{2\nu}{1-\mu_1}$, and $t_2(1,2,\boldsymbol{\mu}) = \frac{\nu}{1-\mu_2}$. From one repetition of the scheme, the user decodes 2 symbols from W_1, hence $L = 2\nu$ symbols. This gives the rate $R(1,2,\boldsymbol{\mu}) = \frac{2}{\frac{2}{1-\mu_1}+\frac{1}{1-\mu_2}}$. The query table of the meaningful queries (without the artificial noise) for one repetition of the scheme is shown in Table 7.6.

Table 7.6: Meaningful queries for $M = 2$, $N = 3$, $\mathbf{n} = (1,2)$.

Database 1	Database 2	Database 3
a_1, b_1	$a_2 + b_1$	

The scheme corresponding to $(n_0, n_1) = (1,3)$: Since $n_1 = 3$, the user exploits the side information in database 2 and database 3. Hence, in one repetition, the user

downloads a_1, b_1 from database 1. The user downloads $a_2 + b_1$ from database 2, and $a_3 + b_1$ from database 3. Therefore, $D_1(1,3) = 2$, $D_2(1,3) = 1$, $D_3(1,3) = 1$ which leads to $t_1(1,3,\boldsymbol{\mu}) = \frac{2\nu}{1-\mu_1}$, $t_2(1,3,\boldsymbol{\mu}) = \frac{\nu}{1-\mu_2}$, $t_3(1,3,\boldsymbol{\mu}) = \frac{\nu}{1-\mu_3}$. From one repetition of the scheme, the user decodes 3 symbols from W_1, hence $L = 3\nu$ symbols. This corresponds to the rate $R(1,3,\boldsymbol{\mu}) = \frac{3}{\frac{2}{1-\mu_1}+\frac{1}{1-\mu_2}+\frac{1}{1-\mu_3}}$. The query table of the meaningful queries (without the artificial noise) for one repetition of the scheme is shown in Table 7.7.

Table 7.7: Meaningful queries for $M = 2$, $N = 3$, $\mathbf{n} = (1,3)$.

Database 1	Database 2	Database 3
a_1, b_1	$a_2 + b_1$	$a_3 + b_1$

The scheme corresponding to $(n_0, n_1) = (2,2)$: In this case, the user applies the symmetric scheme at databases 1 and 2, and deactivates database 3. Consequently, the user downloads a_1, b_1 from database 1. From database 2, the user downloads new symbols a_2, b_2. The user exploits the side information generated in the first round of download by downloading $a_3 + b_2$, and $a_4 + b_1$. Therefore, $D_1(2,2) = 3$, $D_2(2,2) = 3$ which leads to $t_1(2,2,\boldsymbol{\mu}) = \frac{3\nu}{1-\mu_1}$, $t_2(2,2,\boldsymbol{\mu}) = \frac{3\nu}{1-\mu_2}$. From one repetition of the scheme, the user decodes 4 symbols from W_1, hence $L = 4\nu$ symbols. This gives the rate $R(2,2,\boldsymbol{\mu}) = \frac{4}{\frac{3}{1-\mu_1}+\frac{3}{1-\mu_2}}$. The query table of the meaningful queries (without the artificial noise) for one repetition of the scheme is shown in Table 7.8.

Table 7.8: Meaningful queries for $M = 2$, $N = 3$, $\mathbf{n} = (2,2)$.

Database 1	Database 2	Database 3
a_1, b_1	a_2, b_2	
$a_3 + b_2$	$a_4 + b_1$	

The scheme corresponding to $(n_0, n_1) = (2, 3)$: In this case, the user further exploits the side information generated in databases 1 and 2 in database 3. Hence, the user downloads $a_3 + b_1$, $a_4 + b_2$ from database 3. Therefore, $D_1(2, 3) = 3$, $D_2(2, 3) = 3$, $D_3(2, 3) = 2$ which leads to $t_1(2, 3, \boldsymbol{\mu}) = \frac{3\nu}{1-\mu_1}$, $t_2(2, 3, \boldsymbol{\mu}) = \frac{3\nu}{1-\mu_2}$, $t_3(2, 3, \boldsymbol{\mu}) = \frac{2\nu}{1-\mu_3}$. From one repetition of the scheme, the user decodes 6 symbols from W_1, hence $L = 6\nu$ symbols. This gives the rate $R(2, 3, \boldsymbol{\mu}) = \frac{6}{\frac{3}{1-\mu_1}+\frac{3}{1-\mu_2}+\frac{2}{1-\mu_3}}$. The query table of the meaningful queries (without the artificial noise) for one repetition of the scheme is shown in Table 7.9.

Table 7.9: Meaningful queries for $M = 2$, $N = 3$, $\mathbf{n} = (2, 3)$.

Database 1	Database 2	Database 3
a_1, b_1	a_2, b_2	$a_3 + b_1$ $a_4 + b_2$
$a_5 + b_2$	$a_6 + b_1$	

The scheme corresponding to $(n_0, n_1) = (3, 3)$: In this case, the user applies the symmetric scheme in [12]. Therefore, $D_n(3, 3) = 4$, where $n = 1, 2, 3$ which leads to $t_n(3, 3, \boldsymbol{\mu}) = \frac{4\nu}{1-\mu_n}$. From one repetition of the scheme, the user decodes 9 symbols from W_1, hence $L = 9\nu$ symbols. This gives the rate $R(3, 3, \boldsymbol{\mu}) = \frac{9}{\frac{4}{1-\mu_1}+\frac{4}{1-\mu_2}+\frac{4}{1-\mu_3}}$. The query table of the meaningful queries (without the artificial noise) for one repetition of the scheme is shown in Table 7.10.

Table 7.10: Meaningful queries for $M = 2$, $N = 3$, $\mathbf{n} = (3, 3)$.

Database 1	Database 2	Database 3
a_1, b_1	a_2, b_2	a_3, b_3
$a_4 + b_2$ $a_5 + b_3$	$a_6 + b_1$ $a_7 + b_3$	$a_8 + b_1$ $a_9 + b_2$

Consequently, the following rate is achievable:

$$R(\boldsymbol{\mu}) = \max\left\{ \frac{1-\mu_1}{2},\ \frac{2}{\frac{2}{1-\mu_1}+\frac{1}{1-\mu_2}},\ \frac{3}{\frac{2}{1-\mu_1}+\frac{1}{1-\mu_2}+\frac{1}{1-\mu_3}},\right.$$
$$\left.\frac{4}{\frac{3}{1-\mu_1}+\frac{3}{1-\mu_2}},\ \frac{6}{\frac{3}{1-\mu_1}+\frac{3}{1-\mu_2}+\frac{2}{1-\mu_3}},\ \frac{9}{\frac{4}{1-\mu_1}+\frac{4}{1-\mu_2}+\frac{4}{1-\mu_3}}\right\} \quad (7.154)$$

7.6 Conclusions

In this chapter, we investigated the PIR-WTC-II problem. We showed that the problem is a concrete example of the PIR problem under asymmetric traffic constraints. We obtained a general upper bound that extends the converse techniques in Chapter 5. The converse proof takes the form of a max-min optimization problem. The inner minimization problem derives the tightest upper bound for the retrieval rate for an arbitrary traffic ratio vector $\boldsymbol{\tau}$, while the outer maximization problem optimizes over $\boldsymbol{\tau}$. The core of the achievability proof is the achievability proof of the corner points of the PIR problem under asymmetric traffic constraints. The security constraint is satisfied by encrypting each returned answering string by an artificial noise vector. To generate the artificial noise vector, the nth database generates a secret key and encodes it into artificial noise by a $(t_n, \mu_n t_n)$ MDS code. The upper and lower bounds match for $M = 2$ and $M = 3$, for any N, and for every eavesdropping capability vector $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)$.

CHAPTER 8

Secure Degrees of Freedom Region of Gaussian MIMO Interference Channel

8.1 Introduction

In this chapter, we consider the two-user multiple-input multiple-output (MIMO) interference channel with confidential messages (ICCM). We determine the exact secure degrees of freedom (s.d.o.f.) region for the symmetric case of M antennas at both transmitters and N antennas at both receivers. We develop the converse by combining the broadcast channel with confidential messages (BCCM) cooperative upper bound, decodability upper bound for the interference channel with no secrecy constraints, and vector extensions of the secrecy penalty and role of a helper lemmas. For the achievability, we first show that the s.d.o.f. region is a four-vertex polytope. For the sum s.d.o.f. point, we propose a novel achievable scheme for the 2×2 ICCM, which combines asymptotic real interference alignment with spatial interference alignment. Using this scheme, we provide achievable schemes for any M and N by proper vector space operations. We achieve the other non-trivial extreme polytope points by employing one of the transmitters as a deaf helper for assisting

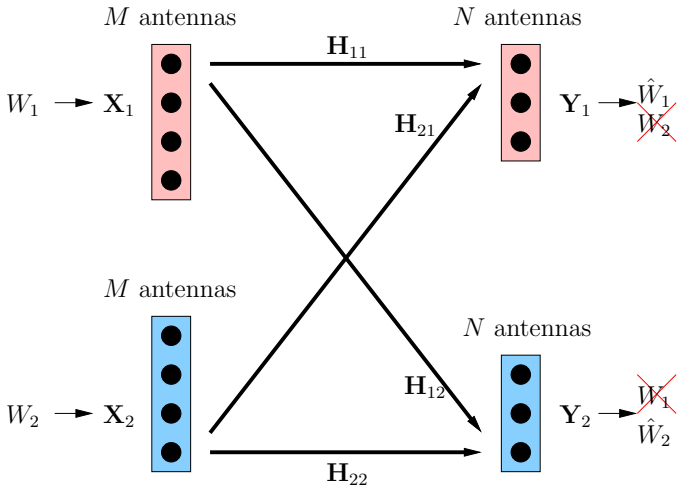


Figure 8.1: Two-user MIMO ICCM.

the secure transmission of the other user. We present simplified achievable schemes for the special case of time-varying MIMO ICCM. The achievable schemes, in this case, make use of the time-varying nature of the channel to construct vector-space alignment counterpart of the real interference alignment used in the static channel case.

8.2 System Model

We consider a two-user symmetric Gaussian MIMO ICCM. Each transmitter has M transmit antennas, and each receiver has N receive antennas. The input-output relationships of a two-user MIMO ICCM (see Fig. 8.1) are:

$$\mathbf{Y}_1(t) = \mathbf{H}_{11}(t)\mathbf{X}_1(t) + \mathbf{H}_{21}(t)\mathbf{X}_2(t) + \mathbf{N}_1(t) \tag{8.1}$$

$$\mathbf{Y}_2(t) = \mathbf{H}_{12}(t)\mathbf{X}_1(t) + \mathbf{H}_{22}(t)\mathbf{X}_2(t) + \mathbf{N}_2(t) \tag{8.2}$$

where $\mathbf{H}_{ij}(t) \in \mathbb{R}^{N\times M}$ is the channel gain matrix from transmitter i to receiver

j (where $i, j \in \{1, 2\}$) at channel use t. We call the ICCM *static*, if $\mathbf{H}_{ij}(t) = \mathbf{H}_{ij}$ for all channel uses t, $\forall i, j$. The ICCM is *time-varying*, if $\mathbf{H}_{ij}(t)$ takes an independent realization at every channel use t, $\forall i, j$. We assume that $\mathbf{H}_{ij}(t)$ is picked from a continuous distribution. Hence, $\mathbf{H}_{ij}(t)$ admits rationally independent elements with probability 1. Furthermore, any finite collection of the channel gains are linearly independent with probability[1] 1. $\mathbf{X}_i(t) \in \mathbb{R}^M$ is the channel input of transmitter i at channel use t, $\mathbf{Y}_i(t) \in \mathbb{R}^N$ is the channel output of receiver i at channel use t, and $\mathbf{N}_i(t) \in \mathbb{R}^N$ is i.i.d. Gaussian noise vector with a finite variance at receiver i.

Transmitter $i \in \{1, 2\}$ sends a message W_i chosen uniformly from a message set $\mathcal{W}_i$ by encoding it into an n-letter channel input $\mathbf{X}_i^n(t)$. The message W_i is to be conveyed reliably to receiver i and to be kept secret from receiver j, where $j \neq i$. Transmitter i performs stochastic encoding f_i over n channel uses $f_i : \mathcal{W}_i \to \mathbf{X}_i^n(t)$ such that for any $\epsilon > 0$, the following reliability and security constraints are satisfied:

$$\mathbb{P}(\hat{W}_1 \neq W_1) \leq \epsilon, \quad \frac{1}{n} I(W_1; \mathbf{Y}_2^n) \leq \epsilon \tag{8.3}$$

$$\mathbb{P}(\hat{W}_2 \neq W_2) \leq \epsilon, \quad \frac{1}{n} I(W_2; \mathbf{Y}_1^n) \leq \epsilon \tag{8.4}$$

where $\hat{W}_i$ is the estimate of W_i at receiver i. The channel inputs are subject to average power constraints $\mathrm{tr}(\mathbb{E}[\mathbf{X}_i(t)\mathbf{X}_i(t)^T]) \leq P$, $i = 1, 2$. The rate of user i is $R_i = \frac{1}{n} \log |\mathcal{W}_i|$. The s.d.o.f. d_i of user i is:

$$d_i = \lim_{P \to \infty} \frac{R_i}{\frac{1}{2} \log P} \tag{8.5}$$

[1] In the exposition of the results, the phrase "for almost all" refers to the rational/linear independence, which occurs with probability 1.

The sum s.d.o.f. d_s is given by $d_s = d_1 + d_2$.

8.3 Preliminaries

In this section, we review the real interference alignment and spatial alignment techniques as they are the main ingredients of our achievable scheme. In this work, we combine both techniques for MIMO ICCM with static channels.

8.3.1 Real Interference Alignment

The real interference alignment technique, which is introduced in [101] and employed in [78] for achieving the s.d.o.f. for one-hop networks, relies on transmitting multiple data streams of PAM signals. Specifically, let $\{b_i\}_{i=1}^{L}$ be a sequence of L symbols. The symbol b_i is picked from PAM constellation $C(a, Q)$, where a is the separation between any two symbols in the constellation set and the number of symbols in the constellation set is given by $2Q + 1$, i.e., $C(a, Q) = a\{-Q, -Q + 1, \cdots, Q - 1, Q\}$. Now, consider transmitting these L symbols simultaneously in the form of a linear combination,

$$x = \sum_{i=1}^{L} \alpha_i b_i \tag{8.6}$$

where $\{\alpha_i : i = 1, \cdots, L\}$ are *rationally independent* real numbers. The rational independence means that if $\sum_{i=1}^{L} \alpha_i q_i = 0$ for some $q_1, \cdots, q_L$ which are rational numbers, then $q_i = 0$ for all i.

Although the signal x is a mixture of $\{b_i\}_{i=1}^{L}$, these symbols lie in separate

rational dimensions if we choose,

$$Q = P^{\frac{1-\delta}{2(L+\delta)}}, \quad a = \gamma \frac{P^{\frac{1}{2}}}{Q} \tag{8.7}$$

for some $\delta > 0$, a positive constant γ which is independent of P that is chosen to satisfy the power constraint [101]. In this case, the constellation observed at the receiver side consists of $(2Q+1)^L$ points and the probability of error can be upper bounded $P_e \leq \exp\left(-\eta_\gamma P^\delta\right)$. To summarize: by careful choice of $C(a, Q)$ (e.g., by choosing the number of points as a function of L as in (8.7)), one can send L separable data streams that satisfy the average power and the reliability constraints. This is done by creating and exploiting *rational dimensions*.

This technique can be effectively used for security as in [78]. To achieve this, the transmitted signals in general consist of two components, namely, the secure signal V, and the cooperative jamming signal U. The cooperative jamming component U_i from transmitter i is utilized to satisfy the security constraint of transmitter j by being aligned with V_j in the same rational dimension. This can be done by scaling both U_i and V_j by real coefficients such that their scaling is the same at the receiver after passing through the channel. More specifically, the ith transmitter sends $X_i = \alpha_i U_i$ and the jth transmitter sends $X_j = \alpha_j V_j$, such that $\alpha_i h_{ij} = \alpha_j h_{jj}$, where h_{ij} is the channel gain from transmitter i to receiver j and h_{jj} is the channel gain from transmitter j to receiver j. This satisfies the security at the jth receiver as the received signal will have a component $\alpha_j h_{jj}(U_i + V_j)$, i.e., the secure signal and the cooperative jamming signal lie in the same rational dimension and hence

the leakage is upper bounded by a constant.

8.3.2 Spatial Alignment

The spatial alignment technique, introduced in [104], can be used for security as well if the system is equipped by multiple antennas. Spatial alignment does not require a specific signaling scheme, i.e., it does not require transmitting PAM signals as in the real interference alignment scheme, instead Gaussian signaling can be used. The spatial alignment exploits the *spatial dimensions* offered by the multiple antennas in contrast to the *rational dimensions* in the real interference alignment scheme.

To achieve this, the ith transmitter transmits precoded version of the cooperative jamming signal $\mathbf{U}_i$ by transmitting $\mathbf{X}_i = \mathbf{Q}_i\mathbf{U}_i$, where $\mathbf{Q}_i$ is a precoding matrix for the cooperative jamming components from the ith transmitter. Furthermore, the jth transmitter sends $\mathbf{X}_j = \mathbf{P}_j\mathbf{V}_j$, where $\mathbf{P}_j$ is the precoding matrix for the secure signal component from the jth transmitter. This is achievable since both transmitters are equipped by multiple transmit antennas. By ensuring that $\mathbf{Q}_i\mathbf{H}_{ij} = \mathbf{P}_j\mathbf{H}_{jj}$, both signal components are aligned in the same *spatial dimension* at the jth receiver, i.e., the received signal has a component $\mathbf{P}_j\mathbf{H}_{jj}(\mathbf{U}_i + \mathbf{V}_j)$. This satisfies the security constraint as well.

Note that in order to ensure reliable decoding at the receiver by a zero forcing decoder, the total number of spatial dimensions spanned by the signal components must be at most N (the number of receive antennas). This is parallel to choosing Q in the real interference alignment scheme. Furthermore, this precoding idea can

be extended as in [104] for time-varying SISO channels by *symbol extension*, i.e., completing the transmission over multiple time slots and dealing with the transmitted symbols across time as a spatial vector. In this case, the alignment technique exploits the *time dimension*.

8.3.3 Comparison of the Two Alignment Techniques

We note that the main strength of the real interference alignment technique is that it creates a potential of performing interference alignment even for SISO channels which do not enjoy time-varying diversity. This technique requires rational independence of the channel coefficients. However, the decoding procedure of this scheme is generally more complex than spatial alignment that uses simple zero-forcing decoder.

On the other hand, the spatial alignment technique requires either the presence of multiple antennas and/or time-varying channels. This hinders the usage of spatial alignment for static channels despite its simplicity.

8.4 Main Results and Discussions

The first result of this chapter characterizes the sum s.d.o.f. d_s of the two-user $M \times N$ MIMO ICCM for arbitrary M and N.

Theorem 8.1 *The sum s.d.o.f. of the two user $M \times N$ MIMO ICCM is given by,*

$$d_s = \begin{cases} \min\{\frac{2N}{3}, [4M-2N]^+\}, & M \leq N \\ \min\{2N, \frac{4M-2N}{3}\}, & M \geq N \end{cases} \tag{8.8}$$

for almost all channel gains.

Remark 8.1 *For a fixed number of receive antennas N, the sum s.d.o.f. d_s is a piece-wise non-decreasing function of the number of transmitting antennas M. d_s in (8.8) consists of five regimes that can be written explicitly as,*

$$d_s = \begin{cases} 0, & M \leq \frac{N}{2} \\ 4M-2N, & \frac{N}{2} \leq M \leq \frac{2N}{3} \\ \frac{2N}{3}, & \frac{2N}{3} \leq M \leq N \\ \frac{4M-2N}{3}, & N \leq M \leq 2N \\ 2N, & M \geq 2N \end{cases} \tag{8.9}$$

i.e., d_s increases linearly with M if $\frac{N}{2} \leq M \leq \frac{2N}{3}$ with slope 4. Then, d_s becomes a constant value of $\frac{2N}{3}$ in the regime $\frac{2N}{3} \leq M \leq N$. Next, d_s increases linearly again with slope $\frac{4}{3}$ until it hits $M = 2N$ and continues as $2N$ afterwards. The sum s.d.o.f. as a function of M for an arbitrary N is shown in Fig. 8.2. We note that when $M = N = 1$ (SISO ICCM), our result reduces to $d_s = 2/3$ in [78].

Remark 8.2 *The term "for almost all channel gains" in Theorem 8.1 refers to the fact that our achievable schemes for the static ICCM depends on real interfer-*

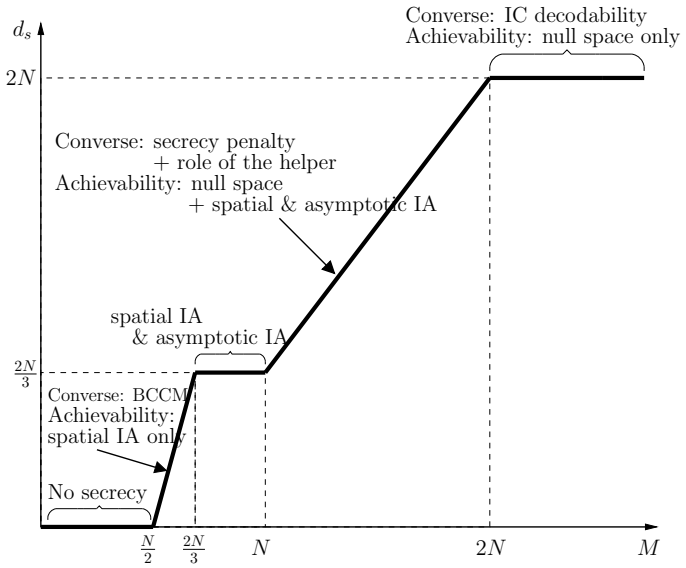


Figure 8.2: Sum s.d.o.f. of $M \times N$ two-user ICCM for a given N.

ence alignment, which necessitates that the channel gains are rationally independent. Since the channel gains are assumed to be drawn randomly from a continuous distribution over $\mathbb{R}^{N\times M}$, the achievable schemes are feasible for almost all channel gains. The same comment holds true for the time-varying ICCM, as the achievable schemes, in this case, assume linear independence of channel gains.

Remark 8.3 *The sum s.d.o.f. in the regime $\frac{N}{2} \leq M \leq \frac{2N}{3}$ coincides with the sum s.d.o.f. of the MIMO BCCM with the transmitter having $2M$ antennas. This implies that, in this regime, there is no loss in the sum s.d.o.f. due to independent coding of the users with respect to the sum s.d.o.f. obtained if cooperation (joint encoding) is allowed.*

Remark 8.4 *The sum s.d.o.f. in the regime $\frac{2N}{3} \leq M \leq N$ is constant. This implies that there is no gain in the sum s.d.o.f. that can be obtained by increasing the number of transmit antennas in this regime.*

Remark 8.5 *The sum s.d.o.f. in the regime $M \geq 2N$ coincides with the sum degrees of freedom (d.o.f.) of the IC with no security constraints. This implies that there is no loss in the sum s.d.o.f. due to enforcing the security constraint, i.e., we achieve security for free in this regime.*

The second result characterizes the entire s.d.o.f. region for the two-user $M \times N$ ICCM.

Theorem 8.2 *The s.d.o.f. region of the two-user $M \times N$ ICCM is given by the set of all pairs (d_1, d_2) that lie in the four-vertex polytope, which is defined as*

$$\mathcal{C} = \left\{ (d_1, d_2) \in \text{conv} \left\{ (0,0), (d_m, 0), (0, d_m), \left(\frac{d_s}{2}, \frac{d_s}{2} \right) \right\} \right\} \tag{8.10}$$

where conv *denotes the convex hull, and d_m is the maximum individual s.d.o.f., which is given by,*

$$d_m = \begin{cases} \min\{\frac{N}{2}, [2M - N]^+\}, & M \leq N \\ \min\{N, \frac{2M-N}{2}\}, & M \geq N \end{cases} \tag{8.11}$$

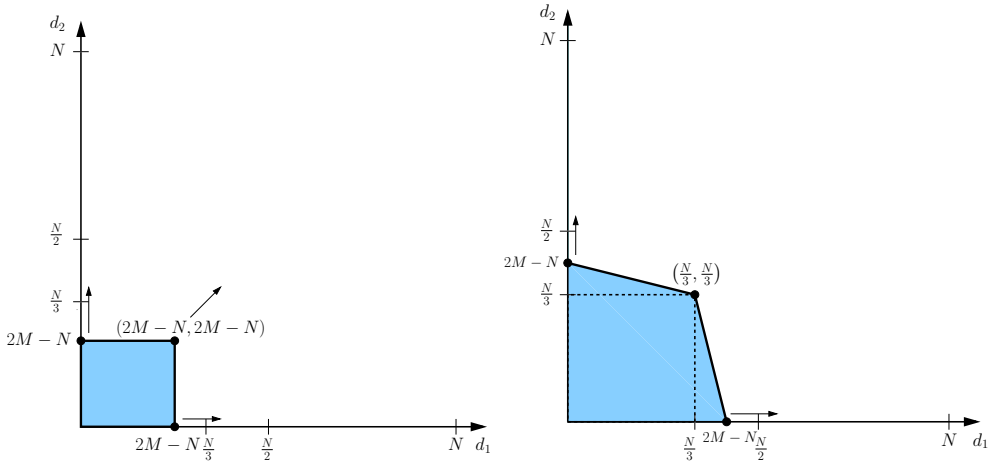
and d_s is defined as in (8.8). The result holds for almost all channel gains.

Remark 8.6 *The s.d.o.f. region $\mathcal{C}$ can be written in an explicit form as*

$$\begin{cases} \{(d_1,d_2): d_1=0,\ d_2=0\}, & M\leq\frac{N}{2} \\ \{(d_1,d_2): d_1\leq 2M-N,\ d_2\leq 2M-N,\ d_1\geq 0,\ d_2\geq 0\}, & \frac{N}{2}\leq M\leq\frac{2N}{3} \\ \{(d_1,d_2): Nd_i+(6M-4N)d_j\leq N(2M-N),\ d_i\geq 0,\ i,\ j=1,2\}, & \frac{2N}{3}\leq M\leq\frac{3N}{4} \\ \{(d_1,d_2): d_1+2d_2\leq N,\ 2d_1+d_2\leq N,\ d_1\geq 0,\ d_2\geq 0\}, & \frac{3N}{4}\leq M\leq N \\ \{(d_1,d_2): d_1+2d_2\leq 2M-N,\ 2d_1+d_2\leq 2M-N,\ d_1\geq 0,\ d_2\geq 0\}, & N\leq M\leq\frac{3N}{2} \\ \{(d_1,d_2): (2M-N)d_i+(4N-2M)d_j\leq N(2M-N),\ d_i\geq 0,\ i,j=1,2\}, & \frac{3N}{2}\leq M\leq 2N \\ \{(d_1,d_2): d_1\leq N,\ d_2\leq N\ d_1\geq 0,\ d_2\geq 0\}, & M\geq 2N \end{cases} \tag{8.12}$$

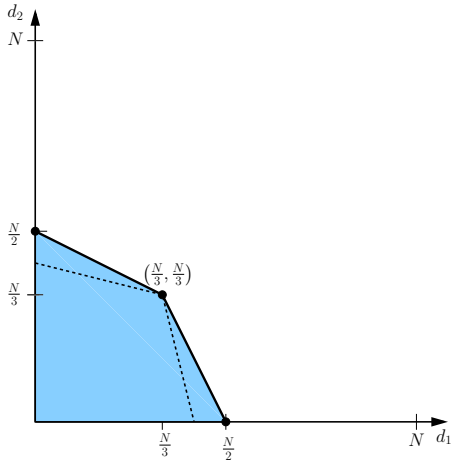
for almost all channel gains.

Remark 8.7 *The maximum individual s.d.o.f. of each user d_m follows a pattern similar to the sum s.d.o.f. in Remark 8.1. d_m coincides with the s.d.o.f. of the MIMO wiretap channel with $2M$ antennas at the transmitter and N receive antennas for $\frac{N}{2}\leq M\leq\frac{3N}{4}$. Then, d_m is constant at $\frac{N}{2}$ for the regime $\frac{3N}{4}\leq M\leq N$. Next, d_m increases linearly with M with slope 1 until $M=\frac{3N}{2}$. The maximum individual s.d.o.f. is constant at N afterwards, which coincides with the maximum individual d.o.f. of MIMO channel with N receive antennas with no security constraints.*

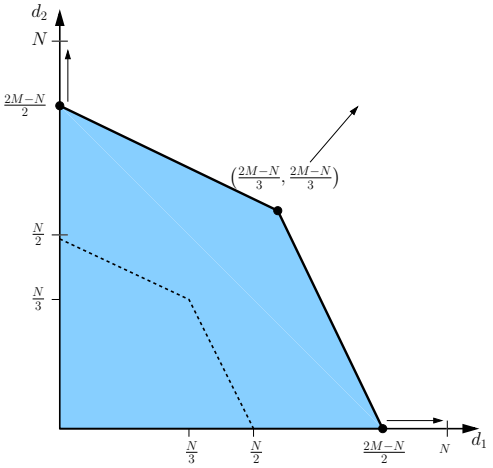


(a) $\frac{N}{2} \leq M \leq \frac{2N}{3}$

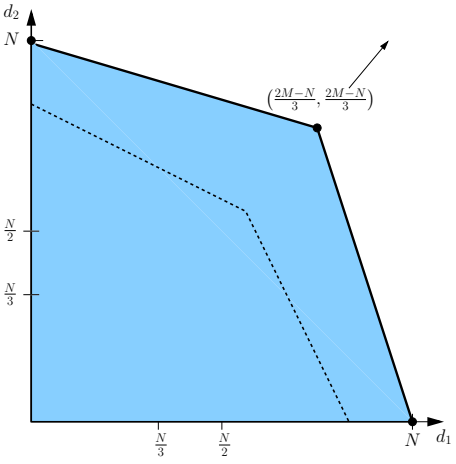
(b) $\frac{2N}{3} \leq M \leq \frac{3N}{4}$



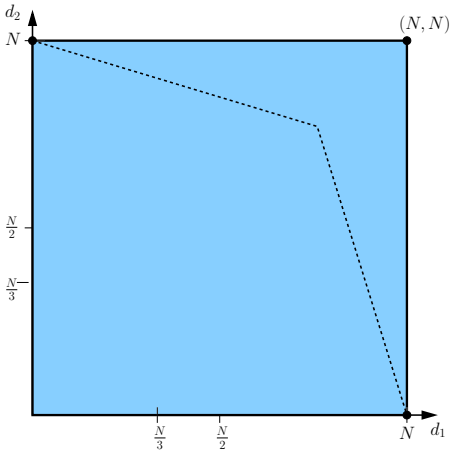
(c) $\frac{3N}{4} \leq M \leq N$



(d) $N \leq M \leq \frac{3N}{2}$



(e) $\frac{3N}{2} \leq M \leq 2N$



(f) $M \geq 2N$

Figure 8.3: Evolution of the s.d.o.f. region with M for a fixed N. The dashed lines in each sub-figure correspond to the rate region in the previous regime for better viewing of how the region evolves.

Remark 8.8 *From Remarks 8.1 and 8.7, we can track the evolution of the s.d.o.f. region by noting the evolution of the extreme points of the corresponding polytope as in Fig. 8.3. We start with a square region with* $d_m = 2M - N$*, this region increases in size while keeping its square shape with the increase of* M *until* $M = \frac{2N}{3}$*. Starting from this point, we cannot support a sum s.d.o.f. larger than* $\frac{2N}{3}$*. Consequently, the sum s.d.o.f. point is kept constant, while the maximum individual s.d.o.f. points can still increase and the s.d.o.f. region is no longer a square region. This continues until* $M = \frac{3N}{4}$*, then the maximum s.d.o.f. points are kept constant to* $\frac{N}{2}$*. This implies that the s.d.o.f. region does not grow in the regime* $\frac{3N}{4} \leq M \leq N$*. The s.d.o.f. region starts increasing in size again from* $M = N$*. The maximum individual s.d.o.f. points increase linearly with slope* 1*, while, the sum s.d.o.f. point increases with slope* $\frac{2}{3}$*. Since slopes are different, the maximum individual s.d.o.f. point hits the* N *bound first at* $M = \frac{3N}{2}$*, while the sum s.d.o.f. point hits this bound at* $M = 2N$ *and we are back to a square region again.*

Remark 8.9 *For the regimes* $\frac{N}{2} \leq M \leq \frac{2N}{3}$ *and* $M \geq 2N$*, the s.d.o.f. region is a square, which implies that both users can transmit with their corresponding maximum s.d.o.f. without sacrificing from their individual s.d.o.f.*

8.5 Outer Bounds for MIMO ICCM

8.5.1 For $M < N$

Allowing cooperation between transmitters yields an upper bound. This results in a BCCM with a single transmitter with $2M$ antennas and two receivers with N antennas each. The s.d.o.f. region of this BCCM is a square whose corner points are $(\min\{[2M-N]^+, N\}, 0), (\min\{[2M-N]^+, N\}, \min\{[2M-N]^+, N\}), (0, \min\{[2M-N]^+, N\})$ [93]. Hence, the individual s.d.o.f. of the two users is upper bounded by:

$$d_i \leq \min\{N, [2M-N]^+\}, \quad i = 1, 2 \tag{8.13}$$

and the sum s.d.o.f. is upper bounded by:

$$d_s \leq 2\min\{N, [2M-N]^+\} = \min\{2N, [4M-2N]^+\} \tag{8.14}$$

Therefore, for $M < N$, the s.d.o.f. region of the MIMO ICCM, $\mathcal{C}$, is upper bounded by the region $\{(d_1, d_2) : d_i \geq 0,\ d_i \leq 2M - N\}$.

8.5.2 For $M \geq N$

We have two distinct upper bounds for the MIMO ICCM when $M \geq N$. From the sum d.o.f. of the two-user IC with no secrecy constraints, $\tilde{d}$, we have the following

bound [105]:

$$d_s \leq \tilde{d} = \min\{M_1 + M_2, N_1 + N_2, \max\{M_1, N_2\}, \max\{M_2, N_1\}\} \tag{8.15}$$

$$= \min\{2M, 2N, \max\{M, N\}\} \tag{8.16}$$

$$= \min\{2N, M\} \tag{8.17}$$

This above upper bound corresponds to the decodability of IC without secrecy constraints. In addition, for the individual s.d.o.f. d_m, we have $d_m \leq N$ if $M \geq N$ from the single-user MIMO channel.

In order to derive an upper bound using the secrecy constraints, we follow the techniques in [78], [79]. From the *secrecy penalty lemma* in [78], we have:

$$nR_i \leq h(\tilde{\mathbf{X}}_1^n) + h(\tilde{\mathbf{X}}_2^n) - h(\mathbf{Y}_j^n) + nc_1 \tag{8.18}$$

where $i \neq j$, and $\tilde{\mathbf{X}}_i^n = \mathbf{X}_i^n + \tilde{\mathbf{N}}_i^n$ is a finite-variance Gaussian perturbed channel input; here small Gaussian perturbation is introduced in order to avoid mixing continuous and discrete entropies, see [78]. In addition, we have the following vectorized version of the *role of a helper lemma* of [78] (see also [79]).

Lemma 8.1 (MIMO role of a helper lemma) *For $M \geq N$, reliable decoding of the jth transmitter at the ith receiver, $i \neq j$, is guaranteed if the perturbed channel input $\tilde{\mathbf{X}}_i^n$ satisfies*

$$h(\tilde{\mathbf{X}}_i^n) \leq h(\tilde{\mathbf{X}}_i^{n^{(2)}}) + h(\mathbf{Y}_j^n) - nR_j + nc_2, \quad i \neq j \tag{8.19}$$

where $\tilde{\mathbf{X}}_i^{n^{(2)}} = [\tilde{X}_i^n(N+1) \quad \tilde{X}_i^n(N+2) \quad \dots \quad \tilde{X}_i^n(M)]$.

Remark 8.10 $\tilde{\mathbf{X}}_i^{n^{(1)}} = [\tilde{X}_i^n(1) \quad \tilde{X}_i^n(2) \quad \dots \quad \tilde{X}_i^n(N)]$ *represents the first* N *(perturbed) antenna inputs, and* $\tilde{\mathbf{X}}_i^{n^{(2)}} = [\tilde{X}_i^n(N+1) \quad \tilde{X}_i^n(N+2) \quad \dots \quad \tilde{X}_i^n(M)]$ *represents the* $M-N$ *extra (perturbed) antenna inputs that can be used for null space transmission. Note that, here we have* $M \geq N$*, therefore,* $\tilde{\mathbf{X}}_i^{n^{(2)}}$ *is well-defined. We note also that intuitively we should separate the upper bounding of differential entropies of* $\tilde{\mathbf{X}}_i^{n^{(1)}}$ *and* $\tilde{\mathbf{X}}_i^{n^{(2)}}$ *because the null space components do not hurt the other receiver (in fact, they are invisible to the other receiver) as* $\tilde{\mathbf{X}}_i^{n^{(1)}}$ *components do. Consequently, we upper bound the differential entropy of these components directly using Gaussian entropy bounds.*

Proof: Let $\tilde{\mathbf{X}}_i^n = [\tilde{\mathbf{X}}_i^{n^{(1)}} \tilde{\mathbf{X}}_i^{n^{(2)}}]$. Using Fano's inequality, the rate of user j, where $j \neq i$, is upper bounded by

$$
\begin{aligned}
nR_j &\leq I(\mathbf{X}_j^n; \mathbf{Y}_j^n) + nc_3 && (8.20)\\
&= h(\mathbf{Y}_j^n) - h(\mathbf{Y}_j^n | \mathbf{X}_j^n) + nc_3 && (8.21)\\
&\leq h(\mathbf{Y}_j^n) - h(\mathbf{Y}_j^n | \mathbf{X}_j^n, \tilde{\mathbf{X}}_i^{n^{(2)}}) + nc_3 && (8.22)\\
&= h(\mathbf{Y}_j^n) - h(\mathbf{H}_{jj}\mathbf{X}_j^n + \mathbf{H}_{ij}^{(1)}\mathbf{X}_i^{n^{(1)}} + \mathbf{H}_{ij}^{(2)}\mathbf{X}_i^{n^{(2)}} + \mathbf{N}_j^n | \mathbf{X}_j^n, \tilde{\mathbf{X}}_i^{n^{(2)}}) + nc_3 && (8.23)\\
&\leq h(\mathbf{Y}_j^n) - h(\mathbf{H}_{ij}^{(1)}\tilde{\mathbf{X}}_i^{n^{(1)}} + \mathbf{H}_{ij}^{(2)}\tilde{\mathbf{X}}_i^{n^{(2)}} | \mathbf{X}_j^n, \tilde{\mathbf{X}}_i^{n^{(2)}}) + nc_3 && (8.24)\\
&= h(\mathbf{Y}_j^n) - h(\tilde{\mathbf{X}}_i^{n^{(1)}} | \tilde{\mathbf{X}}_i^{n^{(2)}}) + nc_2 && (8.25)
\end{aligned}
$$

where $\tilde{\mathbf{X}}_i^n = \mathbf{X}_i^n + \tilde{\mathbf{N}}_i^n$ such that $\tilde{\mathbf{N}}_i^n \sim \mathcal{N}(\mathbf{0},\ \rho_i \mathbf{I}_M)$, where $\rho_i < \min_j \frac{1}{||\mathbf{H}_{ij}||^2}$. (8.24) follows from considering a stochastically equivalent version of $\mathbf{Y}_j$ given by $\tilde{\mathbf{Y}}_j =$

$\mathbf{H}_{jj}\mathbf{X}_j+\mathbf{H}_{ij}\tilde{\mathbf{X}}_i+\bar{\mathbf{N}}_j$, where $\bar{\mathbf{N}}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_M-\rho_i\mathbf{H}_{ij}\mathbf{H}_{ij}^T)$, hence $h(\mathbf{H}_{jj}\mathbf{X}_j^n+\mathbf{H}_{ij}^{(1)}\mathbf{X}_i^{n^{(1)}}+\mathbf{H}_{ij}^{(2)}\mathbf{X}_i^{n^{(2)}}+\mathbf{N}_j|\mathbf{X}_j^n,\tilde{\mathbf{X}}_i^{n^{(2)}}) \geq h(\mathbf{H}_{jj}\mathbf{X}_j^n+\mathbf{H}_{ij}^{(1)}\tilde{\mathbf{X}}_i^{n^{(1)}}+\mathbf{H}_{ij}^{(2)}\tilde{\mathbf{X}}_i^{n^{(2)}}|\mathbf{X}_j^n,\tilde{\mathbf{X}}_i^{n^{(2)}})$. (8.25) follows from the scaling property of the differential entropy which results in an additional constant that does not depend on P. Hence, the conditional entropy of the ith user's channel input is upper bounded by

$$h(\tilde{\mathbf{X}}_i^{n^{(1)}}|\tilde{\mathbf{X}}_i^{n^{(2)}}) \leq h(\mathbf{Y}_j^n) - nR_j + nc_2 \tag{8.26}$$

By applying chain rule for users' inputs $h(\tilde{\mathbf{X}}_i^n) = h(\tilde{\mathbf{X}}_i^{n^{(2)}}) + h(\tilde{\mathbf{X}}_i^{n^{(1)}}|\tilde{\mathbf{X}}_i^{n^{(2)}})$, we have (8.19). ■

By applying the secrecy penalty and MIMO role of a helper lemmas in (8.18), (8.19) for user 1, we have the following upper bound

$$nR_1 \leq h(\tilde{\mathbf{X}}_1^{n^{(2)}}) + h(\tilde{\mathbf{X}}_2^{n^{(2)}}) + h(\mathbf{Y}_1^n) - nR_1 - nR_2 + nc_4 \tag{8.27}$$

which is equivalent to

$$n(2R_1 + R_2) \leq h(\tilde{\mathbf{X}}_1^{n^{(2)}}) + h(\tilde{\mathbf{X}}_2^{n^{(2)}}) + h(\mathbf{Y}_1^n) + nc_4 \tag{8.28}$$

Using the fact that Gaussian random variables maximize the differential entropy, we obtain:

$$\begin{aligned} n(2R_1 + R_2) \leq & h(\tilde{\mathbf{X}}_1^{n^{(2)}}) + h(\tilde{\mathbf{X}}_2^{n^{(2)}}) + h(\mathbf{Y}_1^n) + nc_4 & (8.29)\\ \leq & (M-N)\cdot\frac{n}{2}\log P + (M-N)\cdot\frac{n}{2}\log P + N\cdot\frac{n}{2}\log P + nc_5 & (8.30) \end{aligned}$$

$$=(2M-N)\cdot\frac{n}{2}\log P+nc_5 \tag{8.31}$$

Dividing by n yields,

$$2R_1+R_2\leq(2M-N)\cdot\frac{1}{2}\log P+c_5 \tag{8.32}$$

and by dividing by $\frac{1}{2}\log P$ and taking the limit as $P\to\infty$, we obtain:

$$2d_1+d_2\leq 2M-N \tag{8.33}$$

By symmetry, we obtain the following upper bound by writing the secrecy penalty and role of a helper lemmas for user 2

$$d_1+2d_2\leq 2M-N \tag{8.34}$$

Also, adding (8.33) and (8.34), we obtain the following upper bound on the sum s.d.o.f. d_s

$$d_1+d_2\leq\frac{4M-2N}{3} \tag{8.35}$$

Consequently, the s.d.o.f. region $\mathcal{C}$ is upper bounded by the region $\{(d_1,d_2):d_i+2d_j\leq 2M-N,\ d_i\geq 0,\quad i,\ j=1,2,j\neq i\}$ for $M\geq N$.

Focusing on the sum s.d.o.f., from (8.35) and (8.17) we have the upper bound

on the sum s.d.o.f. as

$$d_s \leq \min\left\{\frac{4M-2N}{3}, M, 2N\right\} \tag{8.36}$$

If the first term in the upper bound is not active, then $M \leq \frac{4M-2N}{3}$ or $2N \leq \frac{4M-2N}{3}$, which both lead to $M \geq 2N$ and hence the M term in the upper bound is never active, and the sum s.d.o.f. upper bound is

$$d_s \leq \min\left\{\frac{4M-2N}{3}, 2N\right\} \tag{8.37}$$

Focusing on the maximum individual s.d.o.f. points, from (8.33) and (8.34), we have $d_m \leq \frac{2M-N}{2}$. Including the maximum d.o.f. upper bound for the MIMO channel, we have

$$d_m \leq \min\left\{\frac{2M-N}{2}, N\right\} \tag{8.38}$$

8.5.3 Combining Both Bounds

First, we note that since the outer bounds in (8.13)-(8.14) and (8.36)-(8.37) define a bounded polyhedron in $\mathbb{R}^2$, the outer bounds form a polytope as in [82]. Thus, it is sufficient to characterize upper bounds for its extreme points.

Now, we note that increasing the number of transmit antennas of both transmitters cannot decrease the s.d.o.f. of ICCM for a fixed number of receiver antennas. Therefore, $d_s \leq \frac{2N}{3}$ corresponding to the case of $M = N$ for both the sum s.d.o.f. point and the maximum individual s.d.o.f. point. For the sum s.d.o.f. point, the upper bound in (8.35) is $\frac{2N}{3}$ for the case $M = N$. Combining the bounds (8.14)

and $d_s \leq \frac{2N}{3}$, we have $d_s \leq \min\{\frac{2N}{3}, 4M-2N\}$ for $M \leq N$. Consequently, the upper bound for the sum s.d.o.f. of the ICCM for any arbitrary M and N is,

$$d_s = \begin{cases} \min\{\frac{2N}{3}, [4M-2N]^+\}, & M \leq N \\ \min\{2N, \frac{4M-2N}{3}\}, & M \geq N \end{cases} \tag{8.39}$$

Similarly, for the maximum individual s.d.o.f. point, the upper bound in (8.38) for the case $M = N$ is $\frac{N}{2}$. Hence, combining this with (8.13), $d_m \leq \min\{\frac{N}{2}, [2M-N]^+\}$ for $M \leq N$. Consequently, the maximum individual s.d.o.f. of the ICCM for any arbitrary M and N is,

$$d_m = \begin{cases} \min\{\frac{N}{2}, [2M-N]^+\}, & M \leq N \\ \min\{N, \frac{2M-N}{2}\}, & M \geq N \end{cases} \tag{8.40}$$

Since the problem is symmetric with respect to the two users, there exists a symmetric sum s.d.o.f. point $\left(\frac{d_s}{2}, \frac{d_s}{2}\right)$ and two maximum individual s.d.o.f. points $(0, d_m), (d_m, 0)$.

8.6 Achievable Scheme for Sum s.d.o.f. of the 2×2 ICCM for Static Channels

In this section, we develop optimal achievable schemes to match the presented upper bounds. First, we focus on the sum s.d.o.f. point $\left(\frac{d_s}{2}, \frac{d_s}{2}\right)$ for the case of static channels, i.e., $\mathbf{H}_{ij}(t) = \mathbf{H}_{ij}, \ \forall t$. We start by proposing a novel achievable scheme for the 2×2 ICCM system using asymptotic real interference alignment. Then,

we build on this achievable scheme to obtain achievable schemes for any M, N by combining spatial alignment and exploiting the null space (whenever possible, i.e., $M > N$) with the 2×2 scheme. Real interference alignment is not needed in regimes that correspond to integer s.d.o.f., i.e., it suffices to use Gaussian codebooks along with spatial alignment and/or null space transmission in these cases. To carry out the secure rate calculation, we use the following result from [70] which states that the following secure rates are achievable for the ICCM[2]:

$$R_1 \leq I(\mathbf{V}_1; \mathbf{Y}_1) - I(\mathbf{V}_1; \mathbf{Y}_2|\mathbf{V}_2) \tag{8.41}$$

$$R_2 \leq I(\mathbf{V}_2; \mathbf{Y}_2) - I(\mathbf{V}_2; \mathbf{Y}_1|\mathbf{V}_1) \tag{8.42}$$

8.6.1 Basic System: 2×2 MIMO ICCM

The basic building blocks of all achievable schemes for the sum s.d.o.f. point when the channel is static are the 1×1 SISO ICCM and the 2×2 MIMO ICCM systems. We can reduce all other regimes to one of these cases by proper vector space manipulations. The achievable scheme for the 1×1 SISO ICCM is given in [78]. In this section, we give an achievable scheme for the 2×2 MIMO ICCM. The achievable scheme for the 2×2 system combines spatial alignment with asymptotic real interference alignment. To use asymptotic real interference alignment, the secure signal $\mathbf{V}_i$ and the cooperative jamming signal $\mathbf{U}_i$ are constructed as a linear combination

[2]Interestingly, the rate region in (8.41) is also achievable under the strong security constraint as shown in [126, Theorem 1] and [127, Remark 1]. This implies that our s.d.o.f. region is in fact valid if we changed the security constraint to the strong security constraint, i.e., $I(W_i; \mathbf{Y}_j^n) \leq \epsilon$, for $i, j \in \{1, 2\}$ without normalization with n. Note that any scheme that achieves the strong security constraint is a valid achievable scheme under the weak security constraint as well.

of structured signals picked from PAM constellation $C(a, Q)$ with proper parameters that will be identified shortly. The transmitted signals are:

$$\mathbf{X}_1 = \mathbf{H}_{12}^{-1}\mathbf{V}_1 + \mathbf{H}_{11}^{-1}\mathbf{U}_1 \tag{8.43}$$

$$\mathbf{X}_2 = \mathbf{H}_{21}^{-1}\mathbf{V}_2 + \mathbf{H}_{22}^{-1}\mathbf{U}_2 \tag{8.44}$$

The received signals are:

$$\begin{aligned}\mathbf{Y}_1 &= \mathbf{H}_{11}\mathbf{H}_{12}^{-1}\mathbf{V}_1 + (\mathbf{U}_1 + \mathbf{V}_2) + \mathbf{H}_{21}\mathbf{H}_{22}^{-1}\mathbf{U}_2 + \mathbf{N}_1 \\ &= \mathbf{A}\mathbf{V}_1 + (\mathbf{U}_1 + \mathbf{V}_2) + \mathbf{B}\mathbf{U}_2 + \mathbf{N}_1 \end{aligned} \tag{8.45}$$

and

$$\begin{aligned}\mathbf{Y}_2 &= (\mathbf{V}_1 + \mathbf{U}_2) + \mathbf{H}_{12}\mathbf{H}_{11}^{-1}\mathbf{U}_1 + \mathbf{H}_{22}\mathbf{H}_{21}^{-1}\mathbf{V}_2 + \mathbf{N}_2 \\ &= (\mathbf{V}_1 + \mathbf{U}_2) + \bar{\mathbf{B}}\mathbf{U}_1 + \bar{\mathbf{A}}\mathbf{V}_2 + \mathbf{N}_2 \end{aligned} \tag{8.46}$$

Considering the first receiver without loss of generality, we note that $\mathbf{A} = \mathbf{H}_{11}\mathbf{H}_{12}^{-1}$, $\mathbf{B} = \mathbf{H}_{21}\mathbf{H}_{22}^{-1}$ are generally non-diagonal with rationally independent elements almost surely. Using exact real interference alignment requires constructing 5 irrational dimensions in order to decode $\mathbf{V}_i$ with arbitrary small probability of error. However, this wastes the observation space of the second antenna and achieves an s.d.o.f. of 2/5 from only one antenna.

To see this, let a_{ij}, $i, j \in \{1, 2\}$ be the (i, j)th element of matrix $\mathbf{A}$, and

similarly for the other matrices $\mathbf{B}, \bar{\mathbf{A}} = \mathbf{H}_{22}\mathbf{H}_{21}^{-1}, \bar{\mathbf{B}} = \mathbf{H}_{12}\mathbf{H}_{11}^{-1}$. Then, the received signal at receiver 1 is

$$\mathbf{Y}_1 = \begin{bmatrix} a_{11}v_{11} + a_{12}v_{12} + (u_{11} + v_{21}) + b_{11}u_{21} + b_{12}u_{22} \\ a_{21}v_{11} + a_{22}v_{12} + (u_{12} + v_{22}) + b_{21}u_{21} + b_{22}u_{22} \end{bmatrix} + \mathbf{N}_1 \tag{8.47}$$

where $\mathbf{V}_1 = [v_{11}\ v_{12}]^T, \mathbf{U}_2 = [u_{21}\ u_{22}]^T$. The scaling factors $\{a_{ij}\}_{i,j=1,2}, \{b_{ij}\}_{i,j=1,2}$, and 1 are rationally independent almost surely. Thus, in order to decode v_{11}, v_{12} with arbitrarily small probability of error using exact real interference alignment as in [101], [78], we need to construct at least 5 irrational dimensions. We note also that from antenna 2, the same symbols v_{11}, v_{12} can be decoded. Hence, by using exact real interference alignment, we exploit the observation of the first antenna only, as the second antenna does not give any new information. Consequently, from the first antenna, we achieve an s.d.o.f. of 2/5, as 2 components of the secure signal can be decoded out of the 5 irrational dimensions needed for correct decoding. To minimize the required irrational dimensions, we need to leave one of v_{11} or v_{12} to be in a separate irrational dimension at each antenna, while the other component is aligned with u_{21}, u_{22}. This type of alignment can be done asymptotically by breaking $\{v_{ij}\}_{i,j=1,2}, \{u_{ij}\}_{i,j=1,2}$ into sufficiently large number of components. Hence, for the first antenna, the components of signal v_{11} are in separate irrational dimensions that cover 1/3 of the total dimensions, and the signal components of $(u_{11}+v_{21})$ cover 1/3 of the total dimensions, while the signal components of v_{12}, u_{21}, u_{22} are asymptotically aligned together and cover slightly larger than 1/3 of the total irrational

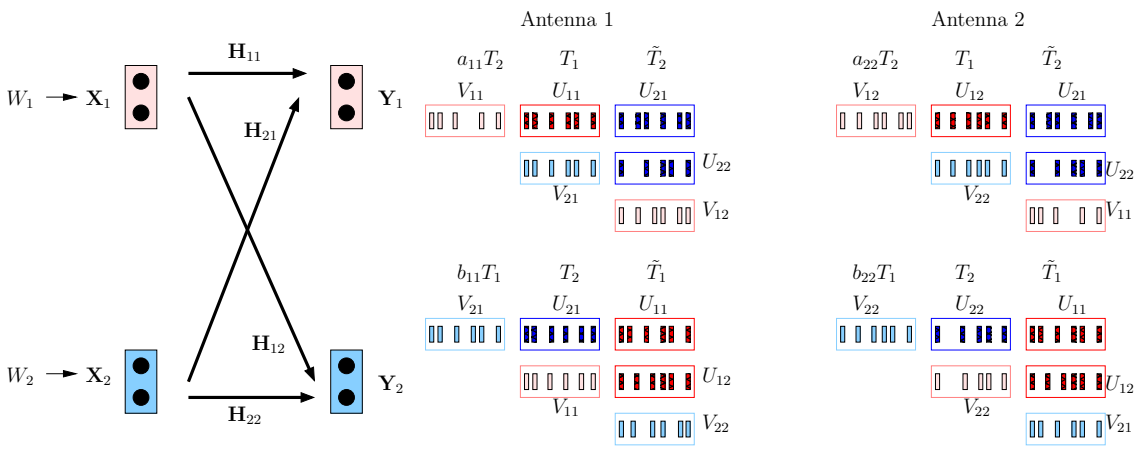


Figure 8.4: Illustration of asymptotic real interference alignment for the 2×2 system.

dimensions. Consequently, user 1 can achieve 1/3 s.d.o.f. from the first antenna. A similar argument holds for the second antenna with switching the roles of v_{11} and v_{12}. This scheme is illustrated in Fig. 8.4.

We begin discussing the details of the asymptotic real interference alignment [86] by defining sets of irrational dimensions T_i

$$T_1 = \left\{ \prod_{i,j=1,i\neq j}^{2} \bar{a}_{ij}^{r_{ij}} \prod_{i,j=1}^{2} \bar{b}_{ij}^{s_{ij}} : r_{ij}, s_{ij} = 1, \ldots, m \right\} \tag{8.48}$$

$$T_2 = \left\{ \prod_{i,j=1,i\neq j}^{2} a_{ij}^{r_{ij}} \prod_{i,j=1}^{2} b_{ij}^{s_{ij}} : r_{ij}, s_{ij} = 1, \ldots, m \right\} \tag{8.49}$$

We define $\mathbf{t}_1, \mathbf{t}_2$ to be the vectors constructed by enumerating all elements of T_1, T_2 sets, respectively. The cardinality of T_i (which is also the length of the $\mathbf{t}_i$ vector) is given by

$$M_T = |T_i| = m^6, \quad i = 1, 2 \tag{8.50}$$

We note that T_i set does not contain the gains $a_{ii}, \bar{a}_{ii}$ and hence multiplying by these channel gains produces new M_T irrational dimensions. On the other hand, multi-

plying with any channel gain that appears in T_i results in asymptotically aligning this signal within $\tilde{T}_i$ set which is defined as

$$\tilde{T}_1 = \left\{ \prod_{i,j=1, i\neq j}^{2} \bar{a}_{ij}^{r_{ij}} \prod_{i,j=1}^{2} \bar{b}_{ij}^{s_{ij}} : r_{ij}, s_{ij} = 1, \ldots, m+1 \right\} \tag{8.51}$$

$$\tilde{T}_2 = \left\{ \prod_{i,j=1, i\neq j}^{2} a_{ij}^{r_{ij}} \prod_{i,j=1}^{2} b_{ij}^{s_{ij}} : r_{ij}, s_{ij} = 1, \ldots, m+1 \right\} \tag{8.52}$$

with cardinality of

$$M_R = |\tilde{T}_i| = (m+1)^6, \quad i = 1, 2 \tag{8.53}$$

Now, we give the explicit structure of the transmitted signals. The vectors $\mathbf{V}_i$, $\mathbf{U}_i$ are 2×1 vectors. Each component is constructed out of irrational combinations of M_T PAM signals $\mathbf{v}_{ij} = [v_{ij1}\, v_{ij2} \cdots v_{ijM_T}]^T$ representing secure signal components of user i from antenna j. Generate $\mathbf{u}_i = [u_{ij1}\, u_{ij2} \cdots u_{ijM_T}]^T$ as cooperative jamming signal as follows

$$\mathbf{V}_1 = \begin{bmatrix} \mathbf{t}_2^T \mathbf{v}_{11} \\ \mathbf{t}_2^T \mathbf{v}_{12} \end{bmatrix}, \quad \mathbf{U}_1 = \begin{bmatrix} \mathbf{t}_1^T \mathbf{u}_{11} \\ \mathbf{t}_1^T \mathbf{u}_{12} \end{bmatrix} \tag{8.54}$$

$$\mathbf{V}_2 = \begin{bmatrix} \mathbf{t}_1^T \mathbf{v}_{21} \\ \mathbf{t}_1^T \mathbf{v}_{22} \end{bmatrix}, \quad \mathbf{U}_2 = \begin{bmatrix} \mathbf{t}_2^T \mathbf{v}_{21} \\ \mathbf{t}_2^T \mathbf{v}_{22} \end{bmatrix} \tag{8.55}$$

This means that the alignment of $\mathbf{V}_1$ and $\mathbf{U}_2$ is carried over the T_2 set, while that of $\mathbf{V}_2$ and $\mathbf{U}_1$ over the T_1 set. Using this construction, the received signal at receiver

1 is

$$\mathbf{Y}_1 = \begin{bmatrix} a_{11}\mathbf{t}_2^T\mathbf{v}_{11} + \mathbf{t}_1^T(\mathbf{u}_{11} + \mathbf{v}_{21}) + \mathbf{t}_2^T(a_{12}\mathbf{v}_{12} + b_{11}\mathbf{u}_{21} + b_{12}\mathbf{u}_{22}) \\ a_{22}\mathbf{t}_2^T\mathbf{v}_{12} + \mathbf{t}_1^T(\mathbf{u}_{12} + \mathbf{v}_{22}) + \mathbf{t}_2^T(a_{21}\mathbf{v}_{11} + b_{21}\mathbf{u}_{21} + b_{22}\mathbf{u}_{22}) \end{bmatrix} + \mathbf{N}_1 \quad (8.56)$$

Lemma 8.2 *The sum s.d.o.f. of $\frac{4}{3}$ is achievable using the combination of asymptotic alignment and spatial alignment shown in this section with signals picked from PAM constellation $C(a, Q)$ with $Q = P^{\frac{1-\delta}{2(M_\Sigma+\delta)}}, a = \gamma\frac{P^{\frac{1}{2}}}{Q}$, where $M_\Sigma = 2m^6 + (m+1)^6$ for arbitrarily large integer m, and any $\delta > 0$.*

Remark 8.11 *From (8.56), we first note that using this type of alignment ensures exact alignment of user 2's secure signals with cooperative jamming signal generated by user 1 as in $(\mathbf{u}_{11} + \mathbf{v}_{21})$ terms. This exact alignment guarantees security as in the SISO case in [78]. In addition, at each antenna, only one secure signal component lies in a separate irrational dimension for decodability as in $a_{11}\mathbf{t}_2^T\mathbf{v}_{11}$ and $a_{22}\mathbf{t}_2^T\mathbf{v}_{12}$, while the other component aligns with user 2's cooperative jamming signal over the set $\tilde{T}_2$. Therefore, the intended secure signal at each antenna covers M_T dimensions out of M_Σ dimensions. Consequently, achievable s.d.o.f. per antenna is approximately $\frac{M_T}{M_\Sigma}$ which approaches $1/3$ as m gets large. Hence, we achieve a total of $2/3$ s.d.o.f. per user, and a total of $d_s = 4/3$ s.d.o.f. for the system.*

Proof: The total number of dimensions at antenna 1 (and similarly antenna 2) needed in this case is

$$M_\Sigma = |a_{11}T_2 \cup T_1 \cup \tilde{T}_2| = 2m^6 + (m+1)^6 \quad (8.57)$$

By choosing the parameters of the PAM constellation as

$$Q = P^{\frac{1-\delta}{2(M_\Sigma+\delta)}}, \qquad a = \gamma\frac{P^{\frac{1}{2}}}{Q} \tag{8.58}$$

the average power constraint is satisfied, and the probability of error can be made arbitrarily small as $P \to \infty$ as in [101], [78]. We can also decode $\mathbf{U}_2$ perfectly at receiver 1 after decoding $\mathbf{V}_1$. By subtracting $\mathbf{V}_1$ from $\mathbf{Y}_1$, we have

$$\mathbf{Y}_1' = (\mathbf{U}_1 + \mathbf{V}_2) + \mathbf{B}\mathbf{U}_2 + \mathbf{N}_1 \tag{8.59}$$

By filtering the received observations using $\mathbf{C} = \mathbf{B}^{-1}$, we have

$$\mathbf{Y}_1'' = \mathbf{B}^{-1}(\mathbf{U}_1 + \mathbf{V}_2) + \mathbf{U}_2 + \mathbf{N}_1'' \tag{8.60}$$

$$= \begin{bmatrix} c_{11}\mathbf{t}_1^T(\mathbf{u}_{11} + \mathbf{v}_{21}) + c_{12}\mathbf{t}_1^T(\mathbf{u}_{12} + \mathbf{v}_{22}) + \mathbf{t}_2^T\mathbf{u}_{21} \\ c_{21}\mathbf{t}_1^T(\mathbf{u}_{11} + \mathbf{v}_{21}) + c_{22}\mathbf{t}_1^T(\mathbf{u}_{12} + \mathbf{v}_{22}) + \mathbf{t}_2^T\mathbf{u}_{22} \end{bmatrix} + \mathbf{N}_1'' \tag{8.61}$$

where $\mathbf{N}_1'' = \mathbf{B}^{-1}\mathbf{N}_1$. Since no specific alignment procedure has been designed for the $\mathbf{C}$ matrix, all these signals are received in separate irrational dimensions. The total required dimensions in this case is $3M_T = 3m^6 < M_\Sigma$, and hence decodable.

Now, we evaluate the rates in (8.41) focusing on user 1. Using the parameters chosen in (8.58), $\mathbf{V}_1$ is received with asymptotically vanishing probability of error. Consequently, the first term of (8.41) can be lower bounded using data processing

and Fano's inequality as

$$I(\mathbf{V}_1;\mathbf{Y}_1) \geq I(\mathbf{V}_1;\hat{\mathbf{V}}_1) \tag{8.62}$$

$$= H(\mathbf{V}_1) - H(\mathbf{V}_1|\hat{\mathbf{V}}_1) \tag{8.63}$$

$$\geq (1-P_e)\log(2Q+1)^{2M_T} - 1 \tag{8.64}$$

$$= (1-P_e)\frac{2M_T(1-\delta)}{M_\Sigma+\delta}\cdot\frac{1}{2}\log P + o(\log P) \tag{8.65}$$

We can upper bound the leakage as

$$I(\mathbf{V}_1;\mathbf{Y}_2|\mathbf{V}_2) \leq I(\mathbf{V}_1;(\mathbf{V}_1+\mathbf{U}_2)+\bar{\mathbf{B}}\mathbf{U}_1+\bar{\mathbf{A}}\mathbf{V}_2|\mathbf{V}_2) \tag{8.66}$$

$$= H((\mathbf{V}_1+\mathbf{U}_2)+\bar{\mathbf{B}}\mathbf{U}_1) - H(\mathbf{U}_2+\bar{\mathbf{B}}\mathbf{U}_1) \tag{8.67}$$

The first term in (8.67) can be upper bounded by

$$H(\mathbf{V}_1+\mathbf{U}_2)+\bar{\mathbf{B}}\mathbf{U}_1) = H((\bar{\mathbf{B}}^{-1}(\mathbf{V}_1+\mathbf{U}_2)+\mathbf{U}_1) \tag{8.68}$$

$$= H\left(\begin{bmatrix}\bar{c}_{11}\mathbf{t}_2^T(\mathbf{u}_{21}+\mathbf{v}_{11})+\bar{c}_{12}\mathbf{t}_2^T(\mathbf{u}_{22}+\mathbf{v}_{12})+\mathbf{t}_1^T\mathbf{u}_{11}\\ \bar{c}_{21}\mathbf{t}_2^T(\mathbf{u}_{21}+\mathbf{v}_{11})+\bar{c}_{22}\mathbf{t}_2^T(\mathbf{u}_{22}+\mathbf{v}_{12})+\mathbf{t}_1^T\mathbf{u}_{12}\end{bmatrix}\right) \tag{8.69}$$

$$= H\left(\begin{bmatrix}\mathbf{V}_1+\mathbf{U}_2\\ \\ \mathbf{U}_1\end{bmatrix}\right) \tag{8.70}$$

$$= \log\left((4Q+1)^{2M_T}(2Q+1)^{2M_T}\right) \tag{8.71}$$

where $\bar{\mathbf{C}} = \bar{\mathbf{B}}^{-1}$, (8.68) holds since $\bar{\mathbf{C}}$ is invertible, and (8.70) follows from the fact

that all signal components lie in different irrational dimensions with a total number of dimensions of $3M_T < M_\Sigma$, which in turn makes these signals decodable for large enough P. Thus, the transformation $\begin{bmatrix}\bar{\mathbf{B}}^{-1} & \mathbf{I}\end{bmatrix}$ is invertible. Similarly, the second term in (8.67) which solely contains cooperative jamming signals, is

$$H(\mathbf{U}_2 + \bar{\mathbf{B}}\mathbf{U}_1) = H\left(\begin{bmatrix}\mathbf{U}_2 \\ \\ \mathbf{U}_1\end{bmatrix}\right) = \log\left((2Q+1)^{4M_T}\right) \tag{8.72}$$

Then, the leakage in (8.67) is upper bounded by

$$I(\mathbf{V}_1; \mathbf{Y}_2|\mathbf{V}_2) \leq \log\left(\frac{4Q+1}{2Q+1}\right)^{2M_T} \tag{8.73}$$

$$\leq 2M_T + o(\log P) \tag{8.74}$$

Therefore, user 1's rate is lower bounded by

$$R_1 \geq I(\mathbf{V}_1; \mathbf{Y}_1) - I(\mathbf{V}_1; \mathbf{Y}_2|\mathbf{V}_2) \tag{8.75}$$

$$= 2M_T\left((1-P_e)\frac{(1-\delta)}{M_\Sigma+\delta}\cdot\frac{1}{2}\log P - 1\right) + o(\log P) \tag{8.76}$$

By normalizing by $\frac{1}{2}\log P$ and taking $P \to \infty$,

$$d_1 \geq \frac{2M_T(1-\delta)}{M_\Sigma+\delta} \tag{8.77}$$

$$= \frac{2m^6(1-\delta)}{2m^6+(m+1)^6+\delta} \tag{8.78}$$

$$\geq \frac{2(1-\delta)}{2+\left(1+\frac{1}{m}\right)^6+\delta} \tag{8.79}$$

As $m \to \infty$ and $\delta \to 0$, the achievable s.d.o.f is $2/3$ for each user, and hence the sum s.d.o.f. is $\frac{4}{3}$. ■

Remark 8.12 *We note that for the SISO system, we do not need any asymptotic alignment. By specializing the spatial alignment presented here to the SISO case, i.e.,*

$$X_1 = \frac{1}{h_{12}}V_1 + \frac{1}{h_{11}}U_1 \tag{8.80}$$

$$X_2 = \frac{1}{h_{21}}V_2 + \frac{1}{h_{22}}U_2 \tag{8.81}$$

we see that the received signals fit exactly into 3 irrational dimensions. Hence, 1/3 s.d.o.f. per user is achievable as in [78]. Therefore, we focus our attention to the presentation of achievable schemes for the cases that result in a 2×2 system, since the SISO case can be obtained as a special case of the 2×2 achievable scheme by ignoring the asymptotic alignment step and replacing with an exact real interference alignment step.

8.7 Achievable Scheme for Sum s.d.o.f. of the $M \times N$ MIMO ICCM

8.7.1 $\frac{N}{2} \leq M \leq \frac{2N}{3}$

In this case, the sum s.d.o.f. is an integer. Hence, we use Gaussian codebooks for transmission of the secure signal $\mathbf{V}_i$ and the cooperative jamming signal $\mathbf{U}_i$. We

precode these signals such that the secure signal of one user lies in the same subspace as the cooperative jamming signal of the other user.

Transmitted Signals: Each user transmits a Gaussian secure signal $\mathbf{V}_i$, and a Gaussian cooperative jamming signal $\mathbf{U}_i$. The signals $\mathbf{V}_i, \mathbf{U}_i \sim \mathcal{N}(\mathbf{0}, \eta_1 P \mathbf{I}_{2M-N})$ are of $2M-N$ dimensions, and independent from each other, where η_1 is a constant, which is chosen to satisfy the power constraint. Let $\mathbf{P}_i, \mathbf{Q}_i \in \mathbb{R}^{M \times (2M-N)}$ be the precoding matrices for $\mathbf{V}_i$, $\mathbf{U}_i$, respectively. Then, the transmitted signals are,

$$\mathbf{X}_1 = \mathbf{P}_1 \mathbf{V}_1 + \mathbf{Q}_1 \mathbf{U}_1 \tag{8.82}$$

$$\mathbf{X}_2 = \mathbf{P}_2 \mathbf{V}_2 + \mathbf{Q}_2 \mathbf{U}_2 \tag{8.83}$$

The received signals in this case are:

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{H}_{11}\mathbf{P}_1\mathbf{V}_1 + (\mathbf{H}_{11}\mathbf{Q}_1\mathbf{U}_1 + \mathbf{H}_{21}\mathbf{P}_2\mathbf{V}_2) + \mathbf{H}_{21}\mathbf{Q}_2\mathbf{U}_2 + \mathbf{N}_1 \\ \mathbf{Y}_2 &= (\mathbf{H}_{12}\mathbf{P}_1\mathbf{V}_1 + \mathbf{H}_{22}\mathbf{Q}_2\mathbf{U}_2) + \mathbf{H}_{12}\mathbf{Q}_1\mathbf{U}_1 + \mathbf{H}_{22}\mathbf{P}_2\mathbf{V}_2 + \mathbf{N}_2 \end{aligned} \tag{8.84}$$

We choose the precoding matrices $\mathbf{P}_i, \mathbf{Q}_i$ such that they satisfy the following alignment equations

$$\text{span}\{\mathbf{H}_{21}\mathbf{P}_2\} \subseteq \text{span}\{\mathbf{H}_{11}\mathbf{Q}_1\} \tag{8.85}$$

$$\text{span}\{\mathbf{H}_{12}\mathbf{P}_1\} \subseteq \text{span}\{\mathbf{H}_{22}\mathbf{Q}_2\} \tag{8.86}$$

Feasibility of Alignment: The alignment can be achieved by choosing $\mathbf{P}_i, \mathbf{Q}_i$

such that

$$\begin{bmatrix} \mathbf{H}_{11} & -\mathbf{H}_{21} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{P}_2 \end{bmatrix} = \mathbf{0} \tag{8.87}$$

$$\begin{bmatrix} \mathbf{H}_{12} & -\mathbf{H}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{Q}_2 \end{bmatrix} = \mathbf{0} \tag{8.88}$$

i.e., by choosing $\mathbf{P}_i$, $\mathbf{Q}_i$ to be in the null space of the combined channel of the two users. Note that $\begin{bmatrix} \mathbf{H}_{11} & -\mathbf{H}_{21} \end{bmatrix}$ is an $\mathbb{R}^{N \times 2M}$ matrix. Hence, the null space of this matrix, $\begin{bmatrix} \mathbf{H}_{11} & -\mathbf{H}_{21} \end{bmatrix}^{\perp}$, is an $\mathbb{R}^{2M \times 2M-N}$ matrix. Thus, choosing $\mathbf{P}_i$, $\mathbf{Q}_i$ as $\mathbb{R}^{M \times 2M-N}$ is feasible.

Decodability: By this alignment scheme, we have

$$\mathbf{Y}_1 = \mathbf{H}_{11}\mathbf{P}_1\mathbf{V}_1 + \mathbf{H}_{11}\mathbf{Q}_1(\mathbf{U}_1 + \mathbf{V}_2) + \mathbf{H}_{21}\mathbf{Q}_2\mathbf{U}_2 + \mathbf{N}_1 \tag{8.89}$$

$$= \begin{bmatrix} \mathbf{H}_{11}\mathbf{P}_1 & \mathbf{H}_{11}\mathbf{Q}_1 & \mathbf{H}_{21}\mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{U}_1 + \mathbf{V}_2 \\ \mathbf{U}_2 \end{bmatrix} + \mathbf{N}_1 \tag{8.90}$$

Similarly, for receiver 2, we have

$$\mathbf{Y}_2 = \begin{bmatrix} \mathbf{H}_{22}\mathbf{Q}_2 & \mathbf{H}_{12}\mathbf{Q}_1 & \mathbf{H}_{22}\mathbf{P}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1 + \mathbf{U}_2 \\ \mathbf{U}_1 \\ \mathbf{V}_2 \end{bmatrix} + \mathbf{N}_2 \tag{8.91}$$

In order to decode $\mathbf{Y}_i$ using a zero forcing receiver, the total dimensions $3(2M - N)$

should be at most N. This holds true since $\frac{M}{N} \leq \frac{2}{3}$. Thus, we can decode $\mathbf{V}_1$ using zero forcing as

$$\begin{bmatrix} \mathbf{V}_1 \\ \mathbf{U}_1 + \mathbf{V}_2 \\ \mathbf{U}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{11}\mathbf{P}_1 & \mathbf{H}_{11}\mathbf{Q}_1 & \mathbf{H}_{21}\mathbf{Q}_2 \end{bmatrix}^{\dagger} \mathbf{Y}_1 \tag{8.92}$$

where $(.)^{\dagger}$ is the pseudo-inverse of a matrix.

Security: Since each secure signal is aligned with a cooperative jamming signal from the other user, the leakage rate is upper bounded by a constant, and hence the system is secure from the s.d.o.f. perspective, i.e., for user 1 without loss of generality, using Fano's and data processing inequality,

$$R_1 \geq I(\mathbf{V}_1; \mathbf{Y}_1) - I(\mathbf{V}_1; \mathbf{Y}_2|\mathbf{V}_2) \tag{8.93}$$

$$\geq I(\mathbf{V}_1; \hat{\mathbf{V}}_1) - h(\mathbf{Y}_2|\mathbf{V}_2) + h(\mathbf{Y}_2|\mathbf{V}_1, \mathbf{V}_2) \tag{8.94}$$

$$\begin{aligned} &\geq h(\mathbf{V}_1) - h(\mathbf{V}_1|\hat{\mathbf{V}}_1) - h\left(\begin{bmatrix} \mathbf{H}_{22}\mathbf{Q}_2 & \mathbf{H}_{12}\mathbf{Q}_1 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1 + \mathbf{U}_2 \\ \mathbf{U}_1 \end{bmatrix}\right) \\ &\quad + h\left(\begin{bmatrix} \mathbf{H}_{12}\mathbf{Q}_1 & \mathbf{H}_{22}\mathbf{P}_2 \end{bmatrix} \begin{bmatrix} \mathbf{U}_2 \\ \mathbf{U}_1 \end{bmatrix}\right) \end{aligned} \tag{8.95}$$

$$\geq (1 - P_e)(2M - N) \cdot \frac{1}{2} \log P - h\left(\begin{bmatrix} \mathbf{V}_1 + \mathbf{U}_2 \\ \mathbf{U}_1 \end{bmatrix}\right) + h\left(\begin{bmatrix} \mathbf{U}_2 \\ \mathbf{U}_1 \end{bmatrix}\right) + o(\log P) \tag{8.96}$$

$$= (1 - P_e)(2M - N) \cdot \frac{1}{2} \log P - 2(2M - N) \cdot \frac{1}{2} \log P$$

$$+ 2(2M - N) \cdot \frac{2}{2} \log P + o(\log P) \tag{8.97}$$

By dividing by $\frac{1}{2} \log P$ and taking $P \to \infty$, the $P_e \to 0$ and hence $d_1 \geq 2M - N$.

8.7.2 $\frac{2N}{3} \leq M \leq N$

In this regime, we combine the achievable scheme of the previous regime with the achievable scheme of the basic 2×2 system (or the 1×1 SISO system).

Transmitted Signals: Let $\mathbf{V}_i = \begin{bmatrix} \mathbf{V}_i^{(1)} \\ \\ \mathbf{V}_i^{(2)} \end{bmatrix}$ and $\mathbf{U}_i = \begin{bmatrix} \mathbf{U}_i^{(1)} \\ \\ \mathbf{U}_i^{(2)} \end{bmatrix}$. $\mathbf{V}_i^{(1)}, \mathbf{U}_i^{(1)} \sim \mathcal{N}(\mathbf{0}, \eta_2 P \mathbf{I}_{\lfloor \frac{N}{3} \rfloor})$ are Gaussian signals of size $\lfloor \frac{N}{3} \rfloor$. $\mathbf{V}_i^{(1)}, \mathbf{U}_i^{(1)}$ correspond to the part that can be protected using spatial alignment only without any real interference alignment. The vectors $\mathbf{V}_i^{(2)}, \mathbf{U}_i^{(2)}$ are structured signals of size $N \bmod 3$ which is either 1 or 2. $\mathbf{V}_i^{(2)}, \mathbf{U}_i^{(2)}$ are picked from PAM constellation $C(a, Q)$, with proper parameters. This separation effectively reduces the problem into designing spatial alignment precoders as in the previous regime and the basic 2×2 system (or the 1×1 SISO system). We consider the case of $N \bmod 3 = 2$, without loss of generality. Let $\mathbf{P}_i$, $\mathbf{Q}_i$ be precoding matrices in $\mathbb{R}^{M \times (\lfloor \frac{N}{3} \rfloor + N \bmod 3)}$, then the transmitted signals

are

$$\mathbf{X}_1 = \mathbf{P}_1 \begin{bmatrix} v_{1,1}^{(1)} \\ v_{1,2}^{(1)} \\ \vdots \\ v_{1,\lfloor \frac{N}{3} \rfloor}^{(1)} \\ \mathbf{t}_2^T \mathbf{v}_{11}^{(2)} \\ \mathbf{t}_2^T \mathbf{v}_{12}^{(2)} \end{bmatrix} + \mathbf{Q}_1 \begin{bmatrix} u_{1,1}^{(1)} \\ u_{1,2}^{(1)} \\ \vdots \\ u_{1,\lfloor \frac{N}{3} \rfloor}^{(1)} \\ \mathbf{t}_1^T \mathbf{u}_{11}^{(2)} \\ \mathbf{t}_1^T \mathbf{u}_{12}^{(2)} \end{bmatrix} \tag{8.98}$$

$$\mathbf{X}_2 = \mathbf{P}_2 \begin{bmatrix} v_{2,1}^{(1)} \\ v_{2,2}^{(1)} \\ \vdots \\ v_{2,\lfloor \frac{N}{3} \rfloor}^{(1)} \\ \mathbf{t}_1^T \mathbf{v}_{21}^{(2)} \\ \mathbf{t}_1^T \mathbf{v}_{22}^{(2)} \end{bmatrix} + \mathbf{Q}_2 \begin{bmatrix} u_{2,1}^{(1)} \\ u_{2,2}^{(1)} \\ \vdots \\ u_{2,\lfloor \frac{N}{3} \rfloor}^{(1)} \\ \mathbf{t}_2^T \mathbf{u}_{21}^{(2)} \\ \mathbf{t}_2^T \mathbf{u}_{22}^{(2)} \end{bmatrix} \tag{8.99}$$

where $\mathbf{P}_i, \mathbf{Q}_i$ are designed using (8.87), (8.88).

Feasibility of Alignment: Similar to the previous section, this alignment is possible if the null space of the combined channel $\begin{bmatrix} \mathbf{H}_{11} & -\mathbf{H}_{21} \end{bmatrix}^{\perp}$ has dimension $2M - N \geq \lfloor \frac{N}{3} \rfloor + N \bmod 3$, which implies that $\frac{M}{N} \geq \frac{2}{3} + \frac{N \bmod 3}{3N}$. This condition always holds in this regime.

Decodability: Partition $\mathbf{P}_i = \begin{bmatrix} \mathbf{P}_{i_{N \times \lfloor \frac{N}{3} \rfloor}}^{(1)} & \mathbf{P}_{i_{N \times N \bmod 3}}^{(2)} \end{bmatrix}$ and similarly for $\mathbf{Q}_i$.

Then, the received signal at receiver 1 is

$$\mathbf{Y}_1 = \begin{bmatrix} \mathbf{H}_{11}\mathbf{P}_1^{(1)} & \mathbf{H}_{11}\mathbf{Q}_1^{(1)} & \mathbf{H}_{21}\mathbf{Q}_2^{(1)} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^{(1)} \\ \mathbf{V}_2^{(1)} + \mathbf{U}_1^{(1)} \\ \mathbf{U}_2^{(1)} \end{bmatrix} + \begin{bmatrix} \mathbf{H}_{11}\mathbf{P}_1^{(2)} & \mathbf{H}_{11}\mathbf{Q}_1^{(2)} & \mathbf{H}_{21}\mathbf{Q}_2^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^{(2)} \\ \mathbf{V}_2^{(2)} + \mathbf{U}_1^{(2)} \\ \mathbf{U}_2^{(2)} \end{bmatrix} + \mathbf{N}_1 \tag{8.100}$$

Define matrix $\mathbf{F}_1 = \begin{bmatrix} \mathbf{H}_{11}\mathbf{P}_1^{(1)} & \mathbf{H}_{11}\mathbf{Q}_1^{(1)} & \mathbf{H}_{21}\mathbf{Q}_2^{(1)} \end{bmatrix}$ as an $\mathbb{R}^{N \times 3\lfloor \frac{N}{3} \rfloor}$ matrix. We null out the effect of the first components from $\mathbf{Y}_1$ by multiplying by the nulling matrix $\mathbf{Z}_1^T$, which is defined as the right null space of $\mathbf{F}_1$

$$\mathbf{Z}_1 = \left(\mathbf{F}_1^T\right)^{\perp} \tag{8.101}$$

The nulling matrix $\mathbf{Z}_1^T$ is $\mathbb{R}^{N \bmod 3 \times N}$. Then, the filtered observation is

$$\tilde{\mathbf{Y}}_1 = \mathbf{Z}_1^T \mathbf{Y}_1 \tag{8.102}$$

$$= \begin{bmatrix} \mathbf{Z}_1^T\mathbf{H}_{11}\mathbf{P}_1^{(2)} & \mathbf{Z}_1^T\mathbf{H}_{11}\mathbf{Q}_1^{(2)} & \mathbf{Z}_1^T\mathbf{H}_{21}\mathbf{Q}_2^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^{(2)} \\ \mathbf{V}_2^{(2)} + \mathbf{U}_1^{(2)} \\ \mathbf{U}_2^{(2)} \end{bmatrix} + \tilde{\mathbf{N}}_1 \tag{8.103}$$

where $\tilde{\mathbf{N}}_1 = \mathbf{Z}_1^T\mathbf{N}_1$. Orthogonalize $\mathbf{V}_2^{(2)} + \mathbf{U}_1^{(2)}$ components by multiplying by

$(\mathbf{Z}_1^T\mathbf{H}_{11}\mathbf{Q}_1^{(2)})^{-1}$,

$$\tilde{\tilde{\mathbf{Y}}}_1 = (\mathbf{Z}_1^T\mathbf{H}_{11}\mathbf{Q}_1^{(2)})^{-1}\tilde{\mathbf{Y}}_1 \tag{8.104}$$

$$= \mathbf{A}\mathbf{V}_1^{(2)} + (\mathbf{V}_2^{(2)} + \mathbf{U}_1^{(2)}) + \mathbf{B}\mathbf{U}_2^{(2)} + \tilde{\tilde{\mathbf{N}}}_1 \tag{8.105}$$

where $\mathbf{A} = (\mathbf{Z}_1^T\mathbf{H}_{11}\mathbf{Q}_1^{(2)})^{-1}\mathbf{Z}_1^T\mathbf{H}_{11}\mathbf{P}_1^{(2)}$, $\mathbf{B} = (\mathbf{Z}_1^T\mathbf{H}_{11}\mathbf{Q}_1^{(2)})^{-1}\mathbf{Z}_1^T\mathbf{H}_{21}\mathbf{Q}_2^{(2)}$, and $\tilde{\tilde{\mathbf{N}}}_1 = (\mathbf{Z}_1^T\mathbf{H}_{11}\mathbf{Q}_1^{(2)})^{-1}\tilde{\mathbf{N}}_1$. $\mathbf{A}$, $\mathbf{B}$ are now $N \bmod 3 \times N \bmod 3$ matrices. By designing $\mathbf{t}_i$ as in the 2×2 system, $\mathbf{V}_1^{(2)}, \mathbf{U}_2^{(2)}$ are decoded without error. Cancelling them from $\mathbf{Y}_1$, we have

$$\bar{\mathbf{Y}}_1 = \begin{bmatrix} \mathbf{H}_{11}\mathbf{P}_1^{(1)} & \mathbf{H}_{11}\mathbf{Q}_1^{(1)} & \mathbf{H}_{21}\mathbf{Q}_2^{(1)} & \mathbf{H}_{11}\mathbf{Q}_1^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^{(1)} \\ \mathbf{V}_2^{(1)} + \mathbf{U}_1^{(1)} \\ \mathbf{U}_2^{(1)} \\ \mathbf{V}_2^{(2)} + \mathbf{U}_1^{(2)} \end{bmatrix} + \mathbf{N}_1 \tag{8.106}$$

To check the decodability, the total number of dimensions is $3\lfloor \frac{N}{3} \rfloor + N \bmod 3 = N$, and hence signals are decodable by a zero-forcing receiver as in the previous section.

Security: Similar to the analysis in the previous sections, the secure signals of user 2 at receiver 1 are exactly aligned with the cooperative jamming signals of user 1. Consequently, the leakage rate is bounded by a constant, and each user achieves an s.d.o.f. of $\lfloor \frac{N}{3} \rfloor + \frac{N \bmod 3}{3} = \frac{N}{3}$ with a total s.d.o.f. $d_s \geq \frac{2N}{3}$.

8.7.3 $N \leq M \leq 2N$

In this regime, we note the availability of a null space for each cross-channel matrix. Consequently, the achievable scheme combines null space transmission with the achievable scheme of the square system $M = N$, which includes spatial and asymptotic real interference alignment. The upper bound suggests that each user sends $M - N$ signals in the null space of the other user, so that they become invisible, and use the rest of the antennas as a square system of dimension $2N - M$ (recall that $d_s \leq \frac{2(2M-N)}{3} = 2(M-N) + 2\frac{2N-M}{3}$). To separate the square system components from contaminating the null space components, we further precode the signals of the square system.

Transmitted Signals: Generate $\mathbf{V}_{i0} \sim \mathcal{N}(\mathbf{0}, \eta_3 P \mathbf{I}_{M-N})$ as Gaussian secure signals of size $M-N$ that are transmitted through the null space of the cross-channel to the jth receiver. $\mathbf{V}_i, \mathbf{U}_i \sim \mathcal{N}(\mathbf{0}, \eta_4 P \mathbf{I}_{\lfloor \frac{2N-M}{3} \rfloor})$ are Gaussian secure signals and Gaussian cooperative jamming signals, respectively, both of size $\lfloor \frac{2N-M}{3} \rfloor$. $\mathbf{v}_{ij}, \mathbf{u}_{ij}$ are structured PAM signals weighted with vectors $\mathbf{t}_i$, which will be defined later. Let $\mathbf{H}_{11}^{(1)}$, $\mathbf{H}_{12}^{(1)}$, $\mathbf{H}_{22}^{(1)}$, $\mathbf{H}_{21}^{(1)}$ are the $\mathbb{R}^{(M-N)\times M}$ channel matrices to the first $M - N$

antennas at the receivers. Therefore, the transmitted signals are

$$\mathbf{X}_1 = \mathbf{H}_{12}^{\perp}\mathbf{V}_{10} + \begin{bmatrix}\mathbf{H}_{11}^{(1)}\\ \mathbf{H}_{12}^{(1)}\end{bmatrix}^{\perp}\left(\mathbf{P}_1\begin{bmatrix} v_{1,1}^{(1)}\\ v_{1,2}^{(1)}\\ \vdots\\ v_{1,\lfloor\frac{\tilde{N}}{3}\rfloor}^{(1)}\\ \mathbf{t}_2^T\mathbf{v}_{11}^{(2)}\\ \mathbf{t}_2^T\mathbf{v}_{12}^{(2)}\end{bmatrix} + \mathbf{Q}_1\begin{bmatrix} u_{1,1}^{(1)}\\ u_{1,2}^{(1)}\\ \vdots\\ u_{1,\lfloor\frac{\tilde{N}}{3}\rfloor}^{(1)}\\ \mathbf{t}_1^T\mathbf{u}_{11}^{(2)}\\ \mathbf{t}_1^T\mathbf{u}_{12}^{(2)}\end{bmatrix}\right) \tag{8.107}$$

$$\mathbf{X}_2 = \mathbf{H}_{21}^{\perp}\mathbf{V}_{20} + \begin{bmatrix}\mathbf{H}_{21}^{(1)}\\ \mathbf{H}_{22}^{(1)}\end{bmatrix}^{\perp}\left(\mathbf{P}_2\begin{bmatrix} v_{2,1}^{(1)}\\ v_{2,2}^{(1)}\\ \vdots\\ v_{2,\lfloor\frac{\tilde{N}}{3}\rfloor}^{(1)}\\ \mathbf{t}_1^T\mathbf{v}_{21}^{(2)}\\ \mathbf{t}_1^T\mathbf{v}_{22}^{(2)}\end{bmatrix} + \mathbf{Q}_2\begin{bmatrix} u_{2,1}^{(1)}\\ u_{2,2}^{(1)}\\ \vdots\\ u_{2,\lfloor\frac{\tilde{N}}{3}\rfloor}^{(1)}\\ \mathbf{t}_2^T\mathbf{u}_{21}^{(2)}\\ \mathbf{t}_2^T\mathbf{u}_{22}^{(2)}\end{bmatrix}\right) \tag{8.108}$$

where $\tilde{N} = 2N - M$. This precoding separates the first $M - N$ antennas at each receiver from the square system signals. This leaves the $\mathbf{V}_{i0}$ vectors to be reliably received via zero-forcing processing.

Decodability and Security: We focus on receiver 1 without loss of generality. $\mathbf{H}_{11}\begin{bmatrix}\mathbf{H}_{11}^{(1)}\\ \mathbf{H}_{12}^{(1)}\end{bmatrix}^{\perp}$ has the dimension of $N \times (2N - M)$. Ignoring the first $M - N$ antennas at the receiver, the remaining system is $(2N - M) \times (2N - M)$, which is a square system as presented in the previous section. By considering the first $M - N$ antennas, $\mathbf{Y}_1^{(1)} = \mathbf{H}_{11}\mathbf{H}_{12}^{\perp}\mathbf{V}_{10} + \mathbf{N}_1^{(1)}$. Consequently, we can decode

$\hat{\mathbf{V}}_{10} = (\mathbf{H}_{11}\mathbf{H}_{12}^{\perp})^{\dagger}\mathbf{Y}_1^{(1)}$. Cancelling $\mathbf{V}_{10}$ from $\mathbf{Y}_1$, we are left with a square system only. Note that the dimensions set and spatial alignment matrices can be constructed in a similar manner by defining $\bar{\mathbf{H}}_{11} = \mathbf{H}_{11}^{(2)} \begin{bmatrix} \mathbf{H}_{11}^{(1)} \\ \mathbf{H}_{12}^{(1)} \end{bmatrix}^{\perp}$, and similarly, $\bar{\mathbf{H}}_{21} = \mathbf{H}_{21}^{(2)} \begin{bmatrix} \mathbf{H}_{21}^{(1)} \\ \mathbf{H}_{22}^{(1)} \end{bmatrix}^{\perp}$, $\bar{\mathbf{H}}_{12} = \mathbf{H}_{12}^{(2)} \begin{bmatrix} \mathbf{H}_{11}^{(1)} \\ \mathbf{H}_{12}^{(1)} \end{bmatrix}^{\perp}$ and $\bar{\mathbf{H}}_{22} = \mathbf{H}_{22}^{(2)} \begin{bmatrix} \mathbf{H}_{21}^{(1)} \\ \mathbf{H}_{22}^{(1)} \end{bmatrix}^{\perp}$. Then the spatial alignment matrices are designed such that

$$\begin{bmatrix} \bar{\mathbf{H}}_{11} & -\bar{\mathbf{H}}_{21} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{P}_2 \end{bmatrix} = \mathbf{0} \tag{8.109}$$

$$\begin{bmatrix} \bar{\mathbf{H}}_{12} & -\bar{\mathbf{H}}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{Q}_2 \end{bmatrix} = \mathbf{0} \tag{8.110}$$

We can now define the dimensions sets on $\bar{\mathbf{H}}_{11}, \bar{\mathbf{H}}_{12}, \bar{\mathbf{H}}_{21}, \bar{\mathbf{H}}_{22}$ as in the previous section. Thus, the alignment process, and the secrecy analysis remain the same as the square system analysis.

8.7.4 $M \geq 2N$

In this case, since $M \geq 2N$, each cross-channel $\mathbf{H}_{12}$, $\mathbf{H}_{21}$ has $M - N$ null space components. Since $M - N \geq N$, each user transmits N secure Gaussian signal components in the null space of the other receiver's channel only. Let $\mathbf{V}_i = [v_{i_1}\, v_{i_2} \ldots v_{i_N} \quad \mathbf{0}_{M-2N}^T]^T = [\bar{\mathbf{V}}_i \quad \mathbf{0}_{M-2N}^T]^T$ be the transmitted Gaussian signal

for user i, where $\bar{\mathbf{V}}_i \sim \mathcal{N}(\mathbf{0}, \eta_4 P \mathbf{I}_N)$. Thus, the transmitted signals are:

$$\mathbf{X}_1 = \mathbf{H}_{12}^{\perp} \mathbf{V}_1 \tag{8.111}$$

$$\mathbf{X}_2 = \mathbf{H}_{21}^{\perp} \mathbf{V}_2 \tag{8.112}$$

Since the channel gains are drawn from a continuous distribution, $\mathbf{H}_{11}\mathbf{H}_{12}^{\perp}$ and $\mathbf{H}_{22}\mathbf{H}_{21}^{\perp}$ are full rank almost surely. Hence, the receiver performs zero-forcing to decode $\mathbf{V}_i$, i.e.,

$$\hat{\mathbf{V}}_1 = (\mathbf{H}_{11}\mathbf{H}_{12}^{\perp})^{\dagger} \mathbf{Y}_1 \tag{8.113}$$

$$\hat{\mathbf{V}}_2 = (\mathbf{H}_{22}\mathbf{H}_{21}^{\perp})^{\dagger} \mathbf{Y}_2 \tag{8.114}$$

At high SNR, the probability of error can be made arbitrarily small. These signals are invisible to the other receiver, i.e., transmitted in perfect security.

8.8 The Entire s.d.o.f. Region for the $M \times N$ ICCM

In this section, we derive the optimal achievable schemes for the entire region of the $M \times N$ ICCM. From the converse proof, we note that the s.d.o.f. region is a four-vertex polytope for any M, N. The non-trivial points of the polytope are the sum s.d.o.f. point and the two symmetric maximum individual s.d.o.f. points. Thus, in this section, we concentrate on characterizing achievable schemes for one of the maximum individual s.d.o.f. points only. Since the s.d.o.f. region is naturally a square region for $\frac{N}{2} \leq M \leq \frac{2N}{3}$ and $M \geq 2N$, the problem of characterizing

the entire s.d.o.f. region reduces to finding the optimal achievable schemes for the maximum individual s.d.o.f. points for $\frac{2N}{3} \leq M \leq 2N$. Any other point that belongs to the s.d.o.f. region can be achieved by time-sharing. In the following, we consider the achievability of the $(d_m, 0)$ point, without loss of generality. We present these schemes in a concise way because these points can be mapped to the achievable schemes of the MIMO wiretap channel with a helper in [79]. The idea in all these schemes is to let user 2 sacrifice his own s.d.o.f. and send only cooperative jamming signals to jam its own receiver, i.e., it acts as a pure helper.

8.8.1 For $\frac{2N}{3} \leq M \leq \frac{3N}{4}$

In this case user 1 sends Gaussian secure signal $\mathbf{V}_1$ of dimension $2M - N$, while user 2 sends pure Gaussian cooperative jamming signal $\mathbf{U}_1$ of the same dimension, i.e., the transmitted signals are

$$\mathbf{X}_1 = \mathbf{P}_1 \mathbf{V}_1 \tag{8.115}$$

$$\mathbf{X}_2 = \mathbf{Q}_2 \mathbf{U}_2 \tag{8.116}$$

Then, the received signals are

$$\mathbf{Y}_1 = \mathbf{H}_{11}\mathbf{P}_1\mathbf{V}_1 + \mathbf{H}_{21}\mathbf{Q}_2\mathbf{U}_2 + \mathbf{N}_1 \tag{8.117}$$

$$\mathbf{Y}_2 = \mathbf{H}_{12}\mathbf{P}_1\mathbf{V}_1 + \mathbf{H}_{22}\mathbf{Q}_2\mathbf{U}_2 + \mathbf{N}_2 \tag{8.118}$$

To ensure security, we align the cooperative jamming of user 2 with the secure signal of user 1 at receiver 2 by designing $\mathbf{P}_1$, $\mathbf{Q}_2$ such that

$$\begin{bmatrix} \mathbf{H}_{12} & -\mathbf{H}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{P}_1 \\ \\ \mathbf{Q}_2 \end{bmatrix} = \mathbf{0} \tag{8.119}$$

Since the null space of the matrix $\begin{bmatrix} \mathbf{H}_{12} & -\mathbf{H}_{22} \end{bmatrix}$ has dimension $2M \times (2M - N)$, this alignment is feasible. The decodability is performed via zero-forcing, if the total dimensions $2(2M - N) \leq N$, which is true for $M \leq \frac{3N}{4}$. The leakage rate is upper bounded by a constant as shown in previous sections, hence, the scheme is secure in the s.d.o.f. sense. Consequently, $d_m = 2M - N$ is achievable.

8.8.2 For $\frac{3N}{4} \leq M \leq N$

We combine the previous achievable scheme with the exact real interference alignment. The signals compose of Gaussian components $\mathbf{V}_1^{(1)}, \mathbf{U}_2^{(1)}$ of dimension $\lfloor \frac{N}{2} \rfloor$ and structured components $v_1^{(2)}, u_2^{(2)}$ of dimension $N \bmod 2 = 0$ or 1 picked from PAM constellation $C(a, Q)$ with $Q = P^{\frac{1-\delta}{2(2+\delta)}}$, $a = \gamma \frac{P^{\frac{1}{2}}}{Q}$. The transmitted signals are

$$\mathbf{X}_1 = \mathbf{P}_1 \begin{bmatrix} \mathbf{V}_1^{(1)} \\ \\ v_1^{(2)} \end{bmatrix}, \qquad \mathbf{X}_2 = \mathbf{Q}_2 \begin{bmatrix} \mathbf{U}_2^{(1)} \\ \\ u_2^{(2)} \end{bmatrix} \tag{8.120}$$

Note that the PAM component is ignored if N is even. By using the same $\mathbf{P}_1, \mathbf{Q}_2$ as in the previous section, the transmission is secure from user 2. This alignment is feasible, because the null space dimension $2M - N \geq \lfloor \frac{N}{2} \rfloor + N \bmod 2$ for an integer

M that satisfies $\frac{3N}{4} \leq M \leq N$. The received signal at receiver 1 is

$$\mathbf{Y}_1 = \begin{bmatrix} \mathbf{H}_{11}\mathbf{P}_1^{(1)} & \mathbf{H}_{21}\mathbf{Q}_2^{(1)} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^{(1)} \\ \mathbf{U}_2^{(1)} \end{bmatrix} + \mathbf{H}_{11}\mathbf{P}_1^{(2)}v_1^{(2)} + \mathbf{H}_{21}\mathbf{Q}_2^{(2)}u_2^{(2)} + \mathbf{N}_1 \tag{8.121}$$

where $\mathbf{P}_1 = [\mathbf{P}^{(1)}_{1_{M\times\lfloor\frac{N}{2}\rfloor}} \quad \mathbf{P}^{(2)}_{1_{M\times N \bmod 2}}]$ and similarly for $\mathbf{Q}_2$. By defining $\mathbf{F}_1 = \begin{bmatrix} \mathbf{H}_{11}\mathbf{P}_1^{(1)} & \mathbf{H}_{21}\mathbf{Q}_2^{(1)} \end{bmatrix}$ and multiplying by the right null space of this matrix, i.e., by multiplying by $\mathbf{Z}_1^T = \left(\left(\mathbf{F}_1^T\right)^{\perp}\right)^T$, we have

$$\tilde{y}_1 = \mathbf{Z}_1^T\mathbf{H}_{11}\mathbf{P}_1^{(2)}v_1^{(2)} + \mathbf{Z}_1^T\mathbf{H}_{21}\mathbf{Q}_2^{(2)}u_2^{(2)} + \tilde{\mathbf{N}}_1 \tag{8.122}$$

where $\tilde{\mathbf{N}}_1 = \mathbf{Z}^T\mathbf{N}_1$. Note that this is a SISO system. Therefore, with the choice of $Q = P^{\frac{1-\delta}{2(2+\delta)}}, a = \gamma\frac{P^{\frac{1}{2}}}{Q}$, the $v_1^{(2)}, u_2^{(2)}$, signals are both decodable, because they lie in rationally independent dimensions almost surely. By cancelling these components from $\mathbf{Y}_1$, we are left with $2\lfloor\frac{N}{2}\rfloor \leq N$ signals, which can be decoded using a zero-forcing receiver. Consequently, $d_m = \frac{N}{2}$ is achievable.

8.8.3 For $N \leq M \leq \frac{3N}{2}$

In this case, user 1 can exploit the null space of $\mathbf{H}_{12}$ to send $M-N$ Gaussian secure signal components. Similarly, user 2 can generate $M-N$ Gaussian cooperative jamming components that are invisible to receiver 1 if transmitted in the null space of $\mathbf{H}_{21}$. Therefore, user 1 sends four signal components, $\mathbf{V}_{10}$ is the Gaussian secure signal that can be transmitted in the null space of dimension $M-N$, $\mathbf{V}_{11}$ is the Gaussian secure signal that can be protected using the invisible cooperative jamming

components of user 2 of dimension $M-N$, $\mathbf{V}_{12}^{(1)}$ is Gaussian secure signal of dimension $\lfloor \frac{3N-2M}{2} \rfloor$ and $v_{12}^{(2)}$ which is a structured secure signal of dimension $N \bmod 2$ picked from PAM constellation $C(a, Q)$ with the same parameters as in the previous section. Similarly, user 2 transmits cooperative jamming signal into $\mathbf{U}_{11}$ of dimension $M-N$ and is sent in the null space of receiver 1, $\mathbf{U}_{12}^{(1)}$ Gaussian of dimension $\lfloor \frac{3N-2M}{2} \rfloor$ and a structured component $u_{12}^{(2)}$ of dimension $N \bmod 2$. The transmitted signals are,

$$\mathbf{X}_1 = \mathbf{H}_{12}^{\perp}\mathbf{V}_{10} + \mathbf{P}_{11}\mathbf{V}_{11} + \mathbf{P}_{12}\begin{bmatrix}\mathbf{V}_{12}^{(1)} \\ v_{12}^{(2)}\end{bmatrix} \tag{8.123}$$

$$\mathbf{X}_2 = \mathbf{H}_{21}^{\perp}\mathbf{Q}_{21}\mathbf{U}_{11} + \mathbf{Q}_{22}\begin{bmatrix}\mathbf{U}_{12}^{(1)} \\ u_{12}^{(2)}\end{bmatrix} \tag{8.124}$$

By forcing, $\mathbf{H}_{12}\mathbf{P}_{11} = \mathbf{H}_{22}\mathbf{H}_{21}^{\perp}\mathbf{Q}_{21}$ and $\mathbf{H}_{12}\mathbf{P}_{12} = \mathbf{H}_{22}\mathbf{Q}_{22}$, the scheme is secure in the s.d.o.f. sense. Then, the received signals are,

$$\mathbf{Y}_1 = \mathbf{H}_{11}\mathbf{H}_{12}^{\perp}\mathbf{V}_{10} + \mathbf{H}_{11}\mathbf{P}_{11}\mathbf{V}_{11} + \mathbf{H}_{11}\mathbf{P}_{12}\begin{bmatrix}\mathbf{V}_{12}^{(1)} \\ v_{12}^{(2)}\end{bmatrix} + \mathbf{H}_{21}\mathbf{Q}_{22}\begin{bmatrix}\mathbf{U}_{12}^{(1)} \\ u_{12}^{(2)}\end{bmatrix} + \mathbf{N}_1 \tag{8.125}$$

$$\mathbf{Y}_2 = \mathbf{H}_{22}\mathbf{H}_{21}^{\perp}\mathbf{Q}_{21}(\mathbf{V}_{11} + \mathbf{U}_{11}) + \mathbf{H}_{22}\mathbf{Q}_{22}\begin{bmatrix}\mathbf{V}_{12}^{(1)} + \mathbf{U}_{12}^{(1)} \\ v_{12}^{(2)} + u_{12}^{(2)}\end{bmatrix} + \mathbf{N}_2 \tag{8.126}$$

This alignment can be designed by choosing $\mathbf{P}_{1i}, \mathbf{Q}_{2i},\ i = 1, 2$ such that

$$\begin{bmatrix}\mathbf{H}_{12} & -\mathbf{H}_{22}\mathbf{H}_{21}^{\perp}\end{bmatrix}\begin{bmatrix}\mathbf{P}_{11} \\ \mathbf{Q}_{21}\end{bmatrix} = \mathbf{0} \tag{8.127}$$

$$\begin{bmatrix}\mathbf{H}_{12} & -\mathbf{H}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{P}_{12}\\ \\ \mathbf{Q}_{22}\end{bmatrix}=\mathbf{0} \tag{8.128}$$

This alignment is feasible if the null space of $\begin{bmatrix}\mathbf{H}_{12} & -\mathbf{H}_{22}\mathbf{H}_{21}^{\perp}\end{bmatrix}$, which has a dimension of $2(M-N)$ is at least accommodating $\mathbf{V}_{11}$ of $M-N$ dimension, i.e., $2(M-N)\geq M-N$. The null space of $\begin{bmatrix}\mathbf{H}_{21} & -\mathbf{H}_{22}\end{bmatrix}$ is also accommodating $\mathbf{V}_{12}$, i.e., $2M-N\geq\lfloor\frac{3N-2M}{2}\rfloor+N \bmod 2$. Both conditions hold true if $M\geq N$. Considering the decodability, we write the received signal at receiver 1 as

$$\mathbf{Y}_1=\begin{bmatrix}\mathbf{H}_{11}\mathbf{H}_{12}^{\perp} & \mathbf{H}_{11}\mathbf{P}_{11} & \mathbf{H}_{11}\mathbf{P}_{12}^{(1)} & \mathbf{H}_{21}\mathbf{Q}_{22}^{(2)}\end{bmatrix}\begin{bmatrix}\mathbf{V}_{10}\\ \\ \mathbf{V}_{11}\\ \\ \mathbf{V}_{12}^{(1)}\\ \\ \mathbf{U}_{12}^{(1)}\end{bmatrix}+\mathbf{H}_{11}\mathbf{P}_{12}^{(2)}v_{12}^{(2)}+\mathbf{H}_{21}\mathbf{Q}_{22}^{(2)}u_{12}^{(2)}+\mathbf{N}_1 \tag{8.129}$$

By defining $\mathbf{F}_1=\begin{bmatrix}\mathbf{H}_{11}\mathbf{H}_{12}^{\perp} & \mathbf{H}_{11}\mathbf{P}_{11} & \mathbf{H}_{11}\mathbf{P}_{12}^{(1)} & \mathbf{H}_{21}\mathbf{Q}_{22}^{(2)}\end{bmatrix}$, we can null out the effect of its symbols by multiplying with the right null space of $\mathbf{F}_1$. This is feasible because $\mathbf{F}_1\in\mathbb{R}^{N\times 2(M-N)+2\lfloor\frac{3N-2M}{2}\rfloor}$ which has right null space of dimension $N \bmod 2\times N$. Hence, we are left with $\mathbf{Y}_1^{(2)}=\mathbf{Z}_1^T(\mathbf{H}_{11}\mathbf{P}_{12}^{(2)}v_{12}^{(2)}+\mathbf{H}_{21}\mathbf{Q}_{22}^{(2)}u_{12}^{(2)})+\tilde{n}_1$. Since, $v_{12}^{(2)},u_{12}^{(2)}$ are picked from structured signals with proper a,Q, these signals are decodable. By cancelling these signals from $\mathbf{Y}_1$ and applying zero-forcing, the rest of the components are also decodable. Consequently, user 1 can transmit $2(M-N)+\lfloor\frac{3N-2M}{2}\rfloor+N \bmod 2$ secure signal components and hence $d_m=\frac{2M-N}{2}$ is achievable.

8.8.4 For $\frac{3N}{2} \leq M \leq 2N$

In this case, user 1 can send $\mathbf{V}_{10}$ Gaussian secure signal of dimension $M-N$ in the null space of cross channel to receiver 2. Therefore, these components are invisible at receiver 2, i.e., perfectly secure. User 2 can send $\mathbf{U}_2$ Gaussian cooperative jamming signals of size $2N-M$ in the null space of the cross channel to receiver 1. These signals ensure the security of signals $\mathbf{V}_1$ of user 1 and at the same time are invisible to receiver 1, and hence leave the space for decodability of the secure signals only. The transmitted signals are

$$\mathbf{X}_1 = \mathbf{H}_{12}^{\perp}\mathbf{V}_{10} + \mathbf{P}_1\mathbf{V}_1 \tag{8.130}$$

$$\mathbf{X}_2 = \mathbf{H}_{21}^{\perp}\mathbf{Q}_2\mathbf{U}_2 \tag{8.131}$$

The received signals are

$$\mathbf{Y}_1 = \mathbf{H}_{11}\mathbf{H}_{12}^{\perp}\mathbf{V}_{10} + \mathbf{H}_{11}\mathbf{P}_1\mathbf{V}_1 + \mathbf{N}_1 \tag{8.132}$$

$$\mathbf{Y}_2 = \mathbf{H}_{12}\mathbf{P}_1\mathbf{V}_1 + \mathbf{H}_{22}\mathbf{H}_{21}^{\perp}\mathbf{Q}_2\mathbf{U}_2 + \mathbf{N}_2 \tag{8.133}$$

By designing $\mathbf{P}_1 \in \mathbb{R}^{M\times 2N-M}$, $\mathbf{Q}_2 \in \mathbb{R}^{M-N\times 2N-M}$ such that

$$\begin{bmatrix} \mathbf{H}_{12} & -\mathbf{H}_{22}\mathbf{H}_{21}^{\perp} \end{bmatrix} \begin{bmatrix} \mathbf{P}_1 \\ \\ \mathbf{Q}_2 \end{bmatrix} = \mathbf{0} \tag{8.134}$$

The leakage rate of $\mathbf{V}_1$ can be upper bounded by a constant, and hence, secure

in the s.d.o.f. sense. This alignment is feasible as long as the null space of user 2 dimension $M-N$ is at least as the dimension of $\mathbf{V}_1$, i.e., $M-N \geq 2N-M$. We have also the condition from the precoder design that the null space of $\begin{bmatrix}\mathbf{H}_{12} & -\mathbf{H}_{22}\mathbf{H}_{21}^{\perp}\end{bmatrix}$ columns (which is $2(M-N)$) should be larger than $2N-M$ which both hold true if $M \geq \frac{3N}{2}$. For the decodability, the number of dimensions at receiver 1 is N and hence decodable using a zero-forcing receiver. Consequently, $d_m = N$ is achievable.

8.9 Special Case: Time-Varying $M \times N$ ICCM

In this section, we consider the special case of time-varying channels. The converse proofs do not change if we change the setting to time-varying channels. Any achievable scheme for the static channel is a valid achievable scheme for the time-varying setting. However, we can use channel variations to simplify the achievable schemes via symbol extension as in [104]. By symbol repetition and coding across multiple channel uses, we can obtain fractional s.d.o.f. in a simpler way. The symbol extension (repetition) replaces the complex real interference alignment (exact or asymptotic) with simplified encoding and decoding schemes. Since symbol extension is used to replace real interference alignment in regimes that have fractional s.d.o.f., it suffices to develop achievable schemes for the sum s.d.o.f. point in the $\frac{2N}{3} \leq M \leq 2N$ regime and the maximum individual s.d.o.f. point for the $\frac{3N}{4} \leq M \leq \frac{3N}{2}$ regime because the remaining points achieve integer s.d.o.f. and do not use real interference alignment for achievability.

8.9.1 Sum s.d.o.f. Point for $\frac{2N}{3} \leq M \leq N$

The users send N secure signal components over 3 channel uses. We call the secure signal $\mathbf{V}_i^{(1)}(t)$ as *time-varying*, if for every channel use $t = 1, 2, 3$, $\mathbf{V}_i^{(1)}(t)$ takes an independent realization from an underlying Gaussian distribution. We call the secure signal $\mathbf{V}_i^{(2)}$ as *fixed*, if for every channel use t, the same realization is transmitted (repeated), i.e., $\mathbf{V}_i^{(2)}(t) = \mathbf{V}_i^{(2)}$, $t = 1, 2, 3$. Each user divides its transmitted secure signals $\mathbf{V}_i$ into two parts. The first part is time-varying $\mathbf{V}_i^{(1)}(t)$, which is a Gaussian vector of dimension $\lfloor \frac{N}{3} \rfloor$. This vector takes new symbols at each channel use t. The second part $\mathbf{V}_i^{(2)}$ is fixed, which is a Gaussian vector of dimension $N \bmod 3$. This vector is repeated over channel uses $t = 1, 2, 3$. Similarly, each user sends cooperative jamming signal $\mathbf{U}_i$ with the same structure. The transmitted signals are

$$\mathbf{X}_1(t) = \mathbf{P}_1(t) \begin{bmatrix} \mathbf{V}_1^{(1)}(t) \\ \mathbf{V}_1^{(2)} \end{bmatrix} + \mathbf{Q}_1(t) \begin{bmatrix} \mathbf{U}_1^{(1)}(t) \\ \mathbf{U}_1^{(2)} \end{bmatrix} \tag{8.135}$$

$$\mathbf{X}_2(t) = \mathbf{P}_2(t) \begin{bmatrix} \mathbf{V}_2^{(1)}(t) \\ \mathbf{V}_2^{(2)} \end{bmatrix} + \mathbf{Q}_2(t) \begin{bmatrix} \mathbf{U}_2^{(1)}(t) \\ \mathbf{U}_2^{(2)} \end{bmatrix} \tag{8.136}$$

where $t = 1, 2, 3$. The precoding matrices vary with t and are designed for every t such that

$$\begin{bmatrix} \mathbf{H}_{11}(t) & -\mathbf{H}_{21}(t) \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1(t) \\ \mathbf{P}_2(t) \end{bmatrix} = \mathbf{0} \tag{8.137}$$

$$\begin{bmatrix} \mathbf{H}_{12}(t) & -\mathbf{H}_{22}(t) \end{bmatrix} \begin{bmatrix} \mathbf{P}_1(t) \\ \mathbf{Q}_2(t) \end{bmatrix} = \mathbf{0} \tag{8.138}$$

This alignment is feasible, since $2M - N \geq \lfloor \frac{N}{3} \rfloor + N \bmod 3$ in this regime. Then, the received signal at receiver 1 in this case is

$$\begin{aligned} \mathbf{Y}_1(t) =& \mathbf{H}_{11}(t)\mathbf{P}_1(t) \begin{bmatrix} \mathbf{V}_1^{(1)}(t) \\ \mathbf{V}_1^{(2)} \end{bmatrix} + \mathbf{H}_{11}(t)\mathbf{Q}_1(t) \begin{bmatrix} \mathbf{U}_1^{(1)}(t) + \mathbf{V}_2^{(1)}(t) \\ \mathbf{U}_1^{(2)} + \mathbf{V}_2^{(2)} \end{bmatrix} \\ &+ \mathbf{H}_{21}(t)\mathbf{Q}_2(t) \begin{bmatrix} \mathbf{U}_2^{(1)}(t) \\ \mathbf{U}_2^{(2)} \end{bmatrix} + \mathbf{N}_1(t) \end{aligned} \tag{8.139}$$

for $t = 1, 2, 3$. By observing $\mathbf{Y}_1(t)$ over the 3 channel uses we can form a linear system with $3N$ unknowns and $3N$ equations. The unknowns are $\mathbf{V}_1^{(1)}(t)$, $\mathbf{U}_1^{(1)}(t) + \mathbf{V}_1^{(1)}(t)$, $\mathbf{U}_2^{(1)}(t)$, $t = 1, 2, 3$, each of dimension $3\lfloor \frac{N}{3} \rfloor$ over the 3 channel uses. $\mathbf{V}_1^{(2)}$, $\mathbf{U}_1^{(2)} + \mathbf{V}_2^{(2)}$, $\mathbf{U}_2^{(2)}$ of dimension of $N \bmod 3$ each. Hence, the total number of unknowns are $3\left(3\lfloor \frac{N}{3} \rfloor + N \bmod 3\right) = 3N$. Since receiver has N antennas, and realizations of channels are independently time-varying, the receiver has $3N$ independent observations almost surely over the 3 channel uses. Using zero-forcing we can decode these unknowns with arbitrarily small probability of error. Each secure signal component of user 2 is aligned with one cooperative jamming signal component from user 1, hence the scheme is secure. Now, since each user transmits $3\lfloor \frac{N}{3} \rfloor + N \bmod 3$ over 3 channel uses, $d_s = \frac{2N}{3}$ is achievable.

8.9.2 Sum s.d.o.f. Point for $N \leq M \leq 2N$

In this case, the users make use of the null spaces to send their time-varying secure signals. Specifically, each user transmits a time-varying $\mathbf{V}_{i0}(t)$, $t = 1, 2, 3$ Gaussian secure signal of dimension $M - N$. Each user transmits a fixed $\mathbf{V}_i$ Gaussian secure signal of dimension $2N - M$, and a fixed $\mathbf{U}_i$ Gaussian cooperative jamming signal. We restrict the first $M-N$ antennas at the receiver for decoding of the time-varying symbols and the rest of the antennas for the fixed symbols. To do this restriction, we precode the transmitted signals as

$$\mathbf{X}_1(t) = \mathbf{H}_{12}(t)^{\perp}\mathbf{V}_{10}(t) + \begin{bmatrix} \mathbf{H}_{11}^{(1)}(t) \\ \mathbf{H}_{12}^{(1)}(t) \end{bmatrix}^{\perp} (\mathbf{P}_1(t)\mathbf{V}_1 + \mathbf{Q}_1(t)\mathbf{U}_1) \tag{8.140}$$

$$\mathbf{X}_2(t) = \mathbf{H}_{21}(t)^{\perp}\mathbf{V}_{20}(t) + \begin{bmatrix} \mathbf{H}_{21}^{(1)}(t) \\ \mathbf{H}_{22}^{(1)}(t) \end{bmatrix}^{\perp} (\mathbf{P}_2(t)\mathbf{V}_2 + \mathbf{Q}_2(t)\mathbf{U}_2) \tag{8.141}$$

where $\mathbf{H}_{ij}^{(1)}(t)$, $\mathbf{H}_{ij}^{(2)}(t)$ correspond to the channel matrix from the ith user to the first $M - N$ antennas, and the rest of the $2N - M$ antennas at receiver j, respectively. Focusing on the first $M - N$ antennas of user 1, without loss of generality. The $\mathbf{Y}_1^{(1)}(t) = \mathbf{H}_{11}^{(1)}(t)\mathbf{H}_{12}(t)^{\perp}\mathbf{V}_{10}(t) + \mathbf{N}_1^{(1)}(t)$. Then, using zero-forcing, the signal $\mathbf{V}_{10}(t)$, $t = 1, 2, 3$ is decodable. After decoding $\mathbf{V}_{10}(t)$, we cancel it from $\mathbf{Y}_1(t)$. By defining $\bar{\mathbf{H}}_{11}(t) = \mathbf{H}_{11}^{(2)}(t) \begin{bmatrix} \mathbf{H}_{11}^{(1)}(t) \\ \mathbf{H}_{12}^{(1)}(t) \end{bmatrix}^{\perp}$, and similarly, $\bar{\mathbf{H}}_{21}(t) = \mathbf{H}_{21}^{(2)}(t) \begin{bmatrix} \mathbf{H}_{21}^{(1)}(t) \\ \mathbf{H}_{22}^{(1)}(t) \end{bmatrix}^{\perp}$,

$\bar{\mathbf{H}}_{12}(t) = \mathbf{H}_{12}^{(2)}(t) \begin{bmatrix} \mathbf{H}_{11}^{(1)}(t) \\ \mathbf{H}_{12}^{(1)}(t) \end{bmatrix}^{\perp}$ and $\bar{\mathbf{H}}_{22}(t) = \mathbf{H}_{22}^{(2)}(t) \begin{bmatrix} \mathbf{H}_{21}^{(1)}(t) \\ \mathbf{H}_{22}^{(1)}(t) \end{bmatrix}^{\perp}$, the received signals after cancelling $\mathbf{V}_{10}$ at the second $2N - M$ antennas are

$$\mathbf{Y}_1^{(2)}(t) = \bar{\mathbf{H}}_{11}(t)\mathbf{P}_1(t)\mathbf{V}_1 + (\bar{\mathbf{H}}_{11}(t)\mathbf{Q}_1(t)\mathbf{U}_1 + \bar{\mathbf{H}}_{21}(t)\mathbf{P}_2\mathbf{V}_2) + \bar{\mathbf{H}}_{21}(t)\mathbf{Q}_2(t)\mathbf{U}_2 + \mathbf{N}_1(t) \tag{8.142}$$

$$\mathbf{Y}_2^{(2)}(t) = (\bar{\mathbf{H}}_{12}(t)\mathbf{P}_1(t)\mathbf{V}_1 + \bar{\mathbf{H}}_{22}(t)\mathbf{Q}_2(t)\mathbf{U}_2) + \bar{\mathbf{H}}_{22}(t)\mathbf{P}_2(t)\mathbf{V}_2 + \bar{\mathbf{H}}_{12}(t)\mathbf{Q}_1\mathbf{U}_1 + \mathbf{N}_2(t) \tag{8.143}$$

Note that $\bar{\mathbf{H}}_{ij}(t)$ is a square matrix $\forall i,\ j$. By choosing the precoding matrices as

$$\mathbf{P}_1(t) = \bar{\mathbf{H}}_{12}(t)^{-1}, \quad \mathbf{Q}_1(t) = \bar{\mathbf{H}}_{11}(t)^{-1} \tag{8.144}$$

$$\mathbf{P}_2(t) = \bar{\mathbf{H}}_{21}(t)^{-1}, \quad \mathbf{Q}_2(t) = \bar{\mathbf{H}}_{22}(t)^{-1} \tag{8.145}$$

the received signals become

$$\mathbf{Y}_1^{(2)}(t) = \bar{\mathbf{H}}_{11}(t)\bar{\mathbf{H}}_{12}(t)^{-1}\mathbf{V}_1 + (\mathbf{U}_1 + \mathbf{V}_2) + \bar{\mathbf{H}}_{21}(t)\bar{\mathbf{H}}_{22}(t)^{-1}\mathbf{U}_2 + \mathbf{N}_1(t) \tag{8.146}$$

$$\mathbf{Y}_2^{(2)}(t) = (\mathbf{V}_1 + \mathbf{U}_2) + \bar{\mathbf{H}}_{22}(t)\bar{\mathbf{H}}_{21}(t)^{-1}\mathbf{V}_2 + \bar{\mathbf{H}}_{12}(t)\bar{\mathbf{H}}_{11}(t)^{-1}\mathbf{U}_1 + \mathbf{N}_2(t) \tag{8.147}$$

Hence, the scheme is secure. $\mathbf{Y}_i^{(2)}(t)$, $t = 1, 2, 3$ correspond to $3(2N - M)$ independent observations, and we have $\mathbf{V}_1$, $\mathbf{V}_2 + \mathbf{U}_1$, $\mathbf{U}_2$ unknowns of $2N - M$ each. Consequently, we can form $3(2N - M) \times 3(2N - M)$ square linear system with unique solution using zero-forcing receiver. Therefore, each user transmits $3(M - N)$ time-varying symbols and $(2N - M)$ fixed symbols over 3 channel uses,

then $d_s = 2\frac{3(M-N)+2N-M}{3} = \frac{2(2M-N)}{3}$ is achievable.

8.9.3 Maximum Individual s.d.o.f. Point for $\frac{3N}{4} \leq M \leq N$

User 1 transmits a time-varying vector $\mathbf{V}_1^{(1)}(t),\ t = 1, 2$ Gaussian secure signals of dimension $\lfloor \frac{N}{2} \rfloor$ and a fixed vector $\mathbf{V}_1^{(2)}$ Gaussian secure signal of dimension $N \bmod 2$. The fixed vector is repeated over 2 channel uses. User 2 transmits signals with the same structure for cooperative jamming signalling. The transmitted signals are

$$\mathbf{X}_1(t) = \mathbf{P}_1(t) \begin{bmatrix} \mathbf{V}_1^{(1)}(t) \\ v_1^{(2)} \end{bmatrix} \tag{8.148}$$

$$\mathbf{X}_2(t) = \mathbf{Q}_2(t) \begin{bmatrix} \mathbf{U}_2^{(1)}(t) \\ u_2^{(2)} \end{bmatrix} \tag{8.149}$$

where $\mathbf{P}_1(t),\ \mathbf{Q}_2(t)$ satisfy

$$\begin{bmatrix} \mathbf{H}_{12}(t) & -\mathbf{H}_{22}(t) \end{bmatrix} \begin{bmatrix} \mathbf{P}_1(t) \\ \mathbf{Q}_2(t) \end{bmatrix} = \mathbf{0} \tag{8.150}$$

where $t = 1, 2$. This alignment is feasible, since $2M - N \geq \lfloor \frac{N}{2} \rfloor + N \bmod 2$ in this regime Hence, the received signals are

$$\mathbf{Y}_1(t) = \mathbf{H}_{11}(t)\mathbf{P}_1(t) \begin{bmatrix} \mathbf{V}_1^{(1)}(t) \\ v_1^{(2)} \end{bmatrix} + \mathbf{H}_{21}(t)\mathbf{Q}_2(t) \begin{bmatrix} \mathbf{U}_2^{(1)}(t) \\ u_2^{(2)} \end{bmatrix} + \mathbf{N}_1(t) \tag{8.151}$$

$$\mathbf{Y}_2(t) = \mathbf{H}_{11}(t)\mathbf{P}_1(t) \begin{bmatrix} \mathbf{U}_2^{(1)}(t) + \mathbf{V}_1^{(1)}(t) \\ u_2^{(2)} + v_1^{(2)} \end{bmatrix} + \mathbf{N}_2(t) \tag{8.152}$$

which implies that the scheme is secure. For decodability, we note that $\mathbf{Y}_1(t), t = 1, 2$ has $2N$ independent observations. The unknowns are $\mathbf{V}_1^{(1)}(t)$, $\mathbf{U}_2^{(1)}(t)$, $t = 1, 2$ with $2\lfloor \frac{N}{2} \rfloor$ dimension each and $\mathbf{V}_1^{(2)}$, $\mathbf{U}_2^{(2)}$ of $N \bmod 2$ dimension each. Hence, the total number of unknowns is $2\left(2\lfloor \frac{N}{2} \rfloor + N \bmod 2\right) = 2N$. Therefore, using this achievable scheme, and observing $\mathbf{Y}_1(t)$ for 2 channel uses, we have $2N$ independent observations. Hence, we can form $2N \times 2N$ independent linear system of equations and hence unknowns are decodable. Therefore, user 1 transmits $2\lfloor \frac{N}{2} \rfloor$ time-varying symbols and $N \bmod 2$ fixed symbol over 2 channel uses and hence $d_m = \frac{2\lfloor \frac{N}{2} \rfloor + N \bmod 2}{2} = \frac{N}{2}$ is achievable.

8.9.4 Maximum Individual s.d.o.f. Point for $N \leq M \leq \frac{3N}{2}$

The same scheme presented for static channels can be used here by replacing the structured signals $v_{12}^{(2)}$, $u_{12}^{(2)}$ by fixed Gaussian signals, which are repeated across 2 channel uses. Hence, the transmitted signals are

$$\mathbf{X}_1(t) = \mathbf{H}_{12}(t)^{\perp}\mathbf{V}_{10}(t) + \mathbf{P}_{11}(t)\mathbf{V}_{11}(t) + \mathbf{P}_{12}(t) \begin{bmatrix} \mathbf{V}_{12}^{(1)}(t) \\ v_{12}^{(2)} \end{bmatrix} \tag{8.153}$$

$$\mathbf{X}_2(t) = \mathbf{H}_{21}(t)^{\perp}\mathbf{Q}_{21}(t)\mathbf{U}_{11}(t) + \mathbf{Q}_{22}(t) \begin{bmatrix} \mathbf{U}_{12}^{(1)}(t) \\ u_{12}^{(2)} \end{bmatrix} \tag{8.154}$$

By applying the same alignment procedure, i.e., designing $\mathbf{P}_{1i}$, $\mathbf{Q}_{2i}$, $i = 1, 2$ such that

$$\begin{bmatrix} \mathbf{H}_{12}(t) & -\mathbf{H}_{22}(t)\mathbf{H}_{21}(t)^{\perp} \end{bmatrix} \begin{bmatrix} \mathbf{P}_{11}(t) \\ \mathbf{Q}_{21}(t) \end{bmatrix} = \mathbf{0} \tag{8.155}$$

$$\begin{bmatrix} \mathbf{H}_{12}(t) & -\mathbf{H}_{22}(t) \end{bmatrix} \begin{bmatrix} \mathbf{P}_{12}(t) \\ \mathbf{Q}_{22}(t) \end{bmatrix} = \mathbf{0} \tag{8.156}$$

where $t = 1, 2$, the scheme is secure. The received signal is

$$\begin{aligned} \mathbf{Y}_1(t) =& \mathbf{H}_{11}(t)\mathbf{H}_{12}(t)^{\perp}\mathbf{V}_{10}(t) + \mathbf{H}_{11}(t)\mathbf{P}_{11}(t)\mathbf{V}_{11}(t) \\ &+ \mathbf{H}_{11}(t)\mathbf{P}_{12}(t) \begin{bmatrix} \mathbf{V}_{12}^{(1)}(t) \\ v_{12}^{(2)} \end{bmatrix} + \mathbf{H}_{21}(t)\mathbf{Q}_{22}(t) \begin{bmatrix} \mathbf{U}_{12}^{(1)}(t) \\ u_{12}^{(2)} \end{bmatrix} + \mathbf{N}_1(t) \end{aligned} \tag{8.157}$$

Then, $\mathbf{Y}_1(t)$, $t = 1, 2$ correspond to $2N$ unknowns by $2N$ equations. Using transmission over 2 channel uses, the receiver has $2N$ independent observations. We have $\mathbf{V}_{10}(t)$, $\mathbf{V}_{11}(t)$, $t = 1, 2$ each with $2(M - N)$ dimensions, $\mathbf{V}_{12}^{(1)}(t)$, $\mathbf{V}_{12}^{(1)}(t)$, $t = 1, 2$ each with $2\lfloor \frac{3N-2M}{2} \rfloor$ dimensions and $v_{12}^{(2)}$, $u_{12}^{(2)}$ with $(3N - 2M) \bmod 2$ dimensions. Consequently, the total number of unknowns is $2\left(2(M - N) + 2\lfloor \frac{3N-2M}{2} \rfloor + (3N - 2M) \bmod 2\right) = 2N$. Hence, we constructed a $2N \times 2N$ system, where symbols are decodable over 2 channel uses, and $d_m = N$ is achievable.

8.10 Conclusions

We determined the exact s.d.o.f. region of a two-user $M \times N$ MIMO ICCM for any arbitrary symmetric antenna configuration. For the converse proof, we showed that the cooperative bound which results in a two-user BCCM system is tight if $M \leq \frac{2N}{3}$. We also constructed another outer bound that uses vectorized versions of the secrecy penalty and role of a helper lemmas. We used these outer bounds together with the IC without secrecy constraints to determine the entire s.d.o.f. region for any M, N.

For the achievability, we showed that the s.d.o.f. region is a four-vertex polytope. Focusing on the sum s.d.o.f. point, if the sum s.d.o.f. is an integer (fractional part is zero), then there is no need for real interference alignment; spatial alignment suffices. If the fractional part is 1/3, then after spatial alignment, real alignment in a single dimension is needed. For the case where the fraction is 2/3, we developed a novel achievable scheme for the basic 2×2 MIMO ICCM. This scheme together with its SISO counterpart are central for achievable schemes for general M and N. The 2×2 scheme combines spatial alignment, which ensures that secure signals and cooperative jamming signals lie in the same rational dimension irrespective of the joint MIMO processing used at the receiver, and asymptotic real interference alignment to minimize the required dimensions needed for decodability and ensuring that observations of all receiving antennas are exploited. We showed the achievability of the other non-trivial polytope points by forcing one of the users to act as a cooperative jammer (helper) that jams its own receiver.

Interestingly, we showed that the s.d.o.f. region starts as a square region if

$M \leq \frac{2N}{3}$, then it takes the shape of an irregular polytope until it returns back to a square region when the number of transmit antennas is at least twice the number of receiving antennas. We showed that if the ICCM channel is time-varying, the achievable schemes can be simplified by using vector space alignment via symbol extension over multiple channel uses instead of real interference alignment that is necessary for static channels.

CHAPTER 9

Secure Degrees of Freedom in Networks with User Misbehavior

9.1 Introduction

In this chapter, we investigate s.d.o.f. of three new channel models: broadcast channel with combating helpers, interference channel with selfish users, and multiple access wiretap channel with deviating users. The goal of introducing these channel models is to investigate various malicious interactions that arise in networks, including active adversaries. That is in contrast with the common assumption in the literature that the users follow a certain protocol altruistically and transmit both message-carrying and cooperative jamming signals in an optimum manner.

In the first model, over a classical broadcast channel with confidential messages (BCCM), there are two helpers, each associated with one of the receivers. In the second model, over a classical interference channel with confidential messages (ICCM), there is a helper and users are selfish. By casting each problem as an extensive-form game and applying recursive real interference alignment, we show that, for the first model, the combating intentions of the helpers are neutralized

and the full s.d.o.f. is retained; for the second model, selfishness precludes secure communication and no s.d.o.f. is achieved.

In the third model, we consider the multiple access wiretap channel (MAC-WTC), where multiple legitimate users wish to have secure communication with a legitimate receiver in the presence of an eavesdropper. We consider the case when a subset of users deviate from the optimum protocol that attains the exact s.d.o.f. of this channel. We consider two kinds of deviation: when some of the users stop transmitting cooperative jamming signals, and when a user starts sending intentional jamming signals. For the first scenario, we investigate possible responses of the remaining users to counteract such deviation. For the second scenario, we use an extensive-form game formulation for the interactions of the deviating and well-behaving users. We prove that a deviating user can drive the s.d.o.f. to zero; however, the remaining users can exploit its intentional jamming signals as cooperative jamming signals against the eavesdropper and achieve an optimum s.d.o.f.

9.2 BCCM with Combating Helpers

9.2.1 System Model and Assumptions

In BCCM, the transmitter has two private messages W_1 and W_2 picked from the message sets $\mathcal{W}_1, \mathcal{W}_2$ uniformly with rates R_1, R_2, respectively, where $R_i = \frac{1}{n} \log |\mathcal{W}_i|$. Each message W_i should be received reliably by the ith receiver, while being kept

secure from the jth receiver, $i \neq j$:

$$\mathbb{P}(\hat{W}_1 \neq W_1) \leq \epsilon, \quad \mathbb{P}(\hat{W}_2 \neq W_2) \leq \epsilon \tag{9.1}$$

$$\frac{1}{n} I(W_2; Y_1^n) \leq \epsilon, \quad \frac{1}{n} I(W_1; Y_2^n) \leq \epsilon \tag{9.2}$$

where $\hat{W}_i$ is the estimate of W_i at the ith receiver. The s.d.o.f. d_i is defined as $d_i = \lim_{P\to\infty} \frac{R_i}{\frac{1}{2}\log P}$, where P is the transmitter power constraint $\mathbb{E}[X^2] \leq P$.

The system has two helpers with inputs Z_1 and Z_2, with the power constraints $\mathbb{E}[Z_i^2] \leq P$. Each helper assists secure transmission to *one* of the receivers. The input/output relations for the BCCM with combating helpers (see Fig. 9.1) are:

$$Y_1[k] = hX[k] + \tilde{h}_1 Z_1[k] + \tilde{h}_2 Z_2[k] + N_1[k] \tag{9.3}$$

$$Y_2[k] = gX[k] + \tilde{g}_1 Z_1[k] + \tilde{g}_2 Z_2[k] + N_2[k] \tag{9.4}$$

where $Y_i[k]$ is the received signal at the ith receiver in the kth transmission frame, h, g are the channel gains from the transmitter to receivers 1, 2, respectively, and $\tilde{h}_i$, $\tilde{g}_i$ are the channel gains from helper i to receivers 1, 2, respectively.

The helpers are *combating* as they maximize the s.d.o.f. of one user only, while hurting the other user by sending jamming signals. The transmitter acts in even transmission frames, and helpers respond in odd frames. Each node has perfect channel state information (CSI) and knows the actions of others at the end of every frame. We require that the action of a helper does not hurt its own receiver (in terms of s.d.o.f.) if no new jamming signals are produced by the other helper.

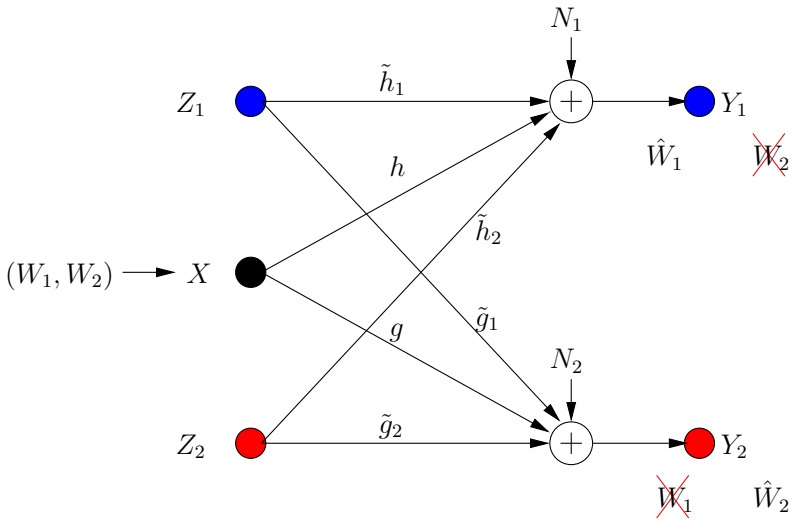


Figure 9.1: BCCM with combating helpers.

Consequently, we formalize the role of the ith helper as:

$$\min \quad d_j(k) \qquad \text{s.t.} \quad d_i(k) = d_i(k-1) \tag{9.5}$$

where $i, j \in \{1, 2\}$, $i \neq j$ and $d_j(k)$ is the s.d.o.f. of the jth user in the kth transmission frame, where k is odd. On the other hand, the transmitter does not take the side of any of the users and maximizes the sum s.d.o.f. of the system, i.e., transmitter's role in even encoding frames is:

$$\max \quad d_1(k) + d_2(k) \tag{9.6}$$

9.2.2 Achievable Scheme: Recursive Real Interference Alignment as Extensive-Form Game

We use recursive real interference alignment as achievable strategy for our model. At encoding frame k, all secure and jamming signals are picked from PAM constellation set $C(a_k, Q_k)$, where a_k is the minimum distance between any two points in the constellation and Q_k is the number of points.

9.2.2.1 For Frames $k = 0$, $k = 1$

Frames 0 and 1 are considered transient frames. For frame 0, the transmitter performs the optimal strategy in the presence of helpers [78], and sends two signal components V_{11}, V_{21} in two irrational dimensions,

$$X[0] = \alpha_1 V_{11} + \alpha_2 V_{21} \tag{9.7}$$

where α_1, α_2 are rationally independent scalars. These message-carrying signals are not secured. None of the helpers expects the other helper to jam its own receiver, hence each helper needs to protect the message of its own receiver at the other receiver. Hence, at $k = 1$, the ith helper sends a structured jamming signal $\tilde{U}_{i1}$ in the irrational dimension where its message-carrying signal lies at the other receiver as

$$Z_1[1] = \frac{\alpha_1 g}{\tilde{g}_1}\tilde{U}_{11}, \quad Z_2[1] = \frac{\alpha_2 h}{\tilde{h}_2}\tilde{U}_{21} \tag{9.8}$$

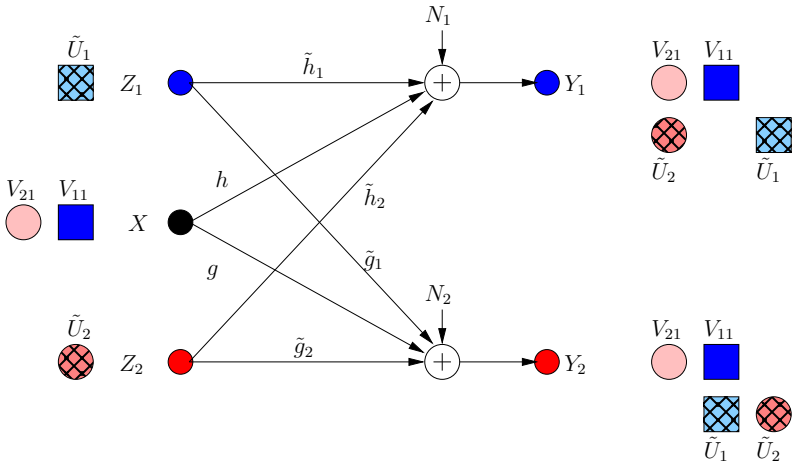


Figure 9.2: BCCM frame $k = 1$. Pink circle and blue square denote user signals, and the hatched circles/squares denote corresponding helper jamming signals.

Then, the received signals are

$$Y_1[1] = \alpha_1 h V_{11} + \frac{\alpha_1 g \tilde{h}_1}{\tilde{g}_1}\tilde{U}_{11} + \alpha_2 h(V_{21} + \tilde{U}_{21}) + N_1 \tag{9.9}$$

$$Y_2[1] = \alpha_2 g V_{21} + \frac{\alpha_2 h \tilde{g}_2}{\tilde{h}_2}\tilde{U}_{21} + \alpha_1 g(V_{11} + \tilde{U}_{11}) + N_2 \tag{9.10}$$

Although V_{11}, V_{21} are now secure, this results in a new irrational dimension at each receiver as in Fig. 9.2. Hence $d_i(1) = 1/3$ for each user as we show formally in Section 9.2.3 (instead of $d_i = 1/2$ in BCCM with coordinating helpers).

9.2.2.2 For Frame $k = 2$

The transmitter knows that a new irrational dimension is generated within frame $k = 1$. The transmitter uses this dimension in its favor, as it can protect more message-carrying signals. It produces two new message-carrying signal components

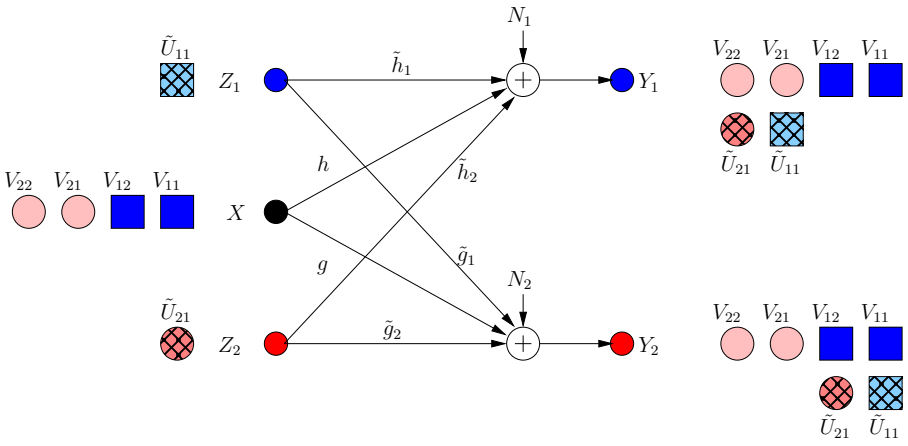


Figure 9.3: BCCM frame $k = 2$.

V_{12}, V_{22} to be aligned with the generated jamming dimensions in frame $k = 1$ as,

$$X[2] = \alpha_1 V_{11} + \alpha_2 V_{21} + \frac{\alpha_2 h \tilde{g}_2}{\tilde{h}_2 g} V_{12} + \frac{\alpha_1 g \tilde{h}_1}{\tilde{g}_1 h} V_{22} \tag{9.11}$$

$$= X[1] + \beta_1 V_{12} + \beta_2 V_{22} \tag{9.12}$$

That is, the transmitter appends its last frame transmission with two new signal components in rationally independent dimensions β_1, β_2 (see Fig. 9.3). The received signals are,

$$Y_1[1] = \alpha_1 h V_{11} + \frac{\alpha_2 h^2 \tilde{g}_2}{\tilde{h}_2 g} V_{12} + \frac{\alpha_1 g \tilde{h}_1}{\tilde{g}_1} (V_{22} + \tilde{U}_{11}) + \alpha_2 h (V_{21} + \tilde{U}_{21}) + N_1 \tag{9.13}$$

$$Y_2[1] = \alpha_2 g V_{21} + \frac{\alpha_1 g^2 \tilde{h}_1}{\tilde{g}_1 h} V_{22} + \frac{\alpha_2 h \tilde{g}_2}{\tilde{h}_2} (V_{12} + \tilde{U}_{12}) + \alpha_1 g (V_{11} + \tilde{U}_{11}) + N_2 \tag{9.14}$$

Consequently, the system retains full s.d.o.f. ($d_i(2) = 1/2$).

9.2.2.3 For Frame $k = 3$

Now, each helper minimizes the s.d.o.f. of the other user by sending jamming signal. However, due to the strong constraint $d_i(3) = d_i(2)$, no helper jams the other receiver directly, as this would create a new jamming dimension at the side of its own receiver, decreasing its own s.d.o.f. Instead, it transmits a jamming signal which aligns with the already jammed dimension at its own receiver as,

$$Z_1[3] = Z_1[1] + \frac{\alpha_2 h}{\tilde{h}_1}\tilde{U}_{12}, \quad Z_2[3] = Z_2[1] + \frac{\alpha_1 g}{\tilde{g}_2}\tilde{U}_{22} \tag{9.15}$$

Consequently, the received signals are,

$$Y_1[3] = Y_1[2] + \alpha_2 h\tilde{U}_{12} + \frac{\alpha_1 \tilde{h}_2 g}{\tilde{g}_2}\tilde{U}_{22} \tag{9.16}$$

$$Y_2[3] = Y_2[2] + \alpha_1 g\tilde{U}_{22} + \frac{\alpha_2 \tilde{g}_2 h}{\tilde{h}_1}\tilde{U}_{12} \tag{9.17}$$

Since the $\alpha_2 h$ dimension is already jammed, the first helper does not create a new irrational dimension. Hence, it does not hurt its own receiver. However, it creates a new jamming dimension $\frac{\alpha_2 \tilde{g}_2 h}{\tilde{h}_1}$ at the second receiver, which decreases the resultant s.d.o.f. From the symmetry, the second helper applies the same strategy and hence the resulting s.d.o.f. is $d_i(3) = 2/5$ as in Fig. 9.4. Note that, neither of the helpers can hold back its original jamming signal (i.e., each helper should append its previous signalling with new jamming signals), because if not, its previous message-carrying signals are compromised.

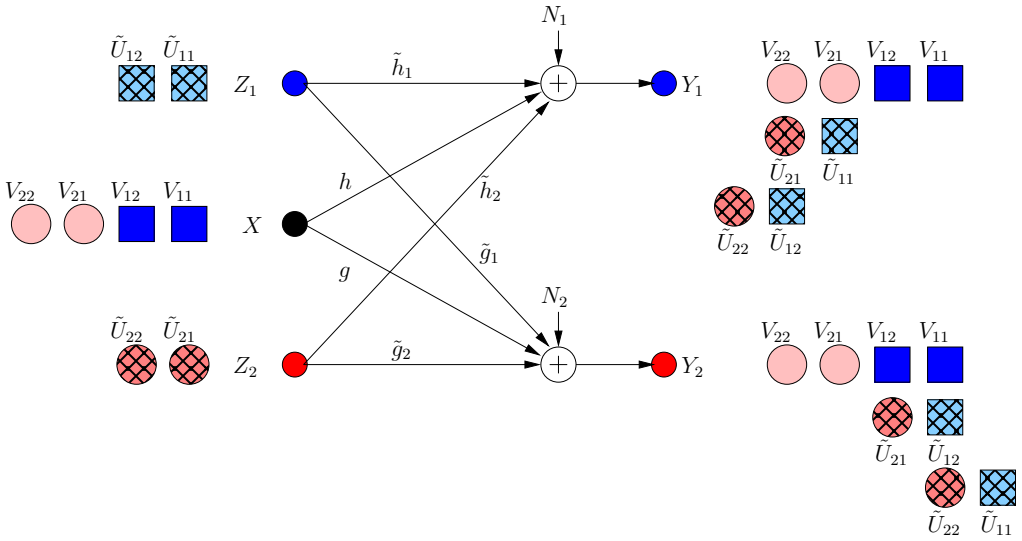


Figure 9.4: BCCM frame $k = 3$.

9.2.2.4 For General kth Frame

If k is odd, the helpers produce one extra jamming component aligned with the last generated jamming signal of the other helper. If k is even, the transmitter makes use of this jamming signal and provides two extra secure signals, achieving the maximum possible s.d.o.f. ($d_i(k) = 1/2$, k is even).

9.2.3 Calculation of the Secure Degrees of Freedom

To calculate the s.d.o.f., we need the following lemma.

Lemma 9.1 *If every message-carrying signal is protected by a cooperative jamming signal, then the s.d.o.f. is given by*

$$d_i(k) = \frac{J_k}{L_k} \tag{9.18}$$

where J_k is the number of irrational dimensions needed to receive the message-carrying signal of user i at the kth frame $\mathbf{V}_i[k] = [V_{i1}, V_{i2}, \dots V_{iJ_k}]^T$ and L_k is the total number of irrational dimensions.

Proof: We give only a sketch of a proof here as it follows standard arguments. At every encoding frame, the transmitter transmits PAM signals with parameters $Q_k = P^{\frac{1-\delta}{2(L_k+\delta)}}$ and $a_k = \gamma P^{\frac{1}{2}}/Q_k$. This satisfies the power constraint and ensures that the probability of error goes to zero as $P \to \infty$. From [70], the following rates are achievable for the BCCM,

$$R_1[k] \geq I(\mathbf{V}_1[k]; Y_1[k]) - I(\mathbf{V}_1[k]; Y_2[k]|\mathbf{V}_2[k]) \tag{9.19}$$

By techniques similar to [78], we calculate $I(\mathbf{V}_1[k]; Y_1[k]) \geq \frac{J_k(1-\delta)}{L_k+\delta}\left(\frac{1}{2}\log P\right) + o(\log P)$, while the leakage rate is upper bounded by $o(\log P)$, as every message-carrying signal is protected by CJ signal. Taking limits concludes the proof. ■

Theorem 9.1 *For BCCM with combating helpers under the constraint of not decreasing the s.d.o.f. of their own receivers due to helper actions, the s.d.o.f. of each user evolves as,*

$$d_i(k) = \begin{cases} 1/2, & k \text{ even} \\ \frac{k+1}{2k+4} \to 1/2, & k \text{ odd} \end{cases} \tag{9.20}$$

I.e., the combating behaviour is asymptotically neutralized.

Proof: Using Lemma 9.1, we have $d_i(k) = \frac{J_k}{L_k}$. We complete the proof by calculating the dimensions J_k, L_k. We prove this by induction on k. For the base step $k = 1$,

we have $J_k = 1$ and $L_k = 3$ which conforms with (9.20). For $k = 2$, we have $J_k = 2$ and $L_k = 4$ and hence $d_i(k) = 1/2$.

For the induction step, assume that k is odd and $d_i(k-2) = \frac{k-1}{2k}$. Then, in the $(k-1)$th frame, transmitter can always add extra 2 message-carrying signals to have $d_i(k-1) = 1/2$. Thus, $J_{k-1} = J_{k-2} + 1$ and $L_{k-1} = L_{k-2} + 1$. This is because the transmitter uses the extra irrational dimension produced by jamming in odd frames in its favor, hence it adds one extra dimension corresponding to the new message-carrying signal. This results in the following simultaneous equations,

$$\frac{J_{k-2}}{L_{k-2}} = \frac{k-1}{2k}, \quad \frac{J_{k-1}}{L_{k-1}} = \frac{J_{k-2}+1}{L_{k-2}+1} = \frac{1}{2} \tag{9.21}$$

Solving these two equations gives $L_{k-2} = k$ and $J_{k-2} = \frac{(k-1)}{2}$. Then, $L_{k-1} = k+1$ and $J_{k-1} = \frac{k+1}{2}$. In the next frame transmission, each helper produces extra jamming component aligned with already jammed dimension. This increases L_k by one at the other receiver without changing J_k. Consequently, $d_i(k) = \frac{J_k}{L_k} = \frac{\frac{k+1}{2}}{k+2} = \frac{k+1}{2k+4}$, which converges to $1/2$. ■

9.3 ICCM with Selfish Users

9.3.1 System Model and Assumptions

In ICCM, each transmitter has a message W_i picked from the message set $\mathcal{W}_i$ uniformly with rate $R_i = \frac{1}{n}\log|\mathcal{W}_i|$ for $i \in \{1, 2\}$. Message W_i should be received reliably by the ith receiver, while being kept secure from the jth receiver, $i \neq j$.

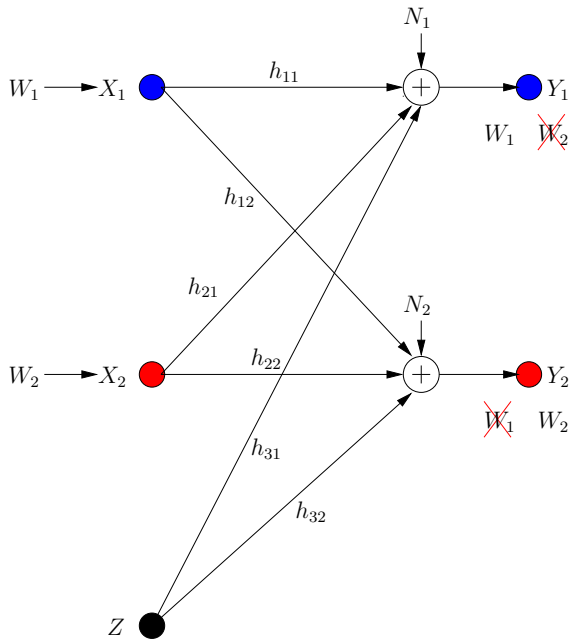


Figure 9.5: ICCM with selfish users.

The system has an external helper with channel input Z. Inputs satisfy power constraints $\mathbb{E}[X_i^2] \leq P$ and $\mathbb{E}[Z^2] \leq P$. The ICCM model depicted in Fig. 9.5 is given by,

$$Y_1[k] = h_{11}X_1[k] + h_{21}X_2[k] + h_{31}Z[k] + N_1[k] \tag{9.22}$$

$$Y_2[k] = h_{12}X_1[k] + h_{22}X_2[k] + h_{32}Z[k] + N_2[k] \tag{9.23}$$

where $Y_i[k]$ is the received signal at the ith receiver in the kth transmission frame, h_{ij} is the channel gain from transmitter $i = 1, 2, 3$ (transmitter 3 is the helper) to receiver $j = 1, 2$.

The users are *selfish* and malicious. User i maximizes the individual s.d.o.f. at receiver Y_i, while maximally hurting the second user. Formally, the ith user's role

is,

$$\max \quad d_i(k) - d_j(k) \tag{9.24}$$

where $i \neq j$, $i, j \in \{1, 2\}$. The role of the users here is *less stringent* than the BCCM model, since in the ICCM model, we allow the users to hurt their own receivers if they hurt the other receiver more. On the other hand, the system helper does not take side of any of the users and maximizes the sum s.d.o.f. of the system,

$$\max \quad d_i(k) + d_j(k) \tag{9.25}$$

9.3.2 Achievable Scheme: Recursive Real Interference Alignment as Extensive Form Game

Similar to the BCCM, we propose to use recursive interference alignment using PAM constellation $C(a_k, Q_k)$.

9.3.2.1 For Frame $k = 0$

All nodes perform the optimal *selfless* strategy as in [78]. The transmitted signals are,

$$X_1[0] = \frac{h_{32}}{h_{12}} V_{11}, \quad X_2[0] = \frac{h_{31}}{h_{21}} V_{21}, \quad Z[0] = \tilde{U}_1 \tag{9.26}$$

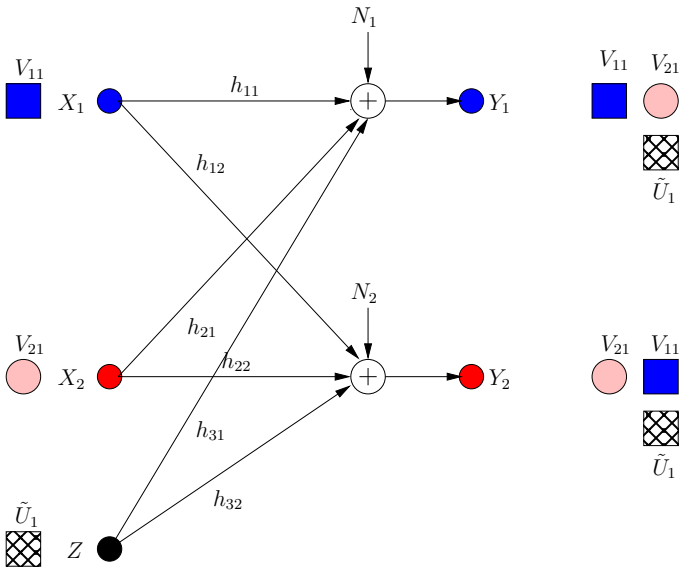


Figure 9.6: ICCM frame $k = 0$. Pink circle and blue square denote user signals, and the hatched squares denote jamming signals.

The received signals at both receivers are (as in Fig. 9.6),

$$Y_1[0] = \frac{h_{32}h_{11}}{h_{12}}V_{11} + h_{31}(V_{21} + \tilde{U}_1) + N_1 \tag{9.27}$$

$$Y_2[0] = \frac{h_{31}h_{22}}{h_{21}}V_{21} + h_{32}(V_{11} + \tilde{U}_1) + N_2 \tag{9.28}$$

which implies that the achievable s.d.o.f. $d_i(0) = 1/2$.

9.3.2.2 For Frame $k = 1$

User i maximizes $d_i(1) - d_j(1)$ assuming that user j keeps its strategy as in frame 0. Each user prefers to jam the other user directly, even if it results in partial decrease of its own s.d.o.f. (by creating extra dimension at its receiver), since in this case it can drive the s.d.o.f. of the other user to zero and maximize the s.d.o.f. difference.

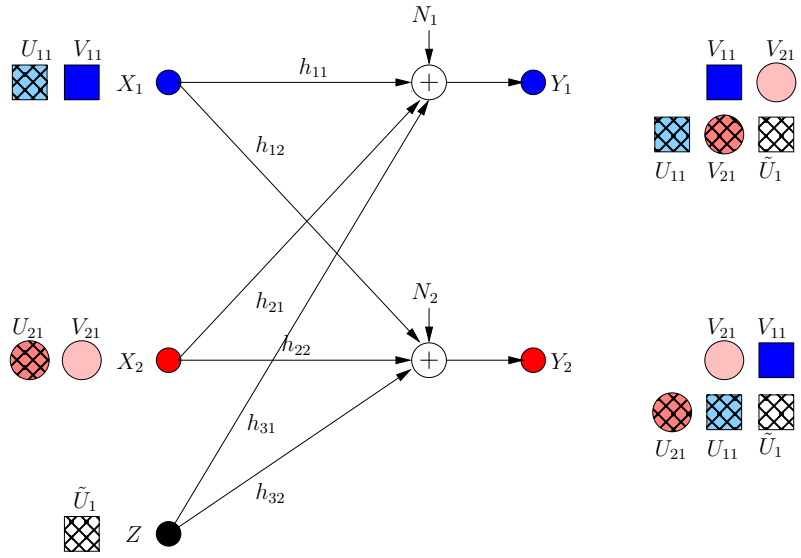


Figure 9.7: ICCM frame $k = 1$.

Thus,

$$X_1[1] = X_1[0] + \frac{h_{31}h_{22}}{h_{12}h_{21}}U_{11} \tag{9.29}$$

$$X_2[1] = X_2[0] + \frac{h_{32}h_{11}}{h_{12}h_{21}}U_{21} \tag{9.30}$$

Hence, the received signals in this case are,

$$Y_1[1] = \frac{h_{32}h_{11}}{h_{12}}(V_{11} + U_{21}) + h_{31}(V_{21} + \tilde{U}_1) + \frac{h_{31}h_{22}h_{11}}{h_{12}h_{21}}U_{11} + N_1 \tag{9.31}$$

$$Y_2[1] = \frac{h_{31}h_{22}}{h_{21}}(V_{21} + U_{11}) + h_{32}(V_{11} + \tilde{U}_1) + \frac{h_{32}h_{12}h_{22}}{h_{12}h_{21}}U_{11} + N_2 \tag{9.32}$$

which implies that all secure signals are jammed and communication is driven to zero s.d.o.f. as in Fig. 9.7.

9.3.2.3 For Frame $k = 2$

Both users know that their communication links are jammed during frame $k = 1$. Therefore, the problem of maximizing the s.d.o.f. difference reduces to maximizing s.d.o.f. of individual user. Since the s.d.o.f. of the other user is zero. Each user benefits from the extra jamming dimension created by the other user to protect extra message-carrying component. Moreover, the helper produces extra jamming component in a new irrational dimension, which allows each user to produce extra secure signal. Thus,

$$X_1[2] = X_1[1] + \frac{\alpha_1 h_{32}}{h_{12}} V_{12} + \frac{h_{32} h_{11} h_{22}}{h_{12}^2 h_{21}} V_{13} \tag{9.33}$$

$$X_2[2] = X_2[1] + \frac{\alpha_1 h_{31}}{h_{21}} V_{22} + \frac{h_{31} h_{22} h_{11}}{h_{21}^2 h_{12}} V_{23} \tag{9.34}$$

$$Z[2] = Z[1] + \alpha_1 \tilde{U}_2 \tag{9.35}$$

where α_1 is irrational number independent from all channel gains. Hence, the received signals are,

$$Y_1[2] = Y_1[1] + \alpha_1 h_{31}(V_{22} + \tilde{U}_2) + \frac{h_{31} h_{22} h_{11}}{h_{21} h_{12}} V_{23} + \frac{\alpha_1 h_{32} h_{11}}{h_{12}} V_{12} + \frac{h_{32} h_{11}^2 h_{22}}{h_{12}^2 h_{21}} V_{13} \tag{9.36}$$

$$Y_2[2] = Y_2[1] + \alpha_1 h_{32}(V_{12} + \tilde{U}_2) + \frac{h_{32} h_{11} h_{22}}{h_{12} h_{21}} V_{13} + \frac{\alpha_1 h_{31} h_{22}}{h_{21}} V_{22} + \frac{h_{31} h_{22} h_{11}}{h_{21}^2 h_{12}} V_{23} \tag{9.37}$$

Consequently, $d_i(2) = 1/3$ as shown in Fig. 9.8.

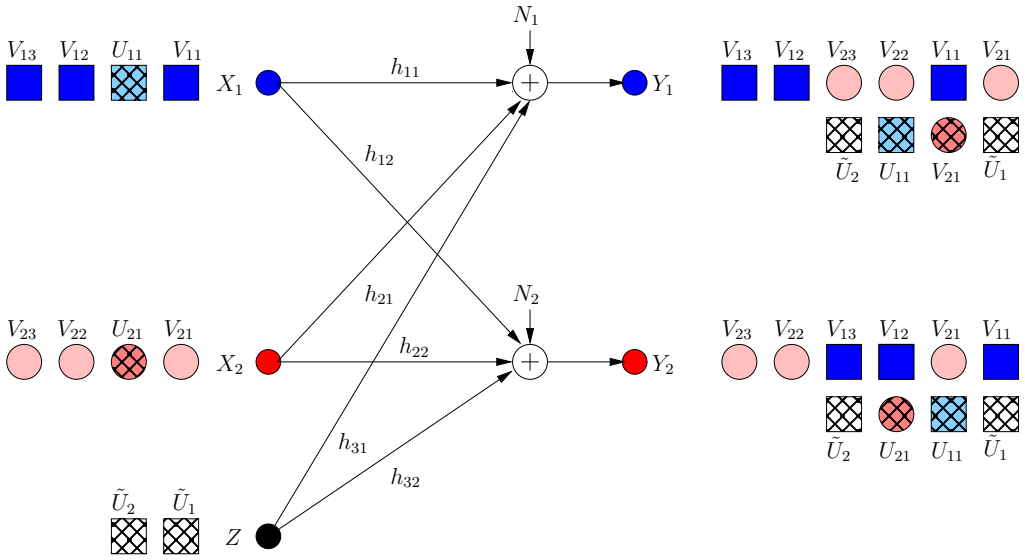


Figure 9.8: ICCM frame $k = 2$.

9.3.2.4 For General kth Frame

The s.d.o.f. differs whether k is odd/even. If k is odd, each user chooses to jam all dimensions of the other user's secure signals. This choice leads to $d_i(k) = 0$ for all odd frames. If k is even, each user takes advantage of the generated jamming by the other user plus extra jamming signal from the system helper to protect more signals.

9.3.3 Calculation of the Secure Degrees of Freedom

Theorem 9.2 *For the ICCM with selfish users in the presence of a system helper, assuming that users maximize the s.d.o.f. difference for every transmission frame, the s.d.o.f. evolves as*

$$d_i(k) = \begin{cases} 0, & k \text{ odd} \\ \frac{2}{k+4} \to 0, & k \text{ even} \end{cases} \tag{9.38}$$

I.e., selfishness eventually precludes secure communication.

Proof: From [70], the rates given in (9.19) are achievable for the ICCM. Then, from Lemma 9.1, we have $d_i(k) = \frac{J_k}{L_k}$. Next, we count $J_k = \frac{k+2}{2}$, when k is even. This follows by induction: For $k = 1$, the number of secure dimensions is 1. Now, assume that the relation holds for any even $k-2$. Then, $J_{k-2} = \frac{k}{2}$. Then, since user i jams all secure dimensions of user j in frame $k-1$, it creates $\frac{k}{2}$ new dimensions. These dimensions are used by user i in frame k to protect $\frac{k}{2}$ new secure signals. The helper produces extra jamming component allowing protection of one extra signals. Then, $J_k = \frac{k}{2} + 1 = \frac{k+2}{2}$.

We use this result in proving s.d.o.f. by induction: For $k = 0$, $J_0 = 1$ and $L_0 = 2$, which leads to $d_i(0) = 1/2$. For $k = 1$, $J_1 = 0$ and $L_1 = 3$, which leads to $d_i(1) = 0$. Now, assume that k is even and expression (9.38) is true, then, $d_i(k-2) = \frac{2}{k+2}$. Then, from above, we have $J_{k-2} = \frac{k}{2}$. Hence, $L_{k-2} = \frac{k(k+2)}{4}$. The total dimensions L_k at any receiver is increased over the $k-2$ frame by $2J_k$, since the increase is caused by the new secure dimensions J_k for the two users which are symmetric. Therefore, the s.d.o.f. for even k is

$$d_i(k) = \frac{J_k}{L_k} = \frac{J_k}{L_{k-2} + 2J_k} = \frac{2}{k+4} \tag{9.39}$$

If k is odd, users make s.d.o.f. zero, completing the proof. ■

Remark 9.1 *Although the previous channel models are different, they have critical similarities: In both models there is a central node, transmitter in BCCM and helper*

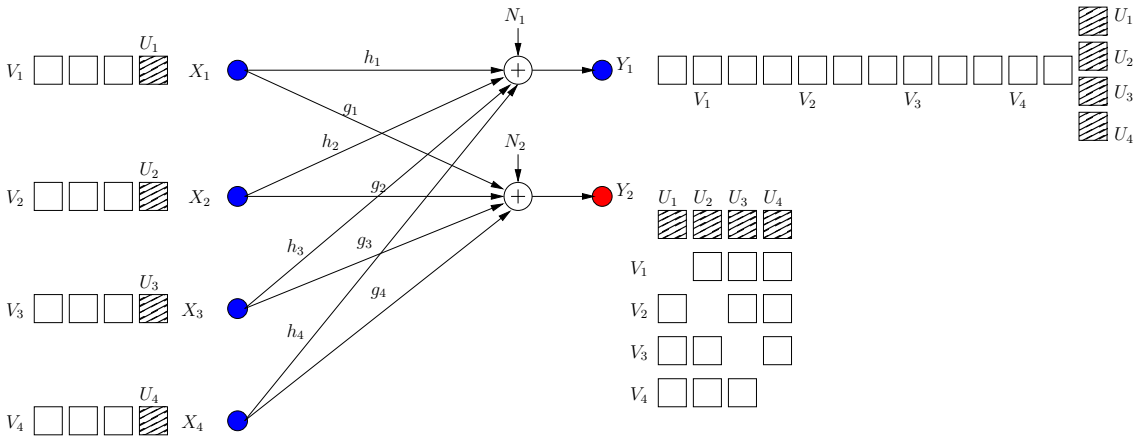


Figure 9.9: Optimal achievable scheme for a $K = 4$ user MAC-WTC.

in ICCM, which altruistically want to maximize the sum s.d.o.f., however, transmitter in BCCM can send useful signals, but helper ICCM can only jam. In both models there are two adversarial/selfish transmitters, helpers in BCCM and users in ICCM, however, helpers in BCCM can only jam, but users in ICCM can send useful signals and/or jam. We observe that this difference in roles drives systems to opposite end results of full s.d.o.f. in BCCM and zero s.d.o.f. in ICCM.

9.4 Multiple Access Wiretap Channel with Deviating Users

9.4.1 System Model and Assumptions

The K-user Gaussian MAC-WTC is given by (see Fig. 9.9),

$$Y_1 = \sum_{i=1}^{K} h_i X_i + N_1 \tag{9.40}$$

$$Y_2 = \sum_{i=1}^{K} g_i X_i + N_2 \tag{9.41}$$

where Y_1, Y_2 are the channel outputs at the legitimate receiver and the eavesdropper, respectively, h_i, g_i are the channel gains from user i to the receiver and the eavesdropper, respectively. User i has a message W_i picked uniformly from the message set $\mathcal{W}_i$, with a rate $R_i = \frac{1}{n}\log|\mathcal{W}_i|$, and sends it in n channel uses using X_i^n reliably and securely, i.e.,

$$\mathbb{P}(\hat{W}_1^K \neq W_1^K) \leq \epsilon, \quad \frac{1}{n} I(W_1^K; Y_2^n) \leq \epsilon \tag{9.42}$$

where $W_1^K = (W_1, \ldots, W_K)$, and $\hat{W}_1^K = (\hat{W}_1, \ldots, \hat{W}_K)$ are the estimates of the messages at the legitimate receiver. The transmitters are subject to power constraints $\mathbb{E}[X_i^2] \leq P$. The sum s.d.o.f. is given by $d_s = \lim_{P\to\infty} \frac{\sum_{i=1}^K R_i}{\frac{1}{2}\log P}$.

In the second part of the chapter, we consider a severe form of deviation where one user transmits intentional jamming signals. To distinguish that user and its jamming signal, we denote its channel input as Z, which also is subject to the power constraint $\mathbb{E}[Z^2] \leq P$, and we designate it as the Kth user without loss of generality, see Fig. 9.13. The malicious user and the remaining users respond to each other in multiple coding frames. The channel inputs/outputs for this model in frame k are:

$$Y_1[k] = \sum_{i=1}^{K-1} h_i X_i[k] + \tilde{h} Z[k] + N_1[k] \tag{9.43}$$

$$Y_2[k] = \sum_{i=1}^{K-1} g_i X_i[k] + \tilde{g} Z[k] + N_2[k] \tag{9.44}$$

where $\tilde{h}, \tilde{g}$ are the channel gains from the malicious user to the legitimate receiver

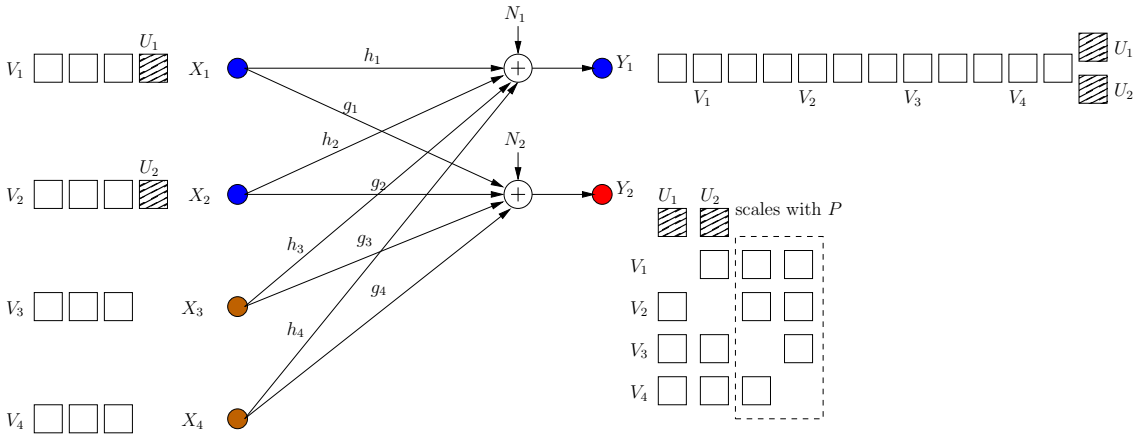


Figure 9.10: The remaining users keep their originally optimum schemes.

and the eavesdropper, respectively.

9.4.2 S.d.o.f. When Remaining Users Do Not Respond

Consider that M users have deviated from the optimum strategy in [78] (see Fig. 9.9) by not sending cooperative jamming signals and that the remaining users have kept their originally optimum strategies, i.e., have not responded to the deviating users (see Fig. 9.10). That is, the user signals are [78],

$$X_i = \begin{cases} \sum_{j=1, j\neq i}^{K} \frac{g_j}{g_i h_j} V_{ij} + \frac{1}{h_i} U_i, & i = 1, \ldots, K-M \\ \sum_{j=1, j\neq i}^{K} \frac{g_j}{g_i h_j} V_{ij}, & i = K-M+1, \ldots, K \end{cases} \tag{9.45}$$

where V_{ij}, U_i are picked uniformly from PAM constellation set $C(a, Q)$ [78]. The constants a, Q are chosen as [78]

$$Q = P^{\frac{1-\delta}{2K(K-1)+1+\delta}}, \quad a = \gamma \frac{P^{1/2}}{Q} \tag{9.46}$$

Consequently, the received signals are (see Fig. 9.10),

$$Y_1 = \sum_{i=1}^{K} \sum_{j=1,j\neq i}^{K} \frac{g_j h_i}{g_j h_j} V_{ij} + \sum_{k=1}^{K-M} U_k + N_1 \tag{9.47}$$

$$Y_2 = \sum_{i=1}^{K} \sum_{j=1,j\neq i}^{K} \frac{g_j}{h_j} V_{ij} + \sum_{j=1}^{K-M} \frac{g_j}{h_j} U_j + N_2 \tag{9.48}$$

$$= \sum_{j=1}^{K-M} \frac{g_j}{h_j} \left(U_j + \sum_{i=1,i\neq j}^{K} V_{ij} \right) + \sum_{j=K-M+1}^{K} \sum_{i=1,i\neq j}^{K} \frac{g_j}{h_j} V_{ij} + N_2 \tag{9.49}$$

Let $\mathbf{V} = \{V_{ij} : i, j = 1, \ldots K, i \neq j\}$. From [78, 80], the following secure rates are achievable,

$$\sum_{i=1}^{K} R_i \geq I(\mathbf{V}; Y_1) - I(\mathbf{V}; Y_2) \tag{9.50}$$

For the first term $I(\mathbf{V}; Y_1)$: we note that the components of vector $\mathbf{V}$ are received in different rational dimensions, and hence we have $(2Q+1)^{K(K-1)}$ separable constellation points, while the cooperative jamming signal components are aligned in the same rational dimension, i.e., $(2(K-M)Q+1)$ constellation points. From data processing and Fano's inequalities,

$$I(\mathbf{V}; Y_1) \geq I(\mathbf{V}; \hat{\mathbf{V}}) = H(\mathbf{V}) - H(\mathbf{V}|\hat{\mathbf{V}}) \tag{9.51}$$

$$\geq [1 - \exp(\eta_\gamma P^\delta)] \log(2Q+1)^{K(K-1)} - 1 \tag{9.52}$$

$$= \frac{K(K-1)(1-\delta)}{K(K-1)+1+\delta} \cdot \frac{1}{2} \log P + o(\log P) \tag{9.53}$$

For the second term $I(\mathbf{V}; Y_2)$: we note that we have $K-M$ dimensions, in which message-carrying signals are aligned with cooperative jamming signals, while M dimensions lack cooperative jamming signals, i.e., we have $(2KQ+1)^{K-M} \cdot (2(K-$

$1)Q+1)^M$ constellation points. Hence,

$$\begin{aligned} I(\mathbf{V};Y_2) \leq & H(Y_2-N_2)-H(Y_2-N_2|\mathbf{V}) && (9.54)\\ \leq & \log(2KQ+1)^{K-M}(2(K-1)Q+1)^M)-\log(2Q+1)^{K-M} && (9.55)\\ = & (K-M)\log\frac{2KQ+1}{2Q+1}+M\log(2(K-1)Q+1) && (9.56)\\ \leq & (K-M)\log K+\frac{M(1-\delta)}{K(K-1)+1+\delta}\cdot\frac{1}{2}\log P+o(\log P) && (9.57)\\ = & \frac{M(1-\delta)}{K(K-1)+1+\delta}\cdot\frac{1}{2}\log P+o(\log P) && (9.58) \end{aligned}$$

Substituting (9.53) and (9.58) into (9.50), and taking the limit as $P\to\infty$, the achievable sum s.d.o.f. is,

$$d_s \geq \frac{K(K-1)-M}{K(K-1)+1} \tag{9.59}$$

That is, the sum s.d.o.f. decreases by $\frac{M}{K(K-1)+1}$ from the optimal in [78]. This affects all users, including the deviating users, hence they do not benefit from their deviation.

9.4.3 S.d.o.f. When Remaining Users Respond

In this section, we consider two achievable schemes resulting from two different responses of the remaining users.

9.4.3.1 Reducing the Secure Rate for Zero Leakage Rate

In this achievable scheme, all users decrease their secure rates, i.e., decrease the number of message-carrying signal components to ensure that all of them are aligned

with cooperative jamming signals. Specifically, the first $K-M$ users send $K-M-1$ message-carrying signals and 1 cooperative jamming signal, while the rest of the users, i.e., the deviating users, send $K-M$ message-carrying signals and no cooperative jamming signals, see Fig. 9.11. Note that the deviating users are motivated to decrease their message-carrying signals from $K-1$ to $K-M$, as otherwise, some of their message-carrying signals would not be protected. The transmitted signals are,

$$X_i = \begin{cases} \sum_{j=1,j\neq i}^{K-M} \frac{g_j}{g_i h_j} V_{ij} + \frac{1}{h_i} U_i, & i = 1, \ldots, K-M \\ \sum_{j=1}^{K-M} \frac{g_j}{g_i h_j} V_{ij}, & i = K-M+1, \ldots, K \end{cases} \tag{9.60}$$

Consequently, the received signals are (see Fig. 9.11),

$$Y_1 = \sum_{i=1}^{K-M} \sum_{j=1,j\neq i}^{K-M} \frac{g_j h_i}{g_j h_j} V_{ij} + \sum_{i=K-M+1}^{K} \sum_{j=1}^{K-M} \frac{g_j h_i}{g_j h_j} V_{ij} + \sum_{k=1}^{K-M} U_k + N_1 \tag{9.61}$$

$$Y_2 = \sum_{i=1}^{K-M} \sum_{j=1,j\neq i}^{K-M} \frac{g_j}{h_j} V_{ij} + \sum_{i=K-M+1}^{K-M} \sum_{j=1}^{K-M} \frac{g_j}{h_j} V_{ij} + \sum_{j=1}^{K-M} \frac{g_j}{h_j} U_j + N_2 \tag{9.62}$$

$$= \sum_{j=1}^{K-M} \frac{g_j}{h_j} \left(U_j + \sum_{i=1,i\neq j}^{K-M} V_{ij} + \sum_{i=K-M+1}^{K} V_{ij} \right) + N_2 \tag{9.63}$$

Let $\mathbf{V} = \{V_{ij} : i = 1, \ldots K, j = 1, \ldots, K-M, i \neq j\}$. We evaluate the secrecy rates using (9.50), after choosing,

$$Q = P^{\frac{1-\delta}{2(K-M)(K-1)+1+\delta}}, \quad a = \gamma \frac{P^{1/2}}{Q} \tag{9.64}$$

The components of $\mathbf{V}$ are received in different dimensions, and hence we have $(2Q+$

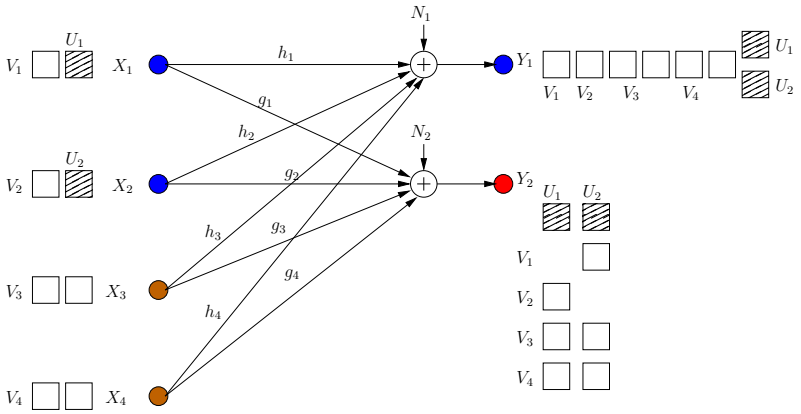


Figure 9.11: All users reduce rates to have zero leakage s.d.o.f.

$1)^{(K-M)(K-M-1)+M(K-M)} = (2Q+1)^{(K-M)(K-1)}$ separable constellation points, while the cooperative jamming signals are aligned in the same dimension, i.e., $(2(K-M)Q+1)$ constellation points. Thus,

$$I(\mathbf{V};Y_1) \geq I(\mathbf{V};\hat{\mathbf{V}}) \tag{9.65}$$

$$= \frac{(K-M)(K-1)(1-\delta)}{(K-M)(K-1)+1+\delta} \cdot \frac{1}{2}\log P + o(\log P) \tag{9.66}$$

Since all message-carrying signals are jammed by cooperative jamming signals, we have $K-M$ dimensions with $(2KQ+1)^{(K-M)}$ overlapping constellation points. Thus,

$$I(\mathbf{V};Y_2) \leq H(Y_2-N_2) - H(Y_2-N_2|\mathbf{V}) \tag{9.67}$$

$$= H\left(\sum_{j=1}^{K-M}\frac{g_j}{h_j}\left(U_j + \sum_{i=1,i\neq j}^{K-M} V_{ij} + \sum_{i=K-M+1}^{K} V_{ij}\right)\right) - H\left(\sum_{j=1}^{K-M}\frac{g_j}{h_j}U_j\right) \tag{9.68}$$

$$= (K-M)\log\frac{2KQ+1}{2Q+1} \tag{9.69}$$

$$\leq (K-M)\log K \tag{9.70}$$

Substituting (9.66) and (9.70) into (9.50), and taking the limit as $P \to \infty$, the achievable sum s.d.o.f. is,

$$d_s \geq \frac{(K-M)(K-1)}{(K-M)(K-1)+1} \tag{9.71}$$

The resultant sum s.d.o.f. is less than the optimal in [78]. However, interestingly, the individual s.d.o.f. of each deviating user is $\frac{K-M}{(K-M)(K-1)+1}$, which is larger than its s.d.o.f. without deviation $\frac{K-1}{K(K-1)+1}$, so long as $M \leq K-1+\frac{1}{K}$, i.e., if at least one user sticks to the optimal strategy in [78].

9.4.3.2 Reducing the Leakage to a Single Dimension

In this achievable scheme, we allow one rational dimension to be leaked. This dimension is not secured by a cooperative jamming signal. This results in the ability of injecting an extra message-carrying signal component for each user. All these extra signals are aligned in the same rational dimension at the eavesdropper. The transmitted signals are (see Fig. 9.12),

$$X_i = \begin{cases} \sum_{j=1,j\neq i}^{K-M} \frac{g_j}{g_i h_j} V_{ij} + \frac{\alpha}{h_i} V_{i0} + \frac{1}{h_i} U_i, i = 1, \ldots, K-M \\ \sum_{j=1}^{K-M} \frac{g_j}{g_i h_j} V_{ij} + \frac{\alpha}{h_i} V_{i0}, \quad i = K-M+1, \ldots, K \end{cases} \tag{9.72}$$

where α is rationally independent from all channel gains. The received signals are shown in Fig. 9.12. By similar steps, we have the following s.d.o.f. for this scheme,

$$d_s \geq \frac{(K-M)^2 + M(K-M+1) - 1}{(K-M)^2 + M(K-M+1) + 1} \tag{9.73}$$

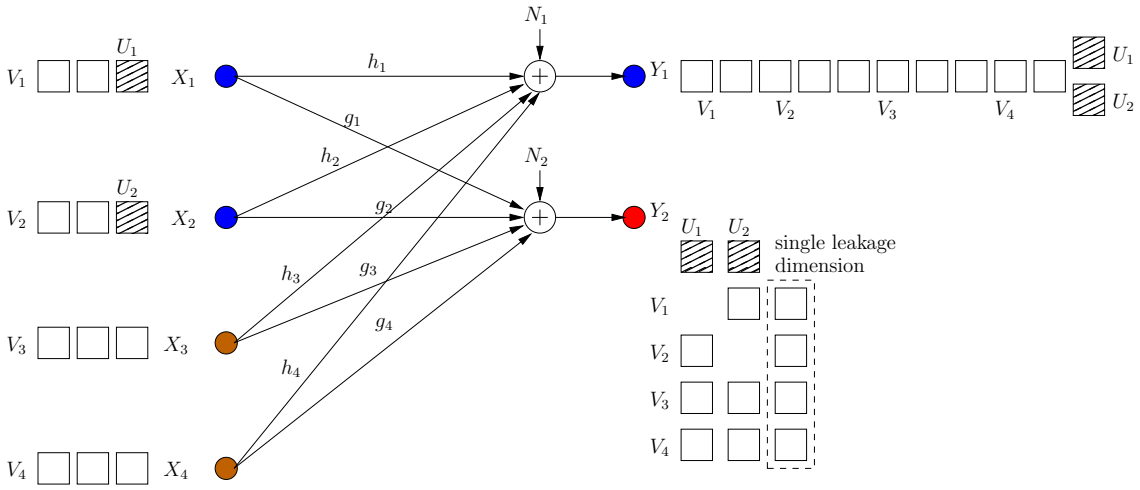


Figure 9.12: All users reduce the leakage dimension to 1.

Although the sum s.d.o.f. in this case is smaller than in (9.71), the individual s.d.o.f. of a well-behaving user is higher and a deviating user is lower than in (9.71).

9.4.4 Malicious Deviation: Intentional Jamming

In this section, we consider a more severe form of deviation, where a user (say the Kth user) sends intentional jamming signals. The deviating (malicious) user is restricted to use structured signals. In this section, we show that, when the malicious user acts, it can drive the sum s.d.o.f. to zero. However, when the remaining users respond, the sum s.d.o.f. is raised to $d_s = \frac{(K-1)^2}{(K-1)^2+1}$, which is the sum s.d.o.f. of a $K-1$ user MAC-WTC with an external altruistic helper.

9.4.4.1 When the Jammer Responds to the Users

In any encoding frame, each user sends its message-carrying signals V_{ij} on N rationally independent dimensions α_{ij} as,

$$X_i[k] = \sum_{j=1}^{N} \alpha_{ij} V_{ij} \tag{9.74}$$

Then, the jammer designs structured jamming signals $\tilde{U}_{ij}$ as a response to users' signals as,

$$Z[k] = \sum_{i=1}^{K-1} \sum_{j=1}^{N} \frac{\alpha_{ij} h_i}{\tilde{h}} \tilde{U}_{ij} \tag{9.75}$$

Consequently, the received signal at the legitimate receiver is,

$$Y_1[k] = \sum_{i=1}^{K-1} \sum_{j=1}^{N} h_i \alpha_{ij} (V_{ij} + \tilde{U}_{ij}) + N_1[k] \tag{9.76}$$

Hence, each message-carrying signal is aligned with a jamming signal. Let $\mathbf{V}[k] = [V_{ij}, i = 1, \ldots, K-1, j = 1, \ldots, N]^T$ to be vectorization of all secure signal components. Then, the secure rate is upper bounded as,

$$\sum_{i=1}^{K-1} R_i \leq I(\mathbf{V}[k]; Y_1[k] - N_1[k]) \tag{9.77}$$

$$= \sum_{i=1}^{K-1} \sum_{j=1}^{N} H(V_{ij} + \tilde{U}_{ij}) - H(\tilde{U}_{ij}) \tag{9.78}$$

$$\leq \sum_{i=1}^{K-1} \sum_{j=1}^{N} \log(4Q+1) - \log(2Q+1) \tag{9.79}$$

$$\leq N(K-1) = o(\log P) \tag{9.80}$$

Hence $d_s = 0$, i.e., whenever the jammer knows the signalling scheme of the users, it nulls the communication by jamming.

9.4.4.2 When the Users Respond to the Jammer

Since structured jamming signalling suffices to jam the system, the jammer sends structured signals in N dimensions,

$$Z[k] = \sum_{j=1}^{N} \alpha_j \tilde{U}_j \tag{9.81}$$

Users make use of the generated jamming signals to hide extra secure signals from the eavesdropper. Users send,

$$X_i[k] = \sum_{j=1}^{N} \sum_{l=1, l\neq i}^{K-1} \frac{\alpha_j \tilde{h} g_l}{g_i h_i} V_{ijl} + \sum_{j=1}^{N} \frac{\alpha_j \tilde{g}}{g_i} V_{ij0} + \sum_{j=1}^{N} \frac{\alpha_j \tilde{h}}{h_i} U_{ij} \tag{9.82}$$

where V_{ijl}, V_{ij0} are the message-carrying signals which are protected by cooperative jamming signals generated by other users, and the jamming signals generated by the malicious user, respectively. Then, the received signal at receiver 1 is,

$$Y_1[k] = \sum_{j=1}^{N} \left(\sum_{i=1}^{K-1} \sum_{l=1, l\neq i}^{K-1} \frac{\alpha_j \tilde{h} g_l h_i}{g_i} V_{ijl} + \sum_{i=1}^{K-1} \frac{\alpha_j \tilde{g} h_i}{g_i} V_{ij0} + \alpha_j \tilde{h} \left(\tilde{U}_j + \sum_{i=1}^{K-1} U_{ij} \right) \right) + N_1 \tag{9.83}$$

i.e., users' jamming signals use the same dimensions as the external jammer to inject extra cooperative jamming signals. The received signal at the eavesdropper is,

$$Y_2[k] = \sum_{i=1}^{K-1} g_i \left[\sum_{j=1}^{N} \sum_{l=1, l\neq i}^{K-1} \frac{\alpha_j \tilde{h} g_l}{g_i h_i} V_{ijl} + \sum_{j=1}^{N} \frac{\alpha_j \tilde{g}}{g_i} V_{ij0} + \sum_{j=1}^{N} \frac{\alpha_j \tilde{h}}{h_i} U_{ij} \right] + \tilde{g} \sum_{j=1}^{N} \alpha_j \tilde{U}_j + N_2 \tag{9.84}$$

$$= \sum_{j=1}^{N} \left[\alpha_j \tilde{g} \left(\sum_{i=1}^{K-1} V_{ij0} + \tilde{U}_j \right) + \sum_{l=1}^{K-1} \frac{\alpha_j \tilde{h} g_l}{h_l} \left(U_{ij} + \sum_{i=1, i\neq l}^{K-1} V_{lji} \right) \right] + N_2 \tag{9.85}$$

i.e., all message-carrying signals are protected from the eavesdropper, as in Fig. 9.13, with $K = 4$, $N = 1$.

We note that the received signals at receiver Y_1 consists of $(2Q + 1)^{N(K-1)(K-2)+N(K-1)}(2NKQ + 1)$ constellation points in $N((K-1)^2 + 1)$ dimensions. Each user is transmitting using PAM constellation $C(a, Q)$. By choosing $Q = P^{\frac{1-\delta}{2N((K-1)^2+1)+\delta}}$ and $a = \gamma P^{\frac{1}{2}}/Q$, we have

$$I(\mathbf{V}; Y_1[k]) \geq \frac{N(K-1)^2(1-\delta)}{N((K-1)^2+1)+\delta} \left(\frac{1}{2} \log P \right) + o(\log P) \tag{9.86}$$

Further, since every message-carrying signal is protected by a cooperative jamming signal, $I(\mathbf{V}; Y_2[k]) \leq o(\log P)$. Thus, the achievable sum s.d.o.f. with one malicious jammer when users respond is $d_s(k) = \frac{(K-1)^2}{(K-1)^2+1}$. Finally, in the Appendix, we determine the sum s.d.o.f. of a K-user MAC-WTC with M external altruistic helpers, as a result on its own. We note that this $d_s(k)$ is in fact equal to the sum s.d.o.f. of a $K-1$ user MAC-WTC with 1 external helper, concluding that the users' action to the jammer is optimal, as they achieve the s.d.o.f. of the case of an altruistic helper

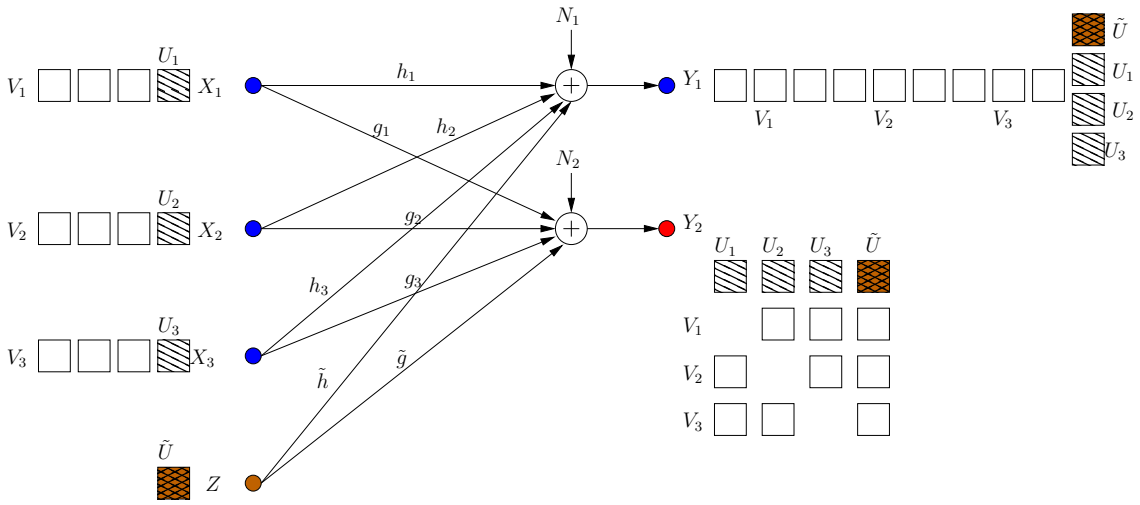


Figure 9.13: A malicious jamming user: users' response.

with a malicious jammer.

9.5 Conclusions

We introduced three new channel models, namely, BCCM with combating helpers, ICCM with selfish users, and MAC-WTC with deviating users. These new models aimed at studying the effects of selfishness and malicious behaviour on the secure rate in networks. We investigated the achievable s.d.o.f. in these models. The presented schemes are only achievable, new role-based converse arguments are needed.

For the BCCM with combating helpers, we formulated the problem as an extensive-form game. We assumed that each helper wants to minimize the s.d.o.f. of the other receiver without sacrificing the s.d.o.f. of its receiver, and analyzed schemes that employs recursive real interference alignment. In this case, we showed that the malicious behaviours of the combating helpers are neutralized and the s.d.o.f. of both users converge to $1/2$, as in the case of altruistic helpers.

Next, For the ICCM with selfish users, we changed the objective function of the users to maximizing the difference of the s.d.o.f. between the two users. By similar analysis to BCCM, we showed that the selfishness precludes any secure communication, and the s.d.o.f. of two users converge to zero.

Finally, for the MAC-WTC with deviating users, we considered two types of deviation: First, in the case when some of the users stopped transmitting cooperative jamming signals as in the optimal scheme, we evaluated the corresponding s.d.o.f. and proposed counter-strategies to respond to the deviation. Second, we investigated an extreme form of deviation, where a user sends intentional jamming signals. We showed that although a deviating user can drive the sum s.d.o.f. to zero, the jamming signals can be exploited as cooperative jamming signals against the eavesdropper to achieve an optimum s.d.o.f.

9.6 Appendix: K-User MAC-WTC with M External Helpers

Theorem 9.3 *The s.d.o.f. of the K-user Gaussian MAC-WTC with M-external helpers is given by* $d_s = \frac{K(K+M-1)}{K(K+M-1)+1}$.

Proof: We give only a sketch of a proof as it follows standard arguments. For the achievability, each user sends $K + M - 1$ message-carrying signals and one cooperative jamming signal to secure the other users. Each helper sends one cooperative jamming signal. The cooperative jamming signals are aligned in the same rational dimension at the receiver.

For the converse, we rely on the techniques in [78]. First, we have the following

upper bound which represents the *secrecy penalty* due to the secrecy constraint on the eavesdropper,

$$n\sum_{i=1}^{K} R_i \leq \sum_{l=2}^{K} h(\tilde{\mathbf{X}}_l) + \sum_{j=1}^{M} h(\tilde{\mathbf{Z}}_j) + nc_1 \tag{9.87}$$

where $\tilde{\mathbf{X}}_i, \tilde{\mathbf{Z}}_j$ are the perturbed inputs of user i and helper j, respectively. Next, we have the *role of the external helper(s)*,

$$\sum_{j=1}^{M} h(\tilde{\mathbf{Z}}_j) \leq Mh(\mathbf{Y}_1) - nM\sum_{i=1}^{K} R_i + nc_2 \tag{9.88}$$

By considering the rates of all users except one for the $K-1$ users, we have *role of the internal helper(s)*,

$$\sum_{l=2}^{K} h(\tilde{\mathbf{X}}_l) \leq (K-1)h(\mathbf{Y}_1) - n\sum_{l=2}^{K}\sum_{i\neq l} R_i + nc_3 \tag{9.89}$$

We substitute (9.88) and (9.89) in (9.87) to have,

$$n\big(R_1 + (M+K-1)\sum_{i=1}^{K} R_i\big) \leq (M+K-1)h(\mathbf{Y}_1) + nc_4 \tag{9.90}$$

We have written (9.87) by eliminating the first user's channel input, hence the summation starting at $i=2$. This inequality holds when any other user's channel input is chosen. Writing (9.87) for all K users, and adding the K corresponding bounds,

$$n(K(K+M-1)+1)\sum_{i=1}^{K} R_i$$

$$\leq K(K+M-1)\left(\frac{n}{2}\log P\right)+nc_5 \tag{9.91}$$

Taking the limit as $P\to\infty$, we have $d_s \leq \frac{K(K+M-1)}{K(K+M-1)+1}$. ■

Note that this result is related to the s.d.o.f. region result in [82] for the $K+M$ user MAC-WTC, when we focus on the hyperplane corresponding to zero s.d.o.f. for M of the users; these M users essentially serve as helpers.

CHAPTER 10

MIMO Wiretap Channel under Receiver Side Power Constraints

10.1 Introduction

In this chapter, we consider the MIMO wiretap channel under a minimum receiver-side power constraint in addition to the usual maximum transmitter-side power constraint. This problem is motivated by energy harvesting communications with wireless energy transfer, where an added goal is to deliver a minimum amount of energy to a receiver in addition to delivering secure data to another receiver. In this chapter, we characterize the exact secrecy capacity of the MIMO wiretap channel under transmitter and receiver-side power constraints. We first show that solving this problem is equivalent to solving the secrecy capacity of the wiretap channel under a *double-sided correlation matrix* constraint on the channel input. We show the converse by extending the channel enhancement technique to our case. We present two achievable schemes that achieve the secrecy capacity: the first achievable scheme uses a Gaussian codebook with a fixed mean, and the second achievable scheme uses artificial noise (or cooperative jamming) together with a Gaussian codebook. The

role of the mean or the artificial noise is to enable energy transfer without sacrificing from the secure rate. This is the first instance of a channel model where either the use of a mean signal or the use of channel prefixing via artificial noise is *strictly necessary* for the MIMO wiretap channel. We then extend our work to consider a maximum receiver-side power constraint instead of a minimum receiver-side power constraint. This problem is motivated by cognitive radio applications, where an added goal is to decrease the received signal energy (interference temperature) at a receiver. We further extend our results to: requiring receiver-side power constraints at both receivers; considering secrecy constraints at both receivers to study broadcast channels with confidential messages; and removing the secrecy constraints to study the classical broadcast channel.

10.2 System Model, Preliminaries and the Main Result

The MIMO wiretap channel with N_t antennas at the transmitter, N_r antennas at the legitimate receiver and N_e antennas at the eavesdropper is given by (see Fig. 10.1),

$$\mathbf{Y}_i = \mathbf{H}\mathbf{X}_i + \mathbf{W}_{1,i} \tag{10.1}$$

$$\mathbf{Z}_i = \mathbf{G}\mathbf{X}_i + \mathbf{W}_{2,i} \tag{10.2}$$

where $\mathbf{X}_i \in \mathbb{R}^{N_t}$ is the channel input, $\mathbf{Y}_i \in \mathbb{R}^{N_r}$ is the legitimate receiver's channel output, and $\mathbf{Z}_i \in \mathbb{R}^{N_e}$ is the eavesdropper's channel output at channel use i; $\mathbf{W}_{1,i}$ and $\mathbf{W}_{2,i}$ are independent Gaussian random vectors $\mathcal{N}(\mathbf{0}, \mathbf{I})$. The channel matrices of legitimate receiver $\mathbf{H}$ and the eavesdropper $\mathbf{G}$ are real-valued matrices of dimen-

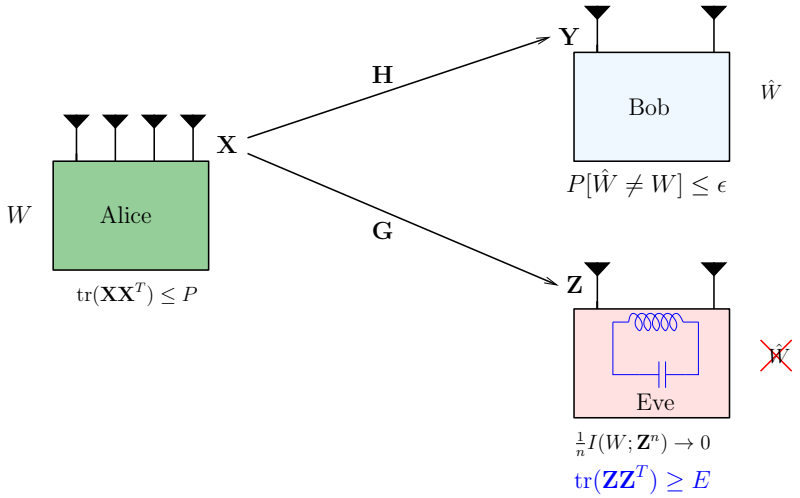


Figure 10.1: Gaussian MIMO wiretap channel with receiver-side power constraint.

sions $N_r \times N_t$ and $N_e \times N_t$, respectively, and are fixed and known to all entities. The transmitter encodes a message W picked from a discrete message set $\mathcal{W}$ to a codeword $\mathbf{X}^n$ over n channel uses via a stochastic encoder $f : \mathcal{W} \to \mathbf{X}^n$. The channel input is constrained by the usual *maximum* average power constraint [47], [128]:

$$\frac{1}{n} \sum_{i=1}^{n} \mathrm{tr}(\mathbf{X}_i \mathbf{X}_i^T) \leq P \tag{10.3}$$

In this chapter, we consider *minimum* and *maximum* power constraints at the receivers. In the initial part of the chapter, we consider a *minimum* power constraint at the eavesdropper only as:

$$\frac{1}{n} \sum_{i=1}^{n} \mathrm{tr}(\mathbf{Z}_i \mathbf{Z}_i^T) \geq E \tag{10.4}$$

As usual, see [47, 128], the actual power constraints in (10.3) and (10.4) will be reflected in the single-letter capacity expressions in the sequel as expectations, i.e.,

$\text{tr}(\mathbb{E}[\mathbf{X}\mathbf{X}^T]) \leq P$ and $\text{tr}(\mathbb{E}[\mathbf{Z}\mathbf{Z}^T]) \geq E$. In addition, for all $\epsilon_n > 0$, we have the following asymptotic reliability and secrecy constraints on W based on n-length observations $\mathbf{Y}^n, \mathbf{Z}^n$ at the receiver and the eavesdropper, respectively:

$$\mathbb{P}[\hat{W} \neq W] \leq \epsilon_n, \quad \lim_{n\to\infty} \frac{1}{n} I(W; \mathbf{Z}^n) = 0 \tag{10.5}$$

where $\epsilon_n \to 0$ as $n \to \infty$, and $\hat{W} = \phi(\mathbf{Y}^n)$ is the estimate of the legitimate receiver of the transmitted message W based on $\mathbf{Y}^n$ by using a decoder $\phi(\cdot)$.

In this case, we have an achievable rate $R_s(E, P, \mathbf{H}, \mathbf{G}) = \lim_{n\to\infty} \frac{1}{n} \log |\mathcal{W}|$ if there exists a code, i.e., a codebook and (f, ϕ) pair such that constraints (10.3)-(10.5) are satisfied. The secrecy capacity $C(E, P, \mathbf{H}, \mathbf{G}) = \sup\ R(E, P, \mathbf{H}, \mathbf{G})$, i.e., the supremum of all achievable rates. Although, we will determine the secrecy capacity under the maximum transmitter-side power constraint in (10.3) and the minimum receiver-side power constraint in (10.4), we initially characterize $C(\mathbf{S}_1, \mathbf{S}_2, \mathbf{H}, \mathbf{G})$, the secrecy capacity, under a general *double-sided correlation matrix constraint*:

$$\mathbf{S}_1 \preceq \mathbf{Q} \preceq \mathbf{S}_2 \tag{10.6}$$

where $\mathbf{Q} = \mathbb{E}[\mathbf{X}\mathbf{X}^T]$ is the channel input correlation matrix, and $\mathbf{S}_1 \preceq \mathbf{S}_2$ are given and fixed positive semi-definite (PSD) matrices, where $\preceq$ denotes the partial ordering of PSD matrices. We will show in a similar way to [102, Section II.B] that the secrecy capacity with power constraints of (10.3)-(10.4) can be obtained from the secrecy capacity with the more general double-sided correlation matrix constraint

in (10.6) by maximizing this secrecy capacity over all correlation matrices $\mathbf{S}_1 \preceq \mathbf{S}_2$ that lie in the compact set $\mathcal{S}_{PE}$:

$$\mathcal{S}_{PE} = \{\mathbf{S} \succeq \mathbf{0} : \mathrm{tr}(\mathbf{S}) \leq P, \quad \mathrm{tr}(\mathbf{G}\mathbf{S}\mathbf{G}^T) \geq \tilde{E}\} \tag{10.7}$$

where $\tilde{E} = E - N_e$. We evaluate the secrecy capacity based on Csiszar-Korner secrecy capacity expression [67]

$$C_s = \max_{V \to \mathbf{X} \to \mathbf{Y},\mathbf{Z}} I(V;\mathbf{Y}) - I(V;\mathbf{Z}) \tag{10.8}$$

where V carries the message signal and $\mathbf{X}$ is the channel input. The maximization is over all jointly distributed $(V, \mathbf{X})$ that satisfy the Markov chain $V \to \mathbf{X} \to \mathbf{Y}, \mathbf{Z}$ and the constraints (10.3), (10.4). Note that although Csiszar-Korner expression is initially given for discrete alphabets, it can be directly extended to alphabets other than discrete, by including the appropriate cost function in the maximization problem; see remarks in [67, Section VI]. This extension can be done via discrete approximations in [129, Chapter 3] and [130, Chapter 7].

The main result of this chapter is the exact characterization of the secrecy capacity of the MIMO wiretap channel under the maximum transmitter-side power constraint in (10.3) and the minimum receiver-side power constraint in (10.4). This result is stated in Theorem 10.1 below. We dedicate Section 10.3 for the achievability proof and Section 10.4 for the converse proof of this theorem. In Section 10.5, we extend this basic proof technique to the cases of: minimum receiver-side power

constraints at both receivers; maximum receiver-side power constraints; no secrecy constraints (classical BC); and double-sided secrecy constraints (BCCM).

Theorem 10.1 *The secrecy capacity of a MIMO wiretap channel with a transmitter-side power constraint P and a receiver-side power constraint E, $C(E, P, \mathbf{H}, \mathbf{G})$, is given as*

$$\begin{aligned} & C(E, P, \mathbf{H}, \mathbf{G}) \\ & \quad = \max_{\mathbf{Q} \succeq \mathbf{0}} \quad \frac{1}{2} \log |\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^T| - \frac{1}{2} \log |\mathbf{I} + \mathbf{G}\mathbf{Q}\mathbf{G}^T| \\ & \qquad s.t. \quad \mathrm{tr}(\mathbf{Q} + \boldsymbol{\mu}\boldsymbol{\mu}^T) \leq P \\ & \qquad\quad \mathrm{tr}(\mathbf{G}(\mathbf{Q} + \boldsymbol{\mu}\boldsymbol{\mu}^T)\mathbf{G}^T) \geq \tilde{E} \end{aligned} \tag{10.9}$$

where $\tilde{E} = E - N_e$. This secrecy capacity is achieved by $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{Q})$, i.e., with a mean but no channel prefixing. Alternatively, the secrecy capacity, $C(E, P, \mathbf{H}, \mathbf{G})$, is also given as

$$\begin{aligned} & C(E, P, \mathbf{H}, \mathbf{G}) \\ & \quad = \max_{\mathbf{Q}_1, \mathbf{Q}_2 \succeq \mathbf{0}} \quad \frac{1}{2} \log \frac{|\mathbf{I} + \mathbf{H}(\mathbf{Q}_1 + \mathbf{Q}_2)\mathbf{H}^T|}{|\mathbf{I} + \mathbf{H}\mathbf{Q}_2\mathbf{H}^T|} \\ & \qquad\qquad\quad - \frac{1}{2} \log \frac{|\mathbf{I} + \mathbf{G}(\mathbf{Q}_1 + \mathbf{Q}_2)\mathbf{G}^T|}{|\mathbf{I} + \mathbf{G}\mathbf{Q}_2\mathbf{G}^T|} \\ & \qquad\quad s.t. \quad \mathrm{tr}(\mathbf{Q}_1 + \mathbf{Q}_2) \leq P \\ & \qquad\qquad\quad \mathrm{tr}(\mathbf{G}(\mathbf{Q}_1 + \mathbf{Q}_2)\mathbf{G}^T) \geq \tilde{E} \end{aligned} \tag{10.10}$$

where $\mathbf{X} = \mathbf{V} + \mathbf{U}$, with jointly Gaussian $\mathbf{V} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_1)$ and $\mathbf{U} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_2)$,

and $\mathbf{V}, \mathbf{U}$ *are independent, i.e., with Gaussian signalling with Gaussian channel prefixing.*

10.3 Achievability Schemes

In this section, we provide two coding schemes that achieve the secrecy capacity of the MIMO wiretap channel with transmitter and receiver-side power constraints given in Theorem 10.1.

10.3.1 Gaussian Coding with Fixed Mean

The first achievable scheme is Gaussian coding with fixed mean, i.e., $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}_1)$. In this case, the fixed mean does not play a role in evaluating the secrecy capacity except for consuming part of the overall correlation matrix and only provides the required power level at the receiver side. Then, we choose $V = \mathbf{X}$, i.e., no channel prefixing. Hence, we have

$$\begin{aligned} & C(\mathbf{S}_1, \mathbf{S}_2, \mathbf{H}, \mathbf{G}) \\ & \quad \geq \max_{\mathbf{Q}_1 \succeq \mathbf{0}, \boldsymbol{\mu}} \quad I(\mathbf{X}; \mathbf{Y}) - I(\mathbf{X}; \mathbf{Z}) \\ & \quad = \max_{\mathbf{Q}_1 \succeq \mathbf{0}, \boldsymbol{\mu}} \quad \frac{1}{2} \log |\mathbf{I} + \mathbf{H}\mathbf{Q}_1\mathbf{H}^T| - \frac{1}{2} \log |\mathbf{I} + \mathbf{G}\mathbf{Q}_1\mathbf{G}^T| \\ & \qquad \text{s.t.} \quad \mathbf{S}_1 \preceq \mathbf{Q}_1 + \boldsymbol{\mu}\boldsymbol{\mu}^T \preceq \mathbf{S}_2 \end{aligned} \tag{10.11}$$

In the converse proof, in place of $\boldsymbol{\mu}\boldsymbol{\mu}^T$, we have a general positive semidefinite matrix $\mathbf{Q}_2$. In order to have a matching feasible coding scheme, $\mathbf{Q}_2$ must be constrained

to unit-rank correlation matrices, as it corresponds to the mean of the transmitted signal. Although, the solution of $\mathbf{Q}_2$ is generally not unit-rank for arbitrary correlation matrices $\mathbf{S}_1, \mathbf{S}_2$, we show in the following lemma that for the special case of a maximum transmitter-side power constraint P and a minimum receiver-side power constraint E, the solution is guaranteed to be of unit-rank, and hence the mean based coding scheme is feasible.

Lemma 10.1 *The coding scheme* $\mathbf{X} \sim \mathcal{N}(\mathbb{V}(\mathbf{Q}_2^*), \mathbf{Q}_1^*)$ *is achievable for the wiretap channel under the transmitter-side power constraint* P *and the receiver-side power constraint* E *given that the matrix* $\mathbf{G}^T\mathbf{G}$ *has a unique maximum eigenvalue. The secrecy rate is characterized by the following optimization problem:*

$$\begin{aligned} \max_{\mathbf{Q}_1, \mathbf{Q}_2 \succeq \mathbf{0}} \quad & \frac{1}{2}\log|\mathbf{I} + \mathbf{H}\mathbf{Q}_1\mathbf{H}^T| - \frac{1}{2}\log|\mathbf{I} + \mathbf{G}\mathbf{Q}_1\mathbf{G}^T| \\ s.t. \quad & \mathrm{tr}(\mathbf{Q}_1 + \mathbf{Q}_2) \leq P \\ & \mathrm{tr}(\mathbf{G}(\mathbf{Q}_1 + \mathbf{Q}_2)\mathbf{G}^T) \geq \tilde{E} \end{aligned} \tag{10.12}$$

where $\mathbf{Q}_1^*, \mathbf{Q}_2^*$ *are the optimal correlation matrices for (10.12) and* $\mathbb{V}(\mathbf{Q}_2^*)$ *denotes the unique eigenvector of matrix* $\mathbf{Q}_2^*$ *with a non-zero eigenvalue.*

Proof: We note that $\mathbf{Q}_2$ does not appear in the objective function; it only appears in the constraint set. Therefore, its only role is to enlarge the feasible set for $\mathbf{Q}_1$ subject to some power constraint $\tilde{P}$, where $\tilde{P} \leq P$. Thus, $\mathbf{Q}_2$ must be chosen such that, when the first constraint of (10.12) is fixed, it maximizes the feasible set for

$\mathbf{Q}_1$ in the second constraint, i.e., $\mathbf{Q}_2$ must be the solution of

$$\max_{\mathbf{Q}_2 \succeq \mathbf{0}} \quad \text{tr}(\mathbf{G}\mathbf{Q}_2\mathbf{G}^T) \quad \text{s.t.} \quad \text{tr}(\mathbf{Q}_2) = \tilde{P} \tag{10.13}$$

The eigenvector decomposition for $\mathbf{Q}_2$, which is symmetric, is

$$\mathbf{Q}_2 = \sum_{i=1}^{r} \lambda_i \mathbf{q}_i \mathbf{q}_i^T \tag{10.14}$$

where r, λ_i, $\mathbf{q}_i$ are the rank, the ith eigenvalue and the corresponding orthonormal eigenvector of $\mathbf{Q}_2$, respectively. Thus, we can write the constraint as $\text{tr}(\mathbf{Q}_2) = \sum_{i=1}^{r} \lambda_i = \tilde{P}$. Moreover, the objective function can be written as

$$\text{tr}(\mathbf{G}\mathbf{Q}_2\mathbf{G}^T) = \text{tr}\left(\mathbf{G}\left(\sum_{i=1}^{r} \lambda_i \mathbf{q}_i \mathbf{q}_i^T\right)\mathbf{G}^T\right) \tag{10.15}$$

$$= \sum_{i=1}^{r} \lambda_i \|\mathbf{G}\mathbf{q}_i\|^2 \tag{10.16}$$

Hence, the optimization problem in (10.13) can be written as

$$\max_{\lambda_i, \mathbf{q}_i} \quad \sum_{i=1}^{r} \lambda_i \|\mathbf{G}\mathbf{q}_i\|^2 \quad \text{s.t.} \quad \sum_{i=1}^{r} \lambda_i = \tilde{P} \tag{10.17}$$

which is a linear program in λ_i. The optimum solution is $\lambda_m = \tilde{P}$, and $\lambda_i = 0$ for $i \neq m$, where

$$m = \arg\max_i \|\mathbf{G}\mathbf{q}_i\|^2 \tag{10.18}$$

Hence, the optimal solution for this problem is to beam-form all the available power $\tilde{P}$ to the direction of the largest $\|\mathbf{G}\mathbf{q}_i\|^2$. This solution is unique if $\mathbf{G}^T\mathbf{G}$ has a unique maximum eigenvalue. Otherwise a unit-rank solution for $\mathbf{Q}_2$ is not guaranteed. In this case, $\mathbf{Q}_2 = \tilde{P}\mathbf{q}_m\mathbf{q}_m^T$, i.e., it is unit-rank with eigenvector $\boldsymbol{\mu} = \sqrt{\tilde{P}}\mathbf{q}_m$, and the problem is feasible. ■

We remark that the same capacity expression in (10.12) can be realized by letting $\mathbf{X} = \mathbf{V} + \mathbf{U}$, where $\mathbf{V} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_1)$ is the message-carrying signal and $\mathbf{U} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_2)$ is the energy-carrying signal that is known causally at both Bob and Eve, so that it can be cancelled prior to information decoding. We note that, with this coding scheme any covariance matrix $\mathbf{Q}_2$ can be realized, and therefore Lemma 1 is not needed with this coding scheme, i.e., that the converse and achievability match for all $\mathbf{S}_1, \mathbf{S}_2$. However, if $\mathbf{Q}_2$ is optimized for this scheme as well for given P, E, then the optimum $\mathbf{Q}_2$ is still unit-rank. If the problem is considered under covariance constraints, as opposed to power constraints, unit-rank requirement of the mean based scheme can be removed by sending known Gaussian signals instead, at the cost of extra overhead of identifying $\mathbf{U}$ causally at Bob and Eve.

10.3.2 Gaussian Coding with Gaussian Artificial Noise

The second achievable scheme is Gaussian coding with Gaussian artificial noise. In this case, we choose $\mathbf{X} = \mathbf{V} + \mathbf{U}$, where $\mathbf{V}$, $\mathbf{U}$ are independent and $\mathbf{V} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_1)$ and $\mathbf{U} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_2)$. Here, $\mathbf{V}$ carries the message, $\mathbf{X}$ is the channel input, and $\mathbf{U}$ is the artificial noise (or cooperative jamming [111]) signal. In this case, we use

channel prefixing, hence $\mathbf{V} \neq \mathbf{X}$. The extra randomness $\mathbf{U}$ is sent by the transmitter to provide extra noise floor at both receivers, and confuses the eavesdropper. The added significance of this artificial noise in our problem is to provide a suitable level of received power at the receiver, i.e., we utilize the artificial noise as a source of power. In this case, the achievable secrecy rate satisfies

$$\begin{aligned} C(\mathbf{S}_1, \mathbf{S}_2, \mathbf{H}, \mathbf{G}) \geq & \max_{\mathbf{Q}_1, \mathbf{Q}_2 \succeq \mathbf{0}} \quad I(\mathbf{V}; \mathbf{Y}) - I(\mathbf{V}; \mathbf{Z}) \\ = & \max_{\mathbf{Q}_1, \mathbf{Q}_2 \succeq \mathbf{0}} \quad \frac{1}{2} \log \frac{|\mathbf{I} + \mathbf{H}(\mathbf{Q}_1 + \mathbf{Q}_2)\mathbf{H}^T|}{|\mathbf{I} + \mathbf{H}\mathbf{Q}_2\mathbf{H}^T|} \\ & \qquad - \frac{1}{2} \log \frac{|\mathbf{I} + \mathbf{G}(\mathbf{Q}_1 + \mathbf{Q}_2)\mathbf{G}^T|}{|\mathbf{I} + \mathbf{G}\mathbf{Q}_2\mathbf{G}^T|} \\ & \text{s.t.} \quad \mathbf{S}_1 \preceq \mathbf{Q}_1 + \mathbf{Q}_2 \preceq \mathbf{S}_2 \end{aligned} \tag{10.19}$$

10.4 Converse Proof

In this section, we prove the reverse implication using the channel enhancement technique [100,102]. We will consider the case of $\mathbf{S}_2 \succeq \mathbf{S}_1 \succ \mathbf{0}$ and the aligned MIMO setting which means that the channel matrices are square and invertible. The general MIMO case follows directly from the limiting arguments in [100], as the additional receiver-side power constraint is irrelevant in the limit. The idea of this limiting argument is to perform singular-value decomposition of the perturbed channels $\bar{\mathbf{H}}, \bar{\mathbf{G}}$ [100, Eqn. (37)]. Our result follows by taking the limit of this perturbation to zero. The argument is introduced in [100, Section II.B] and used for example in [72, Appendix B.2], [76, Section VII]. Therefore, we focus on the aligned case here. The aligned MIMO model is obtained by multiplying the input-output relations

(10.1)-(10.2) by the inverse of the channel matrices:

$$\tilde{\mathbf{Y}} = \mathbf{X} + \mathbf{H}^{-1}\mathbf{W}_1 = \mathbf{X} + \tilde{\mathbf{W}}_1 \tag{10.20}$$

$$\tilde{\mathbf{Z}} = \mathbf{X} + \mathbf{G}^{-1}\mathbf{W}_2 = \mathbf{X} + \tilde{\mathbf{W}}_2 \tag{10.21}$$

where $\tilde{\mathbf{W}}_1$ and $\tilde{\mathbf{W}}_2$ are the equivalent zero-mean Gaussian random vectors with covariance matrices $\mathbf{N}_1 = \mathbf{H}^{-1}\mathbf{H}^{-T}$ and $\mathbf{N}_2 = \mathbf{G}^{-1}\mathbf{G}^{-T}$, respectively.

10.4.1 Equivalence of a Double-Sided Correlation Matrix Constraint

For the MIMO broadcast and wiretap channels under a transmitter-side maximum power constraint, references [100, 102] showed that it is sufficient to prove the converse under a maximum correlation constraint on the channel input. We first note here that in our case with maximum transmitter-side and minimum receiver-side power constraints, a single correlation constraint on the channel input, i.e., $\mathbf{Q} \preceq \mathbf{S}$, is not sufficient. Next, we show the equivalence of solving our problem with a *double-sided* correlation matrix constraint on the channel input, i.e., $\mathbf{S}_1 \preceq \mathbf{Q} \preceq \mathbf{S}_2$. Then, our problem can be solved in two stages: the inner problem finds the capacity under fixed correlation matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ constraints, and the outer problem finds the optimal $\mathbf{S}_1, \mathbf{S}_2 \in \mathcal{S}_{PE}$ in (10.7). Finally, we modify the original channel enhancement technique [100, 102] to prove the optimality of the achievable schemes presented in the previous section.

We first note that solving the problem for $\mathbf{Q} \preceq \mathbf{S}$, where $\mathbf{S} \in \mathcal{S}_{PE}$ is insufficient. Consider solving the secrecy capacity under maximum transmitter-side and

minimum receiver-side power constraints in two stages, first, solving the problem under a fixed correlation matrix $\mathbf{S}$, and then choosing the optimal $\mathbf{S} \in \mathcal{S}_{PE}$, i.e.,

$$\max_{\mathbf{S}\in\mathcal{S}_{PE}} \quad \max_{\mathbf{Q}\preceq\mathbf{S}} \quad R_s(\mathbf{Q},\mathbf{H},\mathbf{G}) \tag{10.22}$$

where $R_s(\mathbf{Q},\mathbf{H},\mathbf{G})$ is the achievable secure rate upon using correlation matrix $\mathbf{Q}$. Since $\mathbf{Q} \preceq \mathbf{S}$, we have $\mathbf{G}\mathbf{Q}\mathbf{G}^T \preceq \mathbf{G}\mathbf{S}\mathbf{G}^T$ and hence $\mathrm{tr}(\mathbf{G}\mathbf{Q}\mathbf{G}^T) \leq \mathrm{tr}(\mathbf{G}\mathbf{S}\mathbf{G}^T)$. Then, although any $\mathbf{S} \in \mathcal{S}_{PE}$ satisfies the minimum receiver-side power constraint, i.e., $\mathrm{tr}(\mathbf{G}\mathbf{S}\mathbf{G}^T) \geq \tilde{E}$, the input correlation matrix $\mathbf{Q}$ is not guaranteed to satisfy $\mathrm{tr}(\mathbf{G}\mathbf{Q}\mathbf{G}^T) \geq \tilde{E}$. Hence, the single correlation constraint is not sufficient for solving problems involving minimum receiver-side power constraints.

Lemma 10.2 *Since $\mathcal{S}_{PE}$ is a compact set of PSD matrices, and $C(\mathbf{S}_1,\mathbf{S}_2,\mathbf{H},\mathbf{G})$ is continuous with respect to $\mathbf{S}_2$, we have*

$$C(E,P,\mathbf{H},\mathbf{G}) = \max_{\mathbf{S}_1,\mathbf{S}_2\in\mathcal{S}_{PE},\mathbf{S}_1\preceq\mathbf{S}_2} \quad C(\mathbf{S}_1,\mathbf{S}_2,\mathbf{H},\mathbf{G}) \tag{10.23}$$

Proof: We follow and extend the proof technique in [102, Lemma 1] to the case of double-sided covariance matrices. We define the wiretap code $\mathcal{C}(n,\mathbf{S},R,\epsilon)$ as a codebook, where the codewords $\{\mathbf{X}_i^n\}_{i=1}^{2^{nR}}$ are such that $\mathbf{S} = \frac{1}{2^{nR}}\sum_{i=1}^{2^{nR}} \mathbf{X}_i^n \mathbf{X}_i^{n^T}$, and accompanying encoding and decoding functions (f,ϕ), such that $\mathbb{P}(\phi(f(W)) \neq W) \leq \epsilon$. The decoder ϕ can be taken as the maximum likelihood decoder.

To see

$$C(E,P,\mathbf{H},\mathbf{G}) \geq \max_{\mathbf{S}_1,\mathbf{S}_2\in\mathcal{S}_{PE},\mathbf{S}_1\preceq\mathbf{S}_2} \quad C(\mathbf{S}_1,\mathbf{S}_2,\mathbf{H},\mathbf{G}) \tag{10.24}$$

we note that for any $\mathbf{S}_1 \preceq \mathbf{Q} \preceq \mathbf{S}_2$ where $\mathbf{S}_1, \mathbf{S}_2 \in \mathcal{S}_{PE}$, we have $\mathbf{Q} \in \mathcal{S}_{PE}$, i.e., every $\mathbf{Q}$ in the feasible set of the optimization problem on the right hand side belongs to the feasible set of the optimization problem $C(E,P,\mathbf{H},\mathbf{G})$. Hence, $C(E,P,\mathbf{H},\mathbf{G})$ is at least as large as $\max_{\mathbf{S}_1,\mathbf{S}_2\in\mathcal{S}_{PE},\mathbf{S}_1\preceq\mathbf{S}_2}\ C(\mathbf{S}_1,\mathbf{S}_2,\mathbf{H},\mathbf{G})$.

To see

$$C(E,P,\mathbf{H},\mathbf{G}) \leq \max_{\mathbf{S}_1,\mathbf{S}_2\in\mathcal{S}_{PE},\mathbf{S}_1\preceq\mathbf{S}_2} \quad C(\mathbf{S}_1,\mathbf{S}_2,\mathbf{H},\mathbf{G}) \tag{10.25}$$

we should prove that $C(E,P,\mathbf{H},\mathbf{G}) = C(\mathbf{S}_1,\mathbf{S}_2,\mathbf{H},\mathbf{G})$ for some $\mathbf{S}_1, \mathbf{S}_2 \in \mathcal{S}_{PE}$ [102]. If $R = C(E,P,\mathbf{H},\mathbf{G})$ is achievable, then there exists an infinite sequence of codes $\mathcal{C}(n_i, \mathbf{S}_{0_i}, R, \epsilon_i)$, $i = 1, \ldots$ with rate R and decreasing probability of error $\epsilon_i \to 0$ as $i \to \infty$. Choose $\mathbf{S}_1 \preceq \mathbf{S}_{0_i}$, $\forall i$ and $\mathbf{S}_1 \in \mathcal{S}_{PE}$. We note that the choice of $\mathbf{S}_1$ is completely arbitrary, thus without loss of generality, we can choose it to be the first element in the sequence, i.e., $\mathbf{S}_{0_1}$. As $\mathcal{S}_{PE}$ is compact [131,132], for any infinite sequence of points in $\mathcal{S}_{PE}$, there must exist a sub-sequence that converges to a point $\mathbf{S}_0 \in \mathcal{S}_{PE}$. Hence, for any arbitrary $\delta > 0$, we can find an increasing subsequence $i(k)$ such that $\mathbf{S}_1 \preceq \mathbf{S}_{0_{i(k)}} \preceq \mathbf{S}_0 + \delta\mathbf{I}$.

This implies that we can find a sequence of codes $\mathcal{C}(n_k, \mathbf{S}_0 + \delta\mathbf{I}, R, \epsilon_k)$ with $\mathbf{S}_0 \in \mathcal{S}_{PE}$, $\mathbf{S}_0 \succeq \mathbf{S}_1$ achieving small probability of error. Therefore, for every $\delta > 0$,

we have $R = C(\mathbf{S}_1, \mathbf{S}_0 + \delta\mathbf{I}, \mathbf{H}, \mathbf{G})$. Since $C(\mathbf{S}_1, \mathbf{S}_0 + \delta\mathbf{I}, \mathbf{H}, \mathbf{G})$ is continuous, see Appendix 10.9, with respect to its second argument, we have that every ϵ-ball around R contains $C(\mathbf{S}_1, \mathbf{S}_0, \mathbf{H}, \mathbf{G})$, since for every $\epsilon > 0$, there exists $\delta > 0$ such that $C(\mathbf{S}_1, \mathbf{S}_0 + \delta\mathbf{I}, \mathbf{H}, \mathbf{G}) - C(\mathbf{S}_1, \mathbf{S}_0, \mathbf{H}, \mathbf{G}) < \epsilon$ as continuity asserts. Therefore R is a limit point of $\mathcal{C}(\mathbf{S}_1, \mathbf{S}_0, \mathbf{H}, \mathbf{G})$ and hence $C(E, P, \mathbf{H}, \mathbf{G}) = C(\mathbf{S}_1, \mathbf{S}_0, \mathbf{H}, \mathbf{G})$. This limit point belongs to $\mathcal{S}_{PE}$ since it is closed. ■

10.4.2 Converse Proof for Gaussian Coding with Fixed Mean

First, we begin with writing the equivalent optimization problem corresponding to the achievability scheme in the aligned MIMO case with Gaussian coding $\mathbf{X} \sim \mathcal{N}(\mathbb{V}(\mathbf{Q}_2^*), \mathbf{Q}_1^*)$:

$$\begin{aligned} \max_{\mathbf{Q}_1, \mathbf{Q}_2 \succeq \mathbf{0}} \quad & \frac{1}{2}\log\frac{|\mathbf{Q}_1 + \mathbf{N}_1|}{|\mathbf{N}_1|} - \frac{1}{2}\log\frac{|\mathbf{Q}_1 + \mathbf{N}_2|}{|\mathbf{N}_2|} \\ \text{s.t.} \quad & \mathbf{Q}_1 + \mathbf{Q}_2 \succeq \mathbf{S}_1, \quad \mathbf{Q}_1 + \mathbf{Q}_2 \preceq \mathbf{S}_2 \end{aligned} \tag{10.26}$$

The Lagrangian of this optimization problem can be written as:

$$\begin{aligned} \mathcal{L} = & \log\frac{|\mathbf{Q}_1 + \mathbf{N}_2|}{|\mathbf{N}_2|} - \log\frac{|\mathbf{Q}_1 + \mathbf{N}_1|}{|\mathbf{N}_1|} - \mathrm{tr}(\mathbf{Q}_1\mathbf{M}_1) - \mathrm{tr}(\mathbf{Q}_2\mathbf{M}_2) \\ & - \mathrm{tr}((\mathbf{Q}_1 + \mathbf{Q}_2 - \mathbf{S}_1)\mathbf{M}_3) + \mathrm{tr}((\mathbf{Q}_1 + \mathbf{Q}_2 - \mathbf{S}_2)\mathbf{M}_4) \end{aligned} \tag{10.27}$$

where $\mathbf{M}_1 \succeq \mathbf{0}, \mathbf{M}_2 \succeq \mathbf{0}, \mathbf{M}_3 \succeq \mathbf{0}$ and $\mathbf{M}_4 \succeq \mathbf{0}$ are the Lagrange multipliers for each constraint. The corresponding KKT complementary slackness conditions are:

$$\mathbf{Q}_1^* \mathbf{M}_1 = \mathbf{0}, \quad \mathbf{Q}_2^* \mathbf{M}_2 = \mathbf{0} \tag{10.28}$$

$$(\mathbf{Q}_1^* + \mathbf{Q}_2^* - \mathbf{S}_1)\mathbf{M}_3 = \mathbf{0} \tag{10.29}$$

$$(\mathbf{S}_2 - \mathbf{Q}_1^* - \mathbf{Q}_2^*)\mathbf{M}_4 = \mathbf{0} \tag{10.30}$$

and the KKT optimality conditions for $\mathbf{Q}_1^*$ and $\mathbf{Q}_2^*$ are:

$$(\mathbf{Q}_1^* + \mathbf{N}_2)^{-1} - (\mathbf{Q}_1^* + \mathbf{N}_1)^{-1} - \mathbf{M}_1 - \mathbf{M}_3 + \mathbf{M}_4 = \mathbf{0} \tag{10.31}$$

$$-\mathbf{M}_2 - \mathbf{M}_3 + \mathbf{M}_4 = \mathbf{0} \tag{10.32}$$

Now, using (10.31) and (10.32), we can construct an enhanced channel that can serve as an upper bound for the original legitimate receiver's channel, while the eavesdropper's channel is degraded with respect to it. The covariance of the enhanced channel is chosen as $\tilde{\mathbf{N}}$ such that

$$(\mathbf{Q}_1^* + \mathbf{N}_2)^{-1} + \mathbf{M}_2 = (\mathbf{Q}_1^* + \mathbf{N}_1)^{-1} + \mathbf{M}_1 = (\mathbf{Q}_1^* + \tilde{\mathbf{N}})^{-1} \tag{10.33}$$

Using this definition of the enhanced channel, we explore various characteristics of $\tilde{\mathbf{N}}$.

First, to prove the validity of the covariance matrix $\tilde{\mathbf{N}}$, we note that

$$\tilde{\mathbf{N}} = [(\mathbf{Q}_1^* + \mathbf{N}_1)^{-1} + \mathbf{M}_1]^{-1} - \mathbf{Q}_1^* \tag{10.34}$$

$$= (\mathbf{I} + \mathbf{N}_1\mathbf{M}_1)^{-1}(\mathbf{Q}_1^* + \mathbf{N}_1) - \mathbf{Q}_1^* \tag{10.35}$$

$$= (\mathbf{I} + \mathbf{N}_1\mathbf{M}_1)^{-1}[(\mathbf{Q}_1^* + \mathbf{N}_1) - (\mathbf{I} + \mathbf{N}_1\mathbf{M}_1)\mathbf{Q}_1^*] \tag{10.36}$$

$$= (\mathbf{I} + \mathbf{N}_1\mathbf{M}_1)^{-1}\mathbf{N}_1 \tag{10.37}$$

$$= (\mathbf{N}_1^{-1} + \mathbf{M}_1)^{-1} \succeq \mathbf{0} \tag{10.38}$$

and hence the covariance matrix of the constructed enhanced channel is positive semi-definite, and therefore it is a feasible covariance matrix.

Next, we want to show that the constructed channel is enhanced with respect to $\mathbf{N}_1$, i.e., $\mathbf{N}_1 \succeq \tilde{\mathbf{N}}$. To show that we note from (10.37) that $\tilde{\mathbf{N}} = (\mathbf{N}_1^{-1}+\mathbf{M}_1)^{-1}$ and hence, $\mathbf{N}_1 \succeq \tilde{\mathbf{N}}$. Similarly by considering $(\mathbf{Q}_1^* + \mathbf{N}_2)^{-1} + \mathbf{M}_2 = (\mathbf{Q}_1^* + \tilde{\mathbf{N}})^{-1}$ we note that $\mathbf{N}_2 \succeq \tilde{\mathbf{N}}$. Hence, we conclude that the enhanced channel has better channel conditions than the original legitimate user's channel, therefore, the constructed channel is an upper bound for the legitimate receiver. Moreover, the eavesdropper's channel is degraded with respect to the constructed channel. Consequently the secrecy capacity of the enhanced channel is known. In other words, we have $\tilde{\mathbf{Y}} = \mathbf{X} + \tilde{\mathbf{W}}$ such that $\tilde{\mathbf{W}} \sim \mathcal{N}(\mathbf{0}, \tilde{\mathbf{N}})$ and $\mathbf{X} \to \tilde{\mathbf{Y}} \to \mathbf{Y}$ and $\mathbf{X} \to \tilde{\mathbf{Y}} \to \mathbf{Z}$.

In order to have a meaningful upper bound, we need to show that the rate is preserved between the original problem and the constructed channel. To show that,

we have

$$(\mathbf{Q}_1^* + \tilde{\mathbf{N}})^{-1}\tilde{\mathbf{N}} = (\mathbf{Q}_1^* + \tilde{\mathbf{N}})^{-1}(\tilde{\mathbf{N}} + \mathbf{Q}_1^* - \mathbf{Q}_1^*) \tag{10.39}$$

$$= \mathbf{I} - (\mathbf{Q}_1^* + \tilde{\mathbf{N}})^{-1}\mathbf{Q}_1^* \tag{10.40}$$

$$= \mathbf{I} - [(\mathbf{Q}_1^* + \mathbf{N}_1)^{-1} + \mathbf{M}_1]\mathbf{Q}_1^* \tag{10.41}$$

$$= \mathbf{I} - (\mathbf{Q}_1^* + \mathbf{N}_1)^{-1}\mathbf{Q}_1^* \tag{10.42}$$

$$= (\mathbf{Q}_1^* + \mathbf{N}_1)^{-1}\mathbf{N}_1 \tag{10.43}$$

where (10.41) follows from the definition of the enhanced channel and (10.42) follows from the complementary slackness condition (10.28). Therefore, we have

$$\frac{|\tilde{\mathbf{N}} + \mathbf{Q}_1^*|}{|\tilde{\mathbf{N}}|} = \frac{|\mathbf{N}_1 + \mathbf{Q}_1^*|}{|\mathbf{N}_1|} \tag{10.44}$$

To show a similar rate preservation argument for the degraded channel $\mathbf{N}_2$, we will need the following lemma.

Lemma 10.3 *The optimal covariance matrix for the achievable scheme with Gaussian signaling with a fixed mean* $\mathbf{Q}_1^*$ *satisfies* $(\mathbf{S}_2 - \mathbf{Q}_1^*)\mathbf{M}_2 = \mathbf{0}$.

Proof: We return to the KKT conditions. Considering the correlation constraint, three cases can possibly occur. The first case: the correlation constraint is satisfied with equality, consequently $\mathbf{S}_2 - \mathbf{Q}_1^* = \mathbf{Q}_2^*$. In this case, $(\mathbf{S}_2 - \mathbf{Q}_1^*)\mathbf{M}_2 = \mathbf{Q}_2^*\mathbf{M}_2 = \mathbf{0}$ from (10.28). The second case: the correlation constraint is strictly loose, i.e, $\mathbf{Q}_1 + \mathbf{Q}_2 \prec \mathbf{S}_2$. In this case, we can define a matrix $\Delta = \mathbf{S}_2 - \mathbf{Q}_1^* - \mathbf{Q}_2^* \succ \mathbf{0}$, and therefore

Δ is a full-rank matrix. Thus, $\mathbf{M}_4 = \mathbf{0}$ and from (10.32), we have $\mathbf{M}_2 = -\mathbf{M}_3$. The matrices $\mathbf{M}_2$, $\mathbf{M}_3$ are both positive semi-definite matrices. Therefore, we must have $\mathbf{M}_2 = \mathbf{M}_3 = \mathbf{0}$. Finally, the third case: the correlation constraint is partially loose, that is, we have $\Delta = \mathbf{S}_2 - \mathbf{Q}_1 - \mathbf{Q}_2 \succeq \mathbf{0}$, hence Δ is not a full-rank matrix. We define $\mathbf{\Sigma} = \mathbf{S}_2 - \mathbf{S}_1 \succ \mathbf{0}$, i.e., $\mathbf{S}_1 = \mathbf{S}_2 - \mathbf{\Sigma}$. In this case, we sum the KKT conditions (10.29) and (10.30) to obtain the following implications:

$$(\mathbf{Q}_1^* + \mathbf{Q}_2^*)(\mathbf{M}_3 - \mathbf{M}_4) - \mathbf{S}_1\mathbf{M}_3 + \mathbf{S}_2\mathbf{M}_4 = \mathbf{0} \tag{10.45}$$

$$(\mathbf{Q}_1^* + \mathbf{Q}_2^*)(\mathbf{M}_3 - \mathbf{M}_4) - \mathbf{S}_2\mathbf{M}_3 + \mathbf{\Sigma}\mathbf{M}_3 + \mathbf{S}_2\mathbf{M}_4 = \mathbf{0} \tag{10.46}$$

$$(\mathbf{S}_2 - \mathbf{Q}_1^* - \mathbf{Q}_2^*)(\mathbf{M}_4 - \mathbf{M}_3) = -\mathbf{\Sigma}\mathbf{M}_3 \tag{10.47}$$

$$(\mathbf{S}_2 - \mathbf{Q}_1^* - \mathbf{Q}_2^*)\mathbf{M}_2 = -\mathbf{\Sigma}\mathbf{M}_3 \tag{10.48}$$

$$(\mathbf{S}_2 - \mathbf{Q}_1^*)\mathbf{M}_2 = -\mathbf{\Sigma}\mathbf{M}_3 \tag{10.49}$$

where (10.48) follows from (10.32), and (10.49) follows from (10.28). Since $(\mathbf{S}_2 - \mathbf{Q}_1^*)\mathbf{M}_2 \succeq \mathbf{0}$ and $\mathbf{\Sigma}\mathbf{M}_3 \succeq \mathbf{0}$, or at least $(\mathbf{S}_2 - \mathbf{Q}_1^*)\mathbf{M}_2$ and $\mathbf{\Sigma}\mathbf{M}_3$ have the same number of non-negative eigenvalues of $\mathbf{M}_2$ and $\mathbf{M}_3$, respectively [133], the only way to satisfy (10.49) is to have all the eigenvalues of both matrices equal zero, i.e., $(\mathbf{S}_2 - \mathbf{Q}_1^*)\mathbf{M}_2 = -\mathbf{\Sigma}\mathbf{M}_3 = \mathbf{0}$. Hence, we conclude that for all three cases we have $(\mathbf{S}_2 - \mathbf{Q}_1^*)\mathbf{M}_2 = \mathbf{0}$ and this completes the proof of Lemma 10.3. ■

Hence, using Lemma 10.3, we write:

$$(\tilde{\mathbf{N}} + \mathbf{S}_2)(\mathbf{Q}_1^* + \tilde{\mathbf{N}})^{-1}$$

$$= (\mathbf{S}_2 - \mathbf{Q}_1^*)(\mathbf{Q}_1^* + \tilde{\mathbf{N}})^{-1} + \mathbf{I} \tag{10.50}$$

$$= (\mathbf{S}_2 - \mathbf{Q}_1^*)[(\mathbf{Q}_1^* + \mathbf{N}_2)^{-1} + \mathbf{M}_2] + \mathbf{I} \tag{10.51}$$

$$= (\mathbf{S}_2 - \mathbf{Q}_1^*)(\mathbf{Q}_1^* + \mathbf{N}_2)^{-1} + \mathbf{I} \tag{10.52}$$

$$= [(\mathbf{N}_2 + \mathbf{S}_2) - (\mathbf{Q}_1^* + \mathbf{N}_2)](\mathbf{Q}_1^* + \mathbf{N}_2)^{-1} + \mathbf{I} \tag{10.53}$$

$$= (\mathbf{N}_2 + \mathbf{S}_2)(\mathbf{Q}_1^* + \mathbf{N}_2)^{-1} \tag{10.54}$$

where (10.51) follows from the definition of the enhanced channel (10.33), and (10.52) follows from Lemma 10.3. Hence, we have:

$$\frac{|\mathbf{S}_2 + \tilde{\mathbf{N}}|}{|\mathbf{S}_2 + \mathbf{N}_2|} = \frac{|\mathbf{Q}_1^* + \tilde{\mathbf{N}}|}{|\mathbf{Q}_1^* + \mathbf{N}_2|} \tag{10.55}$$

We upper bound the secrecy capacity of the MIMO wiretap channel with a receiver-side power constraint by the secrecy capacity of the enhanced channel. Since $\mathbf{S}_2 \in \mathcal{S}_{PE}$, $\mathbf{S}_2$ satisfies the receiver power constraint for the enhanced channel. Hence, the receiver constraint is valid with the upper bounding argument. The secrecy capacity of the enhanced channel $\tilde{C}_s$ is given by

$$\tilde{C}_s = \frac{1}{2}\log\frac{|\mathbf{S}_2 + \tilde{\mathbf{N}}|}{|\tilde{\mathbf{N}}|} - \frac{1}{2}\log\frac{|\mathbf{S}_2 + \mathbf{N}_2|}{|\mathbf{N}_2|} \tag{10.56}$$

$$= \frac{1}{2}\log\frac{|\mathbf{S}_2 + \tilde{\mathbf{N}}|}{|\mathbf{S}_2 + \mathbf{N}_2|} \cdot \frac{|\mathbf{N}_2|}{|\tilde{\mathbf{N}}|} \tag{10.57}$$

$$= \frac{1}{2}\log\frac{|\mathbf{Q}_1^* + \tilde{\mathbf{N}}|}{|\mathbf{Q}_1^* + \mathbf{N}_2|} \cdot \frac{|\mathbf{N}_2|}{|\tilde{\mathbf{N}}|} \tag{10.58}$$

$$= \frac{1}{2}\log\frac{|\mathbf{Q}_1^* + \tilde{\mathbf{N}}|}{|\tilde{\mathbf{N}}|} - \frac{1}{2}\log\frac{|\mathbf{Q}_1^* + \mathbf{N}_2|}{|\mathbf{N}_2|} \tag{10.59}$$

$$= \frac{1}{2}\log\frac{|\mathbf{Q}_1^* + \mathbf{N}_1|}{|\mathbf{N}_1|} - \frac{1}{2}\log\frac{|\mathbf{Q}_1^* + \mathbf{N}_2|}{|\mathbf{N}_2|} \tag{10.60}$$

$$= C(\mathbf{S}_1, \mathbf{S}_2, \mathbf{H}, \mathbf{G}) \tag{10.61}$$

where (10.58) follows from (10.55), and (10.60) follows from (10.44), completing the converse proof for the case of Gaussian signalling with a fixed mean.

10.4.3 Converse Proof for Gaussian Coding with Gaussian Artificial Noise

In this section, we follow a similar channel enhancement technique as in Section 10.4.2. The optimization problem corresponding to the Gaussian coding scheme with artificial noise is:

$$\begin{aligned} \max_{\mathbf{Q}_1, \mathbf{Q}_2 \succeq \mathbf{0}} \quad & \frac{1}{2}\log\frac{|\mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{N}_1|}{|\mathbf{Q}_2 + \mathbf{N}_1|} - \frac{1}{2}\log\frac{|\mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{N}_2|}{|\mathbf{Q}_2 + \mathbf{N}_2|} \\ \text{s.t.} \quad & \mathbf{Q}_1 + \mathbf{Q}_2 \succeq \mathbf{S}_1, \quad \mathbf{Q}_1 + \mathbf{Q}_2 \preceq \mathbf{S}_2 \end{aligned} \tag{10.62}$$

The Lagrangian for this optimization problem is given by:

$$\begin{aligned} \mathcal{L} = & \log\frac{|\mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{N}_2|}{|\mathbf{Q}_2 + \mathbf{N}_2|} - \log\frac{|\mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{N}_1|}{|\mathbf{Q}_2 + \mathbf{N}_1|} - \mathrm{tr}(\mathbf{Q}_1\mathbf{M}_1) - \mathrm{tr}(\mathbf{Q}_2\mathbf{M}_2) \\ & - \mathrm{tr}((\mathbf{Q}_1 + \mathbf{Q}_2 - \mathbf{S}_1)\mathbf{M}_3) + \mathrm{tr}((\mathbf{Q}_1 + \mathbf{Q}_2 - \mathbf{S}_2)\mathbf{M}_4) \end{aligned} \tag{10.63}$$

The complementary slackness conditions (10.28)-(10.30) are still the same due to the same set of constraints for both problems (10.62) and (10.26). The KKT optimality

condition for $\mathbf{Q_1^*}$ and $\mathbf{Q_2^*}$ are:

$$(\mathbf{Q}_1^* + \mathbf{Q}_2^* + \mathbf{N}_2)^{-1} - (\mathbf{Q}_1^* + \mathbf{Q}_2^* + \mathbf{N}_1)^{-1} - \mathbf{M}_1 - \mathbf{M}_3 + \mathbf{M}_4 = \mathbf{0} \tag{10.64}$$

$$(\mathbf{Q}_1^* + \mathbf{Q}_2^* + \mathbf{N}_2)^{-1} - (\mathbf{Q}_2^* + \mathbf{N}_2)^{-1} - (\mathbf{Q}_1^* + \mathbf{Q}_2^* + \mathbf{N}_1)^{-1}$$
$$+(\mathbf{Q}_2^* + \mathbf{N}_1)^{-1} - \mathbf{M}_2 - \mathbf{M}_3 + \mathbf{M}_4 = \mathbf{0} \tag{10.65}$$

Using (10.64), we can write (10.65) as:

$$\mathbf{M}_1 - (\mathbf{Q}_2^* + \mathbf{N}_2)^{-1} + (\mathbf{Q}_2^* + \mathbf{N}_1)^{-1} - \mathbf{M}_2 = \mathbf{0} \tag{10.66}$$

In this case, we again construct an enhanced channel with similar steps as in Section 10.4.2. The enhanced channel is constructed as:

$$(\mathbf{Q}_2^* + \mathbf{N}_1)^{-1} + \mathbf{M}_1 = (\mathbf{Q}_2^* + \mathbf{N}_2)^{-1} + \mathbf{M}_2 = (\mathbf{Q}_2^* + \tilde{\mathbf{N}})^{-1} \tag{10.67}$$

which is the same as in the previous section. Therefore, it follows that $\tilde{\mathbf{N}} \succeq \mathbf{0}$, $\tilde{\mathbf{N}} \preceq \mathbf{N}_1$, $\tilde{\mathbf{N}} \preceq \mathbf{N}_2$. Similarly, we can prove that the rate is preserved for the eavesdropper (as in the set of equations (10.39)-(10.44) with $\mathbf{Q}_2^*$ instead of $\mathbf{Q}_1^*$), i.e.,

$$\frac{|\tilde{\mathbf{N}} + \mathbf{Q}_2^*|}{|\tilde{\mathbf{N}}|} = \frac{|\mathbf{N}_2 + \mathbf{Q}_2^*|}{|\mathbf{N}_2|} \tag{10.68}$$

To prove the rate preservation for the legitimate receiver, we will need the following lemma.

Lemma 10.4 *To achieve a positive secrecy rate using Gaussian coding with artificial noise,* $\mathbf{S}_2$ *must be fully used, i.e.,* $\mathbf{S}_2 = \mathbf{Q}_1^* + \mathbf{Q}_2^*$*, and the optimal covariance matrix used for the artificial noise component,* $\mathbf{Q}_2^*$*, satisfies* $(\mathbf{S}_2 - \mathbf{Q}_2^*)\mathbf{M}_1 = \mathbf{0}$.

Proof: We start by proving the first part of the lemma by contradiction. Assume that a positive secrecy rate can be achieved using artificial noise, and $\mathbf{S}_2$ is partially used. Then, we have two cases. The first case: $\Delta = \mathbf{S}_2 - \mathbf{Q}_1^* - \mathbf{Q}_2^* \succ \mathbf{0}$. Hence, Δ is a full-rank matrix, then $\mathbf{M}_4 = \mathbf{0}$. From (10.64), we can write $(\mathbf{Q}_1^* + \mathbf{Q}_2^* + \mathbf{N}_1)^{-1} + \mathbf{M}_1 + \mathbf{M}_3 = (\mathbf{Q}_1^* + \mathbf{Q}_2^* + \mathbf{N}_2)^{-1}$ and hence, $(\mathbf{Q}_1^* + \mathbf{Q}_2^* + \mathbf{N}_1)^{-1} \preceq (\mathbf{Q}_1^* + \mathbf{Q}_2^* + \mathbf{N}_2)^{-1}$, which results in $\mathbf{N}_2 \preceq \mathbf{N}_1$. This means that the legitimate channel is degraded with respect to the eavesdropper channel, and hence, no positive secrecy rate can be achieved. This contradicts our assumption. The second case: Δ is not full-rank. Due to the similarity of the complementary slackness conditions for the artificial noise and the Gaussian coding with fixed mean settings, we have also (10.47), and from (10.64), we have

$$\mathbf{M}_4 - \mathbf{M}_3 = (\mathbf{Q}_1^* + \mathbf{Q}_2^* + \mathbf{N}_1)^{-1} - (\mathbf{Q}_1^* + \mathbf{Q}_2^* + \mathbf{N}_2)^{-1} + \mathbf{M}_1 \tag{10.69}$$

substituting this in (10.47), we have the following implications:

$$\Delta(\mathbf{Q}_1^* + \mathbf{Q}_2^* + \mathbf{N}_1)^{-1} - \Delta(\mathbf{Q}_1^* + \mathbf{Q}_2^* + \mathbf{N}_2)^{-1} + \Delta\mathbf{M}_1 = -\mathbf{\Sigma}\mathbf{M}_3 \tag{10.70}$$

$$\Delta[(\mathbf{Q}_1^* + \mathbf{Q}_2^* + \mathbf{N}_2)^{-1} - \Delta(\mathbf{Q}_1^* + \mathbf{Q}_2^* + \mathbf{N}_1)^{-1} = \Delta\mathbf{M}_1 + \mathbf{\Sigma}\mathbf{M}_3 \tag{10.71}$$

Then, $[(\mathbf{Q}_1^*+\mathbf{Q}_2^*+\mathbf{N}_2)^{-1}-(\mathbf{Q}_1^*+\mathbf{Q}_2^*+\mathbf{N}_1)]^{-1} \succeq \mathbf{0}$ to have (10.71) hold true [134], and then we have $\mathbf{N}_2 \preceq \mathbf{N}_1$ as in the previous case, which also contradicts the assumption of having a positive secrecy rate. Hence, $\mathbf{Q}_1^* + \mathbf{Q}_2^* = \mathbf{S}_2$. For the second part of the lemma, we now have $\mathbf{S}_2 - \mathbf{Q}_2^* = \mathbf{Q}_1^*$, and from the complementary slackness condition $\mathbf{Q}_1^*\mathbf{M}_1 = \mathbf{0}$. Then, we conclude that $(\mathbf{S}_2 - \mathbf{Q}_2^*)\mathbf{M}_1 = \mathbf{0}$, completing the proof of Lemma 10.4. ■

Using Lemma 10.4, we can prove rate preservation for the legitimate receiver as follows:

$$(\tilde{\mathbf{N}} + \mathbf{S}_2)(\mathbf{Q}_2^* + \tilde{\mathbf{N}})^{-1} = (\mathbf{S}_2 - \mathbf{Q}_2^*)(\mathbf{Q}_2^* + \tilde{\mathbf{N}})^{-1} + \mathbf{I} \tag{10.72}$$

$$= (\mathbf{S}_2 - \mathbf{Q}_2^*)[(\mathbf{Q}_2^* + \mathbf{N}_1)^{-1} + \mathbf{M}_1] + \mathbf{I} \tag{10.73}$$

$$= (\mathbf{S}_2 - \mathbf{Q}_2^*)(\mathbf{Q}_2^* + \mathbf{N}_1)^{-1} + \mathbf{I} \tag{10.74}$$

$$= [(\mathbf{N}_1 + \mathbf{S}_2) - (\mathbf{Q}_2^* + \mathbf{N}_1)](\mathbf{Q}_2^* + \mathbf{N}_1)^{-1} + \mathbf{I} \tag{10.75}$$

$$= (\mathbf{N}_1 + \mathbf{S}_2)(\mathbf{Q}_2^* + \mathbf{N}_1)^{-1} \tag{10.76}$$

where (10.73) follows from the definition of the enhanced channel (10.67), and (10.74) follows from Lemma 10.4. Therefore, we have:

$$\frac{|\mathbf{S}_2 + \tilde{\mathbf{N}}|}{|\mathbf{Q}_2^* + \tilde{\mathbf{N}}|} = \frac{|\mathbf{S}_2 + \mathbf{N}_1|}{|\mathbf{Q}_2^* + \mathbf{N}_1|} \tag{10.77}$$

Hence, the secrecy capacity of the enhanced channel is given by:

$$\tilde{C}_s = \frac{1}{2}\log\frac{|\mathbf{S}_2 + \tilde{\mathbf{N}}|}{|\tilde{\mathbf{N}}|} - \frac{1}{2}\log\frac{|\mathbf{S}_2 + \mathbf{N}_2|}{|\mathbf{N}_2|} \tag{10.78}$$

$$= \frac{1}{2}\log\frac{|\mathbf{S}_2+\tilde{\mathbf{N}}|}{|\mathbf{S}_2+\mathbf{N}_2|}\cdot\frac{|\mathbf{N}_2|}{|\tilde{\mathbf{N}}|} \tag{10.79}$$

$$= \frac{1}{2}\log\frac{|\mathbf{S}_2+\tilde{\mathbf{N}}|}{|\mathbf{S}_2+\mathbf{N}_2|}\cdot\frac{|\mathbf{Q}_2^*+\mathbf{N}_2|}{|\mathbf{Q}_2^*+\tilde{\mathbf{N}}|} \tag{10.80}$$

$$= \frac{1}{2}\log\frac{|\mathbf{S}_2+\tilde{\mathbf{N}}|}{|\mathbf{Q}_2^*+\tilde{\mathbf{N}}|}\cdot\frac{|\mathbf{Q}_2^*+\mathbf{N}_2|}{|\mathbf{S}_2+\mathbf{N}_2|} \tag{10.81}$$

$$= \frac{1}{2}\log\frac{|\mathbf{S}_2+\mathbf{N}_1|}{|\mathbf{Q}_2^*+\mathbf{N}_1|}\cdot\frac{|\mathbf{Q}_2^*+\mathbf{N}_2|}{|\mathbf{S}_2+\mathbf{N}_2|} \tag{10.82}$$

$$= \frac{1}{2}\log\frac{|\mathbf{S}_2+\mathbf{N}_1|}{|\mathbf{Q}_2^*+\mathbf{N}_1|}-\frac{1}{2}\log\frac{|\mathbf{S}_2+\mathbf{N}_2|}{|\mathbf{Q}_2^*+\mathbf{N}_2|} \tag{10.83}$$

$$= \frac{1}{2}\log\frac{|\mathbf{Q}_1^*+\mathbf{Q}_2^*+\mathbf{N}_1|}{|\mathbf{Q}_2^*+\mathbf{N}_1|}-\frac{1}{2}\log\frac{|\mathbf{Q}_1^*+\mathbf{Q}_2^*+\mathbf{N}_2|}{|\mathbf{Q}_2^*+\mathbf{N}_2|} \tag{10.84}$$

$$= C(\mathbf{S}_1,\mathbf{S}_2,\mathbf{H},\mathbf{G}) \tag{10.85}$$

where (10.80) follows from (10.68), (10.82) follows from (10.77), and (10.84) follows from $\mathbf{Q}_1^*+\mathbf{Q}_2^*=\mathbf{S}_2$, completing the converse proof for the case of Gaussian signalling with Gaussian artificial noise.

10.5 Extensions to Related Channel Models

10.5.1 Gaussian MIMO Wiretap Channel Under Dual Minimum Receiver-Side Power Constraints

In this section, we consider the case where we impose dual receiver-side *minimum* power constraints, i.e., receiver-side power constraints both on the legitimate receiver and the eavesdropper. Then, we have the following constraint in addition to the

constraints in (10.3) and (10.4):

$$\mathrm{tr}(\mathbb{E}[\mathbf{Y}\mathbf{Y}^T]) \geq E_2 \tag{10.86}$$

where E_2 is the minimum power level that should be delivered to the legitimate receiver. The following theorem characterizes the secrecy capacity of this model.

Theorem 10.2 *The secrecy capacity of a MIMO wiretap channel with a transmitter-side power constraint P and dual receiver-side power constraints E_1, E_2, $C(E_1, E_2, P, \mathbf{H}, \mathbf{G})$, is given as*

$$\begin{aligned} C(E_1, E_2, P, \mathbf{H}, \mathbf{G}) = \max_{\mathbf{Q}\succeq\mathbf{0}, \boldsymbol{\mu}} \quad & \frac{1}{2}\log|\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^T| - \frac{1}{2}\log|\mathbf{I} + \mathbf{G}\mathbf{Q}\mathbf{G}^T| \\ s.t. \quad & \mathrm{tr}(\mathbf{Q} + \boldsymbol{\mu}\boldsymbol{\mu}^T) \leq P \\ & \mathrm{tr}(\mathbf{G}(\mathbf{Q} + \boldsymbol{\mu}\boldsymbol{\mu}^T)\mathbf{G}^T) \geq \tilde{E}_1 \\ & \mathrm{tr}(\mathbf{H}(\mathbf{Q} + \boldsymbol{\mu}\boldsymbol{\mu}^T)\mathbf{H}^T) \geq \tilde{E}_2 \end{aligned} \tag{10.87}$$

where $\tilde{E}_1 = E_1 - N_e$, and $\tilde{E}_2 = E - N_r$. This secrecy capacity is achieved by $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{Q})$, i.e., with a mean but no channel prefixing. Alternatively, $C(E_1, E_2, P, \mathbf{H}, \mathbf{G})$ is also given as

$$\begin{aligned} C(E_1, E_2, P, \mathbf{H}, \mathbf{G}) \max_{\mathbf{Q}_1, \mathbf{Q}_2 \succeq \mathbf{0}} \quad & \frac{1}{2}\log\frac{|\mathbf{I} + \mathbf{H}(\mathbf{Q}_1 + \mathbf{Q}_2)\mathbf{H}^T|}{|\mathbf{I} + \mathbf{H}\mathbf{Q}_2\mathbf{H}^T|} \\ & - \frac{1}{2}\log\frac{|\mathbf{I} + \mathbf{G}(\mathbf{Q}_1 + \mathbf{Q}_2)\mathbf{G}^T|}{|\mathbf{I} + \mathbf{G}\mathbf{Q}_2\mathbf{G}^T|} \\ s.t. \quad & \mathrm{tr}(\mathbf{Q}_1 + \mathbf{Q}_2) \leq P \end{aligned}$$

$$\begin{aligned}\mathrm{tr}(\mathbf{G}(\mathbf{Q}_1+\mathbf{Q}_2)\mathbf{G}^T) &\geq \tilde{E}_1 \\ \mathrm{tr}(\mathbf{H}(\mathbf{Q}_1+\mathbf{Q}_2)\mathbf{H}^T) &\geq \tilde{E}_2 \end{aligned} \tag{10.88}$$

where $\mathbf{X}=\mathbf{V}+\mathbf{U}$*, with jointly Gaussian* $\mathbf{V}\sim\mathcal{N}(\mathbf{0},\mathbf{Q}_1)$ *and* $\mathbf{U}\sim\mathcal{N}(\mathbf{0},\mathbf{Q}_2)$*, where* $\mathbf{U},\mathbf{V}$ *are independent, i.e., Gaussian signalling with Gaussian channel prefixing.*

Proof: The proof relies on verifying that the *double-sided correlation matrix constraint* constructed in Section 10.4.1 is sufficient for this case also. First, we define the set $\mathcal{S}_{PE_1E_2}$ as:

$$\mathcal{S}_{PE_1E_2}=\{\mathbf{S}\succeq\mathbf{0}:\mathrm{tr}(\mathbf{S})\leq P,\quad \mathrm{tr}(\mathbf{GSG}^T)\geq\tilde{E}_1,\quad \mathrm{tr}(\mathbf{HSH}^T)\geq\tilde{E}_2\} \tag{10.89}$$

To show the direct implication

$$C(E_1,E_2,P,\mathbf{H},\mathbf{G})\geq \max_{\mathbf{S}_1,\mathbf{S}_2\in\mathcal{S}_{PE_1E_2},\mathbf{S}_1\preceq\mathbf{S}_2} C(\mathbf{S}_1,\mathbf{S}_2,\mathbf{H},\mathbf{G}) \tag{10.90}$$

we note that for any $\mathbf{Q}$ such that $\mathbf{S}_1\preceq\mathbf{Q}\preceq\mathbf{S}_1$ where $\mathbf{S}_1,\mathbf{S}_2\in\mathcal{S}_{PE_1E_2}$, we have $\mathrm{tr}(\mathbf{Q})\leq\mathrm{tr}(\mathbf{S}_2)\leq P$, $\mathrm{tr}(\mathbf{GQG}^T)\geq\mathrm{tr}(\mathbf{GS}_1\mathbf{G}^T)\geq E_1$ and $\mathrm{tr}(\mathbf{HQH}^T)\geq\mathrm{tr}(\mathbf{HS}_1\mathbf{H}^T)\geq E_2$. Consequently, $\mathbf{Q}\in\mathcal{S}_{PE_1E_2}$, i.e., the feasible set under $\mathbf{S}_1,\mathbf{S}_2\in\mathcal{S}_{PE_1E_2}$ is a subset of the feasible set under P,E_1,E_2 constraints. Moreover, $\mathcal{S}_{PE_1E_2}\subseteq\mathcal{S}_{PE}$ defined in Section 10.2, and hence $\mathcal{S}_{PE_1E_2}$ is also a compact set. Hence the implication

$$C(E_1,E_2,P,\mathbf{H},\mathbf{G})\leq \max_{\mathbf{S}_1,\mathbf{S}_2\in\mathcal{S}_{PE_1E_2},\mathbf{S}_1\preceq\mathbf{S}_2} C(\mathbf{S}_1,\mathbf{S}_2,\mathbf{H},\mathbf{G}) \tag{10.91}$$

can be proved by following the reverse implication (10.24) of the proof of Lemma 10.2 for the compact set $\mathcal{S}_{PE_1E_2}$, we can show that:

$$C(E_1, E_2, P, \mathbf{H}, \mathbf{G}) = \max_{\mathbf{S}_1, \mathbf{S}_2 \in \mathcal{S}_{PE_1E_2}, \mathbf{S}_1 \preceq \mathbf{S}_2} C(\mathbf{S}_1, \mathbf{S}_2, \mathbf{H}, \mathbf{G}) \tag{10.92}$$

Then, the inner problem under the dual receiver-side power constraints is identical to its counterpart under a single receiver-side power constraint on the eavesdropper side only. Consequently, achievability schemes of mean based and artificial noise based signalling are optimal for the dual receiver-side minimum power constraints.

It only remains to show that the achievable rates with Gaussian signalling with fixed mean match the converse, i.e., that when the covariance matrix representing the mean is left unrestricted for converse purposes, at the optimal, it takes a unit-rank so that it can be implemented with a mean vector in the achievability. That is, we need to show that Lemma 10.1 extends to the current setting under P, E_1, E_2 constraints. To show this, as a generalization of (10.13), we need to solve:

$$\max_{\mathbf{Q}_2 \succeq \mathbf{0}} \quad \alpha_1 \mathrm{tr}(\mathbf{G}\mathbf{Q}_2\mathbf{G}^T) + \alpha_2 \mathrm{tr}(\mathbf{H}\mathbf{Q}_2\mathbf{H}^T) \quad \text{s.t.} \quad \mathrm{tr}(\mathbf{Q}_2) = \tilde{P} \tag{10.93}$$

This optimization problem is equivalent to:

$$\max_{\lambda_i, \mathbf{q}_i} \quad \sum_{i=1}^{r} \lambda_i \left(\alpha_1 \|\mathbf{G}\mathbf{q}_i\|^2 + \alpha_2 \|\mathbf{H}\mathbf{q}_i\|^2 \right) \quad \text{s.t.} \quad \sum_{i=1}^{r} \lambda_i = \tilde{P} \tag{10.94}$$

which has a beam-forming optimal solution of assigning all $\tilde{P}$ to $\mathbf{q}_m$ such that

$$m = \arg\max_i \alpha_1 \|\mathbf{G}\mathbf{q}_i\|^2 + \alpha_2 \|\mathbf{H}\mathbf{q}_i\|^2 \tag{10.95}$$

and hence the optimal $\mathbf{Q}_2$ is unit-rank and the mean-based signalling is feasible. ■

10.5.2 Gaussian MIMO Wiretap Channel Under Maximum Receiver-Side Power Constraints

In this section, we consider the MIMO wiretap channel under *maximum* receiver-side power constraints. This generalizes Gastpar's problem [107] to include a secrecy requirement. In this case, we limit the interference at both receivers instead of maintaining the received power levels at both receivers as in Section 10.2. Then, we impose the following constraints together with (10.3):

$$\mathrm{tr}(\mathbb{E}[\mathbf{Z}\mathbf{Z}^T]) \leq E_1, \quad \mathrm{tr}(\mathbb{E}[\mathbf{Y}\mathbf{Y}^T]) \leq E_2 \tag{10.96}$$

Theorem 10.3 *The secrecy capacity of the MIMO wiretap channel with a transmitter-side power constraint P and maximum receiver-side power constraints E_1, E_2, $C(E_1, E_2, P, \mathbf{H}, \mathbf{G})$, is*

$$\begin{aligned} \max_{\mathbf{Q}\succeq\mathbf{0}} \quad & \frac{1}{2}\log|\mathbf{I}+\mathbf{H}\mathbf{Q}\mathbf{H}^T| - \frac{1}{2}\log|\mathbf{I}+\mathbf{G}\mathbf{Q}\mathbf{G}^T| \\ s.t. \quad & \mathrm{tr}(\mathbf{Q}) \leq P, \quad \mathrm{tr}(\mathbf{G}\mathbf{Q}\mathbf{G}^T) \leq \tilde{E}_1, \quad \mathrm{tr}(\mathbf{H}\mathbf{Q}\mathbf{H}^T) \leq \tilde{E}_2 \end{aligned} \tag{10.97}$$

This secrecy capacity is achieved by $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$*, i.e., neither mean or channel prefixing is required.*

Proof: Similar to the previous section, we construct a suitable correlation matrix set $\mathcal{S}'_{PE_1E_2}$ as:

$$\mathcal{S}'_{PE_1E_2} = \{\mathbf{S} \succeq \mathbf{0} : \mathrm{tr}(\mathbf{S}) \leq P, \quad \mathrm{tr}(\mathbf{GSG}^T) \leq \tilde{E}_1, \quad \mathrm{tr}(\mathbf{HSH}^T) \leq \tilde{E}_2\} \tag{10.98}$$

Now, we show that, using a *single-sided* correlation matrix constraint $\mathbf{Q} \preceq \mathbf{S}$ is sufficient for *maximum* receiver-side power constraints, unlike the *double-sided* correlation constraint that was necessary for *minimum* receiver-side power constraints so far. Since, for all $\mathbf{Q} \preceq \mathbf{S}$, we have $\mathrm{tr}(\mathbf{Q}) \leq \mathrm{tr}(\mathbf{S}) \leq P$, $\mathrm{tr}(\mathbf{GQG}^T) \leq \mathrm{tr}(\mathbf{GSG}^T) \leq \tilde{E}_1$ and $\mathrm{tr}(\mathbf{HQH}^T) \leq \mathrm{tr}(\mathbf{HSH}^T) \leq \tilde{E}_2$, we thus have $\mathbf{Q} \in \mathcal{S}'_{PE_1E_2}$. Moreover, the set $\mathcal{S}'_{PE_1E_2}$ is closed and bounded and hence compact. Consequently, we can find a sequence of codes $\mathcal{C}(n_k, \mathbf{S}_0 + \delta\mathbf{I}, R, \epsilon_k)$ with $\mathbf{S}_0 \in \mathcal{S}'_{PE_1E_2}$, achieving small probability of error, that has a limit point of $C(\mathbf{S}_0, \mathbf{H}, \mathbf{G})$ and hence

$$C(E_1, E_2, P, \mathbf{H}, \mathbf{G}) = \max_{\mathbf{S} \in \mathcal{S}'_{PE_1E_2}} \quad C(\mathbf{S}, \mathbf{H}, \mathbf{G}) \tag{10.99}$$

Consequently, the inner problem under a correlation matrix constraint for the wiretap channel with maximum receiver-side power limitations is identical to the inner problem for the classical wiretap channel without the extra maximum receiver-side power constraints. Hence, the classical Gaussian coding with zero-mean and no channel-prefixing is optimal. ■

10.5.3 Gaussian MIMO Broadcast Channel Under Minimum Receiver-Side Power Constraints

In this section, we consider the MIMO BC with no secrecy constraints under *minimum* receiver-side power constraints. In this setting, the transmitter is required to communicate messages simultaneously and reliably with the largest possible rate, and at the same time, deliver the minimum required powers to the receivers: $\mathrm{tr}(\mathbb{E}[\mathbf{Z}\mathbf{Z}^T]) \geq E_1$, $\mathrm{tr}(\mathbb{E}[\mathbf{Y}\mathbf{Y}^T]) \geq E_2$. The problem without the receiver-side constraints is solved by Weingarten et. al. [102]. The rate region is achieved using DPC along with time sharing. We show in the following theorem that the DPC is optimal even after imposing the receiver-side power constraints.

Theorem 10.4 *The capacity region of a MIMO broadcast channel with a transmitter-side power constraint P and minimum receiver-side power constraints E_1, E_2, $\mathcal{C}(E_1, E_2, P, \mathbf{H}, \mathbf{G})$, is given by the DPC region, which is the convex hull of the union of two regions $\mathcal{R}_1^{DPC}$ and $\mathcal{R}_2^{DPC}$, corresponding to the two orders of encoding, given as:*

$$\mathcal{R}_1^{DPC} = \left\{(R_1, R_2) : R_1 \leq \frac{1}{2}\log|\mathbf{I} + \mathbf{H}\mathbf{Q}_1\mathbf{H}^T|, R_2 \leq \frac{1}{2}\log\frac{|\mathbf{I} + \mathbf{G}(\mathbf{Q}_1 + \mathbf{Q}_2)\mathbf{G}^T|}{|\mathbf{I} + \mathbf{G}\mathbf{Q}_1\mathbf{G}^T|}\right\}$$
$$\mathcal{R}_2^{DPC} = \left\{(R_1, R_2) : R_1 \leq \frac{1}{2}\log\frac{|\mathbf{I} + \mathbf{H}(\mathbf{Q}_1 + \mathbf{Q}_2)\mathbf{H}^T|}{|\mathbf{I} + \mathbf{H}\mathbf{Q}_2\mathbf{H}^T|}, R_2 \leq \frac{1}{2}\log|\mathbf{I} + \mathbf{G}\mathbf{Q}_2\mathbf{G}^T|\right\} \tag{10.100}$$

both of which subject to

$$\begin{aligned} \mathrm{tr}(\mathbf{Q}_1+\mathbf{Q}_2) &\leq P \\ \mathrm{tr}(\mathbf{G}(\mathbf{Q}_1+\mathbf{Q}_2)\mathbf{G}^T) &\geq \tilde{E}_1 \\ \mathrm{tr}(\mathbf{H}(\mathbf{Q}_1+\mathbf{Q}_2)\mathbf{H}^T) &\geq \tilde{E}_2 \end{aligned} \tag{10.101}$$

Proof: We consider, without loss of generality, the region of rates achieved by $\mathcal{R}_1^{DPC}$. We first note that, due to the presence of the minimum receiver-side power constraints, we need to consider a double-sided correlation matrix constraint $\mathbf{S}_1 \preceq \mathbf{Q}_1+\mathbf{Q}_2 \preceq \mathbf{S}_2$, for any fixed $\mathbf{S}_1, \mathbf{S}_2$ in $\mathcal{S}_{PE_1E_2}$ in (10.89). Following the original channel enhancement proof of the aligned MIMO (not necessarily degraded) BC (AMBC) in [102], it suffices to prove that under a double-sided correlation matrix constraint $\mathbf{S}_1 \preceq \mathbf{Q}_1+\mathbf{Q}_2 \preceq \mathbf{S}_2$, there exists an enhanced aligned degraded BC (ADBC) such that for $\alpha_1 \leq \alpha_2$, noise covariances of the enhanced channel satisfy the covariance increment $\tilde{\mathbf{N}}_1 \preceq \tilde{\mathbf{N}}_2$ and supporting hyperplane preservation.

First, the achievable DPC rates in the aligned case with the encoding order in $\mathcal{R}_1^{DPC}$ are

$$\begin{aligned} \max_{\mathbf{Q}_1,\mathbf{Q}_2\succeq\mathbf{0}} \quad & \alpha_1 \cdot \frac{1}{2}\log\frac{|\mathbf{Q}_1+\mathbf{N}_1|}{|\mathbf{N}_1|} + \alpha_2 \cdot \frac{1}{2}\log\frac{|\mathbf{Q}_1+\mathbf{Q}_2+\mathbf{N}_2|}{|\mathbf{Q}_1+\mathbf{N}_2|} \\ \text{s.t.} \quad & \mathbf{Q}_1+\mathbf{Q}_2 \succeq \mathbf{S}_1, \quad \mathbf{Q}_1+\mathbf{Q}_2 \preceq \mathbf{S}_2 \end{aligned} \tag{10.102}$$

The Lagrangian for this problem is:

$$\mathcal{L} = \alpha_1 \cdot \frac{1}{2}\log\frac{|\mathbf{Q}_1+\mathbf{N}_1|}{|\mathbf{N}_1|} + \alpha_2 \cdot \frac{1}{2}\log\frac{|\mathbf{Q}_1+\mathbf{Q}_2+\mathbf{N}_2|}{|\mathbf{Q}_1+\mathbf{N}_2|} + \mathrm{tr}(\mathbf{Q}_1\mathbf{M}_1) + \mathrm{tr}(\mathbf{Q}_2\mathbf{M}_2)$$
$$+ \mathrm{tr}((\mathbf{Q}_1+\mathbf{Q}_2-\mathbf{S}_1)\mathbf{M}_3) - \mathrm{tr}((\mathbf{Q}_1+\mathbf{Q}_2-\mathbf{S}_2)\mathbf{M}_4) \tag{10.103}$$

The KKT optimality conditions for $\mathbf{Q}_1^*, \mathbf{Q}_2^*$ are:

$$\frac{\alpha_1}{2}(\mathbf{Q}_1^*+\mathbf{N}_1)^{-1} + \frac{\alpha_2}{2}(\mathbf{Q}_1^*+\mathbf{Q}_2^*+\mathbf{N}_2)^{-1}$$
$$-\frac{\alpha_2}{2}(\mathbf{Q}_1^*+\mathbf{N}_2)^{-1} + \mathbf{M}_1 + \mathbf{M}_3 - \mathbf{M}_4 = \mathbf{0} \tag{10.104}$$
$$\frac{\alpha_2}{2}(\mathbf{Q}_1^*+\mathbf{Q}_2^*+\mathbf{N}_2)^{-1} + \mathbf{M}_2 + \mathbf{M}_3 - \mathbf{M}_4 = \mathbf{0} \tag{10.105}$$

and the complementary slackness conditions are as in (10.28)-(10.30). From (10.105) and (10.104), we have:

$$\frac{\alpha_1}{2}(\mathbf{Q}_1^*+\mathbf{N}_1)^{-1} + \mathbf{M}_1 = \frac{\alpha_2}{2}(\mathbf{Q}_1^*+\mathbf{N}_2)^{-1} + \mathbf{M}_2 \tag{10.106}$$

Consequently, we construct the enhanced channels as:

$$\frac{\alpha_1}{2}(\mathbf{Q}_1^*+\mathbf{N}_1)^{-1} + \mathbf{M}_1 = \frac{\alpha_1}{2}(\mathbf{Q}_1^*+\tilde{\mathbf{N}}_1)^{-1} \tag{10.107}$$
$$\frac{\alpha_2}{2}(\mathbf{Q}_1^*+\mathbf{N}_2)^{-1} + \mathbf{M}_2 = \frac{\alpha_2}{2}(\mathbf{Q}_1^*+\tilde{\mathbf{N}}_2)^{-1} \tag{10.108}$$

Then, $\tilde{\mathbf{N}}_1 \preceq \mathbf{N}_1$ and $\tilde{\mathbf{N}}_2 \preceq \mathbf{N}_2$, and thus, the constructed channels are enhanced. We need show that the enhanced BC is degraded in favor of receiver 1. Since $\alpha_1 \leq \alpha_2$,

from (10.106)-(10.108),

$$(\mathbf{Q}_1^* + \tilde{\mathbf{N}}_1)^{-1} = \frac{\alpha_2}{\alpha_1}(\mathbf{Q}_1^* + \tilde{\mathbf{N}}_2)^{-1} \succeq (\mathbf{Q}_1^* + \tilde{\mathbf{N}}_2)^{-1} \tag{10.109}$$

and hence $\tilde{\mathbf{N}}_1 \preceq \tilde{\mathbf{N}}_2$. Moreover, we have the rate preservation relation of receiver 1,

$$\frac{|\mathbf{Q}_1^* + \tilde{\mathbf{N}}_1|}{|\tilde{\mathbf{N}}_1|} = \frac{|\mathbf{Q}_1^* + \mathbf{N}_1|}{|\mathbf{N}_1|} \tag{10.110}$$

and the rate preservation for user 2 can be shown as:

$$(\mathbf{Q}_1^* + \mathbf{Q}_2^* + \tilde{\mathbf{N}}_2)(\mathbf{Q}_1^* + \tilde{\mathbf{N}}_2)^{-1} = \mathbf{Q}_2^*(\mathbf{Q}_1^* + \tilde{\mathbf{N}}_2)^{-1} + \mathbf{I} \tag{10.111}$$

$$= \mathbf{Q}_2^*[(\mathbf{Q}_1^* + \mathbf{N}_2)^{-1} + \frac{2}{\alpha_2}\mathbf{M}_2] + \mathbf{I} \tag{10.112}$$

$$= \mathbf{Q}_2^*(\mathbf{Q}_1^* + \mathbf{N}_2)^{-1} + \mathbf{I} \tag{10.113}$$

$$= (\mathbf{Q}_1^* + \mathbf{Q}_2^* + \mathbf{N}_2)(\mathbf{Q}_1^* + \mathbf{N}_2)^{-1} \tag{10.114}$$

leading to:

$$\frac{|\mathbf{Q}_1^* + \mathbf{Q}_2^* + \tilde{\mathbf{N}}_2|}{|\mathbf{Q}_1^* + \tilde{\mathbf{N}}_2|} = \frac{|\mathbf{Q}_1^* + \mathbf{Q}_2^* + \mathbf{N}_2|}{|\mathbf{Q}_1^* + \mathbf{N}_2|} \tag{10.115}$$

Hence, we have an enhanced ADBC whose rate region is achieved by a Gaussian codebook and use full $\mathbf{S}_2$ [102]. Additionally, from (10.110) and (10.115), we conclude that the rate region of the original AMBC coincides with the optimal Gaussian rate region $\mathcal{R}^G(\mathbf{S}_2, \tilde{\mathbf{N}}_1, \tilde{\mathbf{N}}_2)$ of the enhanced ADBC. To complete the proof, we need to show that the supporting hyperplane $\{(R_1, R_2) : \alpha_1 R_1 + \alpha_2 R_2 = b\}$ is

also a supporting hyperplane for the Gaussian rate region of the enhanced ADBC $\mathcal{R}^G(\mathbf{S}_2, \tilde{\mathbf{N}}_1, \tilde{\mathbf{N}}_2)$, i.e., that $\sum_{i=1}^{2} \alpha_i R_i^G(\mathbf{Q}_1, \mathbf{Q}_2, \tilde{\mathbf{N}}_1, \tilde{\mathbf{N}}_2)$ is maximized by the $\mathbf{Q}_i^*$ that solves the AMBC problem. The proof of this follows from [102]. ■

We note that the related work [135] considers a MISO BC with multiple receivers, where each receiver requires either data or energy, but not both. The energy-requiring users are satisfied by the transmission of pseudo-random signals, that are known to all receivers, which can be subtracted out for communication purposes with the information-requiring users. The information-requiring users are served with a DPC scheme, which is optimal in that case due to [102], as energy transfer does not interact with data transfer. The emphasis in [135] is the optimization of the system for this transmission scheme. In our work, all users require both data and information simultaneously. We prove by developing a suitable channel enhancement method using double-sided correlation matrix constraints that DPC is optimal for this system.

10.5.4 Gaussian MIMO Broadcast Channel with Confidential messages Under Minimum Receiver-Side Power Constraints

In this section, we consider the MIMO BCCM where we transmit a message to each receiver secret from the other. In this setting, the transmitter is required to communicate messages reliably, securely and at the same time deliver minimum amounts of energy E_1 and E_2 to the receivers. The problem without receiver-side power constraints was solved in [72], and it was shown that secure DPC (S-DPC)

attains the secrecy capacity region. We show in the following theorem that S-DPC is optimal in the presence of receiver-side power constraints as well.

Theorem 10.5 *The secrecy capacity region of a MIMO broadcast channel with a transmitter-side power constraint P and minimum receiver-side power constraints E_1, E_2 and with secrecy constraints, $\mathcal{C}(E_1, E_2, P, \mathbf{H}, \mathbf{G})$, is given by the S-DPC region,*

$$\begin{aligned} R_1 &\leq \max_{\mathbf{Q}_1,\mathbf{Q}_2 \succeq \mathbf{0}} \quad \frac{1}{2}\log|\mathbf{I}+\mathbf{H}\mathbf{Q}_1\mathbf{H}^T| - \frac{1}{2}\log|\mathbf{I}+\mathbf{G}\mathbf{Q}_1\mathbf{G}^T| \\ R_2 &\leq \max_{\mathbf{Q}_1,\mathbf{Q}_2 \succeq \mathbf{0}} \quad \frac{1}{2}\log\frac{|\mathbf{I}+\mathbf{G}(\mathbf{Q}_1+\mathbf{Q}_2)\mathbf{G}^T|}{|\mathbf{I}+\mathbf{G}\mathbf{Q}_1\mathbf{G}^T|} \\ &\qquad\qquad - \frac{1}{2}\log\frac{|\mathbf{I}+\mathbf{H}(\mathbf{Q}_1+\mathbf{Q}_2)\mathbf{H}^T|}{|\mathbf{I}+\mathbf{H}\mathbf{Q}_1\mathbf{H}^T|} \\ &\qquad s.t. \quad \mathrm{tr}(\mathbf{Q}_1+\mathbf{Q}_2) \leq P \\ &\qquad\qquad \mathrm{tr}(\mathbf{G}(\mathbf{Q}_1+\mathbf{Q}_2)\mathbf{G}^T) \geq \tilde{E}_1 \\ &\qquad\qquad \mathrm{tr}(\mathbf{H}(\mathbf{Q}_1+\mathbf{Q}_2)\mathbf{H}^T) \geq \tilde{E}_2 \end{aligned} \tag{10.116}$$

This region is achieved by S-DPC (Gaussian double binning) using jointly Gaussian random variables $(\mathbf{V}_1, \mathbf{V}_2) \to \mathbf{X} \to (\mathbf{Y}, \mathbf{Z})$ such that $\mathbf{V}_1 = \mathbf{U}_1 + \mathbf{F}\mathbf{U}_2$, $\mathbf{V}_2 = \mathbf{U}_2$, $\mathbf{X} = \mathbf{U}_1 + \mathbf{U}_2$, where $\mathbf{U}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_1)$, $\mathbf{U}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_2)$ are independent and $\mathbf{F} = \mathbf{Q}_1\mathbf{H}^T(\mathbf{I}+\mathbf{H}\mathbf{Q}_1\mathbf{H}^T)^{-1}\mathbf{H}$.

Proof: In this case also, we have a double-sided correlation matrix constraint $\mathbf{S}_1 \preceq \mathbf{Q}_1 + \mathbf{Q}_2 \preceq \mathbf{S}_2$, where $\mathbf{S}_1, \mathbf{S}_2$ in $\mathcal{S}_{PE_1E_2}$ in (10.89). From Lemma 10.4, we know that, to have a positive secrecy rate at receiver 2, we must use the full correlation matrix

$\mathbf{S}_2$, i.e., $\mathbf{Q}_1+\mathbf{Q}_2=\mathbf{S}_2$. Since the outer optimization problem chooses $\mathbf{S}_2$ from the set $\mathcal{S}_{PE_1E_2}$, and $\mathbf{X}$ has the covariance $\mathbf{Q}=\mathbf{Q}_1+\mathbf{Q}_2$, the receiver-side power constraints are satisfied. The achievability of the corner point follows from [72] by using the double binning scheme presented in [70].

We next need to show that the achievable scheme matches the converse. For receiver 2: From Theorem 10.1, noticing that $\mathbf{G}$ in this case corresponds to the main channel and $\mathbf{H}$ corresponds to the eavesdropper channel, the achievable rate $R_{2,\max}$ in (10.116) is equal to the secrecy capacity $C(\mathbf{S}_1,\mathbf{S}_2,\mathbf{G},\mathbf{H})$ in (10.19) proving the converse. For receiver 1: The achievable rate $R_{1,\max}$ in (10.116) is the same as the secrecy capacity $C(\mathbf{S}_1,\mathbf{S}_2,\mathbf{H},\mathbf{G})$ in (10.11) except for the correlation constraint $\mathbf{S}_1 \preceq \mathbf{Q}_1+\boldsymbol{\mu}\boldsymbol{\mu}^T \preceq \mathbf{S}_2$. Recall that, in Section 10.4.2, we proved the converse for arbitrary $\mathbf{Q}_2$, not necessarily unit-rank. Therefore, using S-DPC encoding scheme induces the required extra covariance component $\mathbf{Q}_2$ that supports the receiver-side constraint. Moreover, we observe that

$$C(\mathbf{S}_1,\mathbf{S}_2,\mathbf{G},\mathbf{H}) = C(\mathbf{S}_1,\mathbf{S}_2,\mathbf{H},\mathbf{G}) + \frac{1}{2}\log\frac{|\mathbf{I}+\mathbf{G}\mathbf{S}_2\mathbf{G}^T|}{|\mathbf{I}+\mathbf{H}\mathbf{S}_2\mathbf{H}^T|} \tag{10.117}$$

This implies that $\mathbf{Q}_1$ maximizes the secrecy capacities of both users simultaneously. Consequently, the two users can receive the confidential messages at their respective maximum secrecy rates as individual wiretap channels, i.e., the secrecy rate region is rectangular under the $\mathbf{S}_1,\mathbf{S}_2$ correlation matrix constraints. Hence, the S-DPC scheme is optimal. ■

10.6 Practical Optimization Approaches

In this section, we provide several optimization approaches to evaluate the capacities under receiver-side power constraints stated in Theorems 10.1-10.5. Without loss of generality, we consider the case of a single minimum receiver-side power constraint in the wiretap channel in Theorem 10.1. This is one of the most challenging optimization problems among the results in Theorems 10.1-10.5, as the optimization problem in this case is not convex.

10.6.1 MISO Problem with Gaussian Mean-Based Coding Scheme

The MISO problem with Gaussian mean-based coding scheme can be exactly cast as a convex optimization problem by considering a linear fractional transformation (Charnes-Cooper transformation) [136] as follows:

$$\begin{aligned} \max_{\mathbf{Q}_1,\mathbf{Q}_2\succeq\mathbf{0}} \quad & \frac{1}{2}\log(1+\mathbf{h}^T\mathbf{Q}_1\mathbf{h}) - \frac{1}{2}\log(1+\mathbf{g}^T\mathbf{Q}_1\mathbf{g}) \\ \text{s.t.} \quad & \mathrm{tr}(\mathbf{Q}_1)+\mathrm{tr}(\mathbf{Q}_2)\leq P \\ & \mathbf{g}^T(\mathbf{Q}_1+\mathbf{Q}_2)\mathbf{g}\geq \tilde{E} \end{aligned} \tag{10.118}$$

The objective function is generally not concave. Considering the monotonicity of log, the objective function can be replaced with the linear fractional objective function $\frac{1+\mathbf{h}^T\mathbf{Q}_1\mathbf{h}}{1+\mathbf{g}^T\mathbf{Q}_1\mathbf{g}}$. Following the linear fractional transformation [136] by multiplying by positive variable $t>0$ and defining $\mathbf{Q}_1=\tilde{\mathbf{Q}}_1/t$, $\mathbf{Q}_2=\tilde{\mathbf{Q}}_2/t$, and fixing the resultant denominator as $t+\mathbf{g}^T\tilde{\mathbf{Q}}_1\mathbf{g}=1$, we obtain the convex equivalent of the problem in

(10.118) as

$$\begin{aligned} \max_{\tilde{\mathbf{Q}}_1,\tilde{\mathbf{Q}}_2\succeq\mathbf{0},t>0} \quad & t+\mathbf{h}^T\tilde{\mathbf{Q}}_1\mathbf{h} \\ \text{s.t.} \quad & t+\mathbf{g}^T\tilde{\mathbf{Q}}_1\mathbf{g}=1 \\ & \mathrm{tr}(\tilde{\mathbf{Q}}_1)+\mathrm{tr}(\tilde{\mathbf{Q}}_2)\leq tP \\ & \mathbf{h}^T(\tilde{\mathbf{Q}}_1+\tilde{\mathbf{Q}}_2)\mathbf{h}\geq t\tilde{E} \end{aligned} \tag{10.119}$$

The optimal solution of (10.119) can be obtained efficiently using convex solvers, e.g., CVX.

10.6.2 MISO Problem with Gaussian Artificial Noise Based Coding Scheme

In this case, we cannot fully transform the problem to a convex form. However, we can apply similar techniques together with an extra step of line search [137] to solve the problem. The problem in this case is:

$$\begin{aligned} \max_{\mathbf{Q}_1,\mathbf{Q}_2\succeq\mathbf{0}} \quad & \frac{1}{2}\log\left(1+\frac{\mathbf{h}^T\mathbf{Q}_1\mathbf{h}}{1+\mathbf{h}^T\mathbf{Q}_2\mathbf{h}}\right)-\frac{1}{2}\log\underbrace{\left(1+\frac{\mathbf{g}^T\mathbf{Q}_1\mathbf{g}}{1+\mathbf{g}^T\mathbf{Q}_2\mathbf{g}}\right)}_{\leq\beta} \\ \text{s.t.} \quad & \mathrm{tr}(\mathbf{Q}_1)+\mathrm{tr}(\mathbf{Q}_2)\leq P \\ & \mathbf{g}^T(\mathbf{Q}_1+\mathbf{Q}_2)\mathbf{g}\geq\tilde{E} \end{aligned} \tag{10.120}$$

Next, we upper bound the second term in the optimization problem by $\frac{1}{2}\log\beta$, where β is the line-search variable. This results in an extra constraint $\frac{\mathbf{g}^T\mathbf{Q}_1\mathbf{g}}{1+\mathbf{g}^T\mathbf{Q}_2\mathbf{g}}\leq\beta-1$.

We write the optimization problem by considering the monotonicity of log and rearranging terms as:

$$\begin{aligned} \max_{\mathbf{Q}_1,\mathbf{Q}_2\succeq\mathbf{0}} \quad & \frac{1+\mathbf{h}^T(\mathbf{Q}_1+\mathbf{Q}_2)\mathbf{h}}{\beta(1+\mathbf{h}^T\mathbf{Q}_2\mathbf{h})} \\ \text{s.t.} \quad & \mathbf{g}^T(\mathbf{Q}_1-(\beta-1)\mathbf{Q}_2)\mathbf{g}\leq\beta-1 \\ & \mathrm{tr}(\mathbf{Q}_1)+\mathrm{tr}(\mathbf{Q}_2)\leq P \\ & \mathbf{g}^T(\mathbf{Q}_1+\mathbf{Q}_2)\mathbf{g}\geq\tilde{E} \end{aligned} \tag{10.121}$$

Now, by linear fractional transformation [136], we multiply (10.121) by $t>0$, define $\mathbf{Q}_1=\tilde{\mathbf{Q}}_1/t, \mathbf{Q}_2=\tilde{\mathbf{Q}}_2/t$ and fix $\beta(t+\mathbf{h}^T\tilde{\mathbf{Q}}_2\mathbf{h})=1$. Note that using this transformation, the resultant problem is a convex problem for fixed β. Hence, iterating over β along its range $1\leq\beta\leq 1+P\|\mathbf{h}\|^2$, the problem becomes

$$\max_{\beta} \quad \varphi(\beta), \quad \text{s.t.} \quad 1\leq\beta\leq 1+P\|\mathbf{h}\|^2 \tag{10.122}$$

which together with the following can be solved effectively

$$\begin{aligned} \varphi(\beta)=\max_{\tilde{\mathbf{Q}}_1,\tilde{\mathbf{Q}}_2\succeq\mathbf{0},t>0} \quad & t+\mathbf{h}^T(\tilde{\mathbf{Q}}_1+\tilde{\mathbf{Q}}_2)\mathbf{h} \\ \text{s.t.} \quad & \mathbf{g}^T(\tilde{\mathbf{Q}}_1-(\beta-1)\tilde{\mathbf{Q}}_2)\mathbf{g}\leq t(\beta-1) \\ & \beta(t+\mathbf{h}^T\tilde{\mathbf{Q}}_2\mathbf{h})=1 \\ & \mathrm{tr}(\tilde{\mathbf{Q}}_\mathbf{1})+\mathrm{tr}(\tilde{\mathbf{Q}}_2)\leq tP \\ & \mathbf{g}^T(\tilde{\mathbf{Q}}_1+\tilde{\mathbf{Q}}_2)\mathbf{g}\geq t\tilde{E} \end{aligned} \tag{10.123}$$

10.6.3 General MIMO Problem

For the general MIMO case, we cannot provide a direct convex optimization equivalent as in the MISO case even by adding a line search. This is due to the concavity of log-determinant functions, which result in difference of concave functions. To tackle the problem, we can approximate the objective function using sequential convex optimization techniques [138, 139]. The idea here is to approximate the second term in the objective function by its first order expansion. To show that, first, consider the objective function of the Gaussian coding with fixed mean $\frac{1}{2}\log|\mathbf{I}+\mathbf{H}\mathbf{Q}_1\mathbf{H}^T| - \frac{1}{2}\log|\mathbf{I}+\mathbf{G}\mathbf{Q}_1\mathbf{G}^T|$, which is equivalent to $\log|\mathbf{Q}_1+\mathbf{N}_1| - \log|\mathbf{Q}_1+\mathbf{N}_2|$. We approximate the second term with an affine function using the Taylor series expansion of the $\log\det$ function around $\mathbf{Q}^{(k)}$, where k denotes the kth iteration:

$$\log|\mathbf{Q}_1+\mathbf{N}_2| \cong \log|\mathbf{Q}_1^{(k)}+\mathbf{N}_2| + \mathrm{tr}((\mathbf{Q}_1^{(k)}+\mathbf{N}_2)^{-1}(\mathbf{Q}_1-\mathbf{Q}^{(k)})) \tag{10.124}$$

Since the constant terms do not affect the optimal solution, we can use

$$\log|\mathbf{Q}_1+\mathbf{N}_2| \cong \mathrm{tr}((\mathbf{Q}_1^{(k)}+\mathbf{N}_2)^{-1}\mathbf{Q}_1) \tag{10.125}$$

The optimization problem in the kth iteration is

$$\max_{\mathbf{Q}_1,\mathbf{Q}_2\succeq\mathbf{0}} \quad \log|\mathbf{Q}_1+\mathbf{N}_1| - \mathrm{tr}((\mathbf{Q}_1^{(k)}+\mathbf{N}_2)^{-1}\mathbf{Q}_1)$$

$$\text{s.t.} \quad \text{tr}(\mathbf{Q}_1) + \text{tr}(\mathbf{Q}_2) \leq P$$

$$\text{tr}(\mathbf{G}(\mathbf{Q}_1 + \mathbf{Q}_2)\mathbf{G}^T) \geq \tilde{E} \tag{10.126}$$

which is a convex problem, and can be solved efficiently. We update $\mathbf{Q}_1^{(k)}, \mathbf{Q}_2^{(k)}$ by solving such convex optimization problems until convergence.

Finally, using similar ideas, we can perform linearization in the case of Gaussian with artificial noise coding scheme, where the corresponding optimization problem in the kth iteration is

$$\max_{\mathbf{Q}_2, \mathbf{S} \succeq \mathbf{0}} \quad \log|\mathbf{S} + \mathbf{N}_1| + \log|\mathbf{Q}_2 + \mathbf{N}_2| - \text{tr}((\mathbf{Q}_2^{(k)} + \mathbf{N}_1)^{-1}\mathbf{Q}_2) - \text{tr}((\mathbf{S}^{(k)} + \mathbf{N}_2)^{-1}\mathbf{S})$$

$$\text{s.t.} \quad \text{tr}(\mathbf{S}) \leq P, \quad \text{tr}(\mathbf{G}\mathbf{S}\mathbf{G}^T) \geq \tilde{E} \tag{10.127}$$

10.7 Numerical Results

In this section, we present simple simulation results for the secrecy capacity of the MIMO wiretap channel with maximum transmitter-side power constraint and minimum receiver-side (eavesdropper-side) power constraint. In these simulations, the average transmit power at the transmitter is taken as $P = 10$ and the noise covariance is identity at both receivers.

Fig. 10.2 shows a secrecy capacity receiver-side power constraint region for a MISO 4-1-1 system, i.e, a system with 4 antennas at the transmitter and single antenna at both the legitimate receiver and the eavesdropper. The figure shows the optimality of the Gaussian signalling with a mean and Gaussian coding with

Gaussian artificial noise coding schemes; in particular, the regions corresponding to the mean and artificial noise coding schemes are identical. Moreover, the secrecy rate region with receiver-side power region of the standard Gaussian coding scheme with no mean or no artificial noise is noticeably smaller than the optimal schemes. That is, the standard Gaussian signaling scheme is strictly sub-optimal for the case of receiver-side power constraints. In addition, we observe that, as the receiver-side power constraint is increased, the secrecy capacity decreases, i.e., there is a trade-off between the power that should be delivered to the eavesdropper's receiver and the confidentiality that can be provided to the legitimate receiver. This is because, when the receiver-side power constraint is increased, the problem becomes more confined and more power should be concentrated for the receiver-side power constraint, which decreases the set of signalling choices for the secrecy communications. Fig. 10.3 shows similar observations for the 2-2-2 MIMO wiretap system.

10.8 Conclusions

We considered the MIMO wiretap channel with the usual transmitter-side maximum power constraint and an additional receiver-side minimum power constraint. For the converse, we first proved that the problem is equivalent to solving a secrecy capacity problem with a double-sided correlation matrix constraint on the channel input. We then extended the channel enhancement technique to our setting. For the achievability, we proposed two optimum schemes that achieve the converse rate: Gaussian signalling with a fixed mean and Gaussian signalling with Gaussian channel prefixing

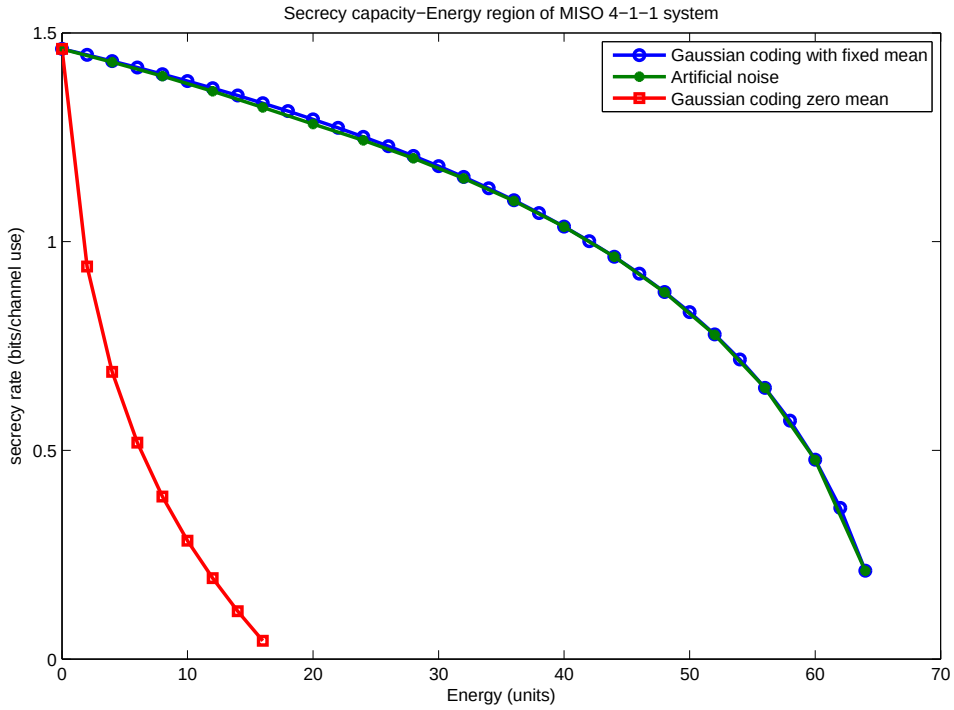


Figure 10.2: Secrecy capacity receiver-side power constraint region for a 4-1-1 MISO wiretap channel.

(artificial noise). This is the first instance of a problem where transmission with a mean or channel prefixing are strictly necessary for a MIMO wiretap channel under power constraints. The transmission scheme with a mean enables us to deliver the needed power to the receiver without creating interference to the legitimate receiver as it is a deterministic signal. On the other hand, the transmission scheme with Gaussian artificial noise, both jams the eavesdropper contributing to the secrecy as well as delivering the needed power to the receiver. We note that the optimal coding scheme for the MIMO wiretap channel under a transmitter-side power constraint only, which is Gaussian signalling with no channel prefixing or mean, is strictly sub-optimal when we impose a receiver-side power constraint, showing similar to

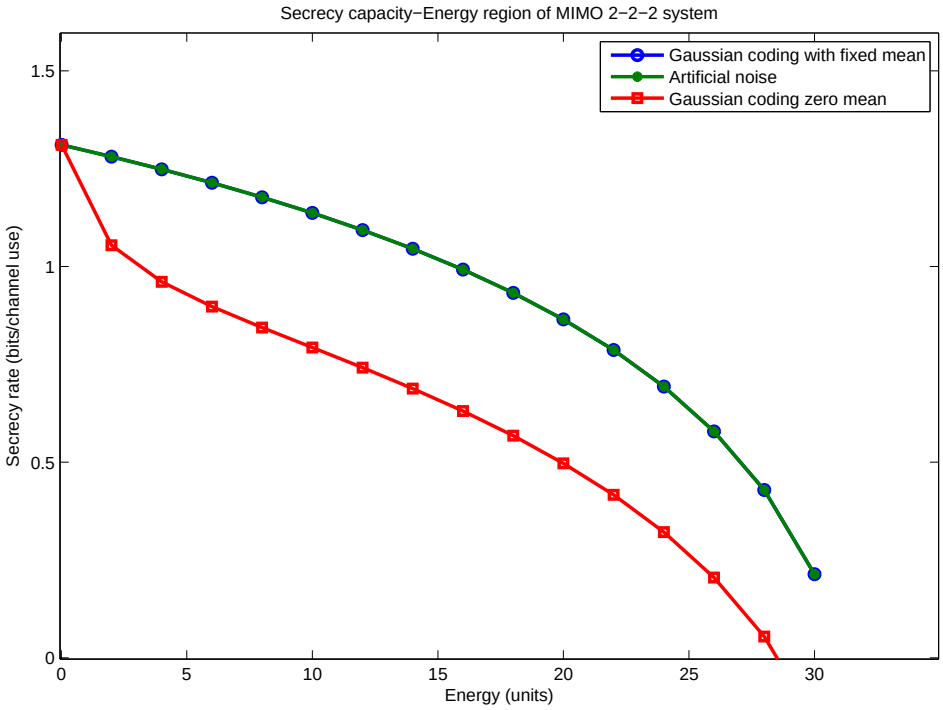


Figure 10.3: Secrecy capacity receiver-side power constraint region for a 2-2-2 MIMO wiretap channel.

the cases of [107, 108], that receiver-side power constraints may change the solution significantly and may introduce non-trivial trade-offs. We then extended our setting to the cases of minimum power constraints at both receivers in a wiretap channel; maximum receiver-side power constraints at both receivers in a wiretap channel; minimum receiver-side power constraints in a broadcast channel (i.e., no secrecy constraints); and minimum receiver-side power constraints in a broadcast channel with confidential messages (i.e., double-sided secrecy constraints).

10.9 Appendix: Continuity of the Capacity Function

We prove our claim in Lemma 10.2 that $C(\mathbf{S}_1, \mathbf{S}_2, \mathbf{H}, \mathbf{G})$ is a continuous function with respect to $\mathbf{S}_2$. Although contiguity defined in [102], which is a weaker notion than continuity, suffices to prove Lemma 10.2, we prove continuity here. To prove this, we begin by writing the optimization problem in a general form as in [102, Appendix IV] by concatenating the rows of $\mathbf{Q}_1, \mathbf{Q}_2$ to form a vector $\mathbf{y} \in \mathbb{R}^{2t^2}$, where $t = \max\{N_t, N_r\}$. We denote the point-to-set map $\Omega(\mathbf{S}_2)$ to be a mapping from $\mathbf{S}_2$ to the power set of all subsets of the corresponding feasible set, i.e.,

$$\Omega(\mathbf{S}_2) = \{\text{row concatenation of } (\mathbf{Q}_1, \mathbf{Q}_2) : \mathbf{Q}_1, \mathbf{Q}_2 \succeq \mathbf{0},\ \mathbf{S}_1 \preceq \mathbf{Q}_1 + \mathbf{Q}_2 \preceq \mathbf{S}_2\} \tag{10.128}$$

Denote $C(\mathbf{S}_1, \mathbf{S}_2, \mathbf{H}, \mathbf{G})$ by $C(\mathbf{S}_2)$ for notational simplicity as we focus on the argument $\mathbf{S}_2$ here. From (10.11) with $\mathbf{Q}_2 = \boldsymbol{\mu}\boldsymbol{\mu}^T$, we write $C(\mathbf{S}_2)$ as

$$C(\mathbf{S}_2) = \max_{\mathbf{y} \in \Omega(\mathbf{S}_2)} f(\mathbf{y}) \tag{10.129}$$

where $f(\mathbf{y}) = \frac{1}{2}\log|\mathbf{I} + \mathbf{H}\mathbf{Q}_1\mathbf{H}^T| - \frac{1}{2}\log|\mathbf{I} + \mathbf{G}\mathbf{Q}_1\mathbf{G}^T|$. Note that in this case $f(\mathbf{y})$ depends only on the first t^2 elements of $\mathbf{y}$. Now, we use [140, Theorem 7], which states conditions on the continuity of the optimal value function in mathematical programming to prove the continuity of $C(\mathbf{S}_2)$. In the sequel, we verify that all requirements of [140, Theorem 7] are satisfied.

Since the determinant of an $n \times n$ matrix $\mathbf{A}$ can be written as $\det(\mathbf{A}) =$

$\sum_\sigma \text{sgn}(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)}$, where the sum is over all $n!$ permutations of $\{1, 2, \cdots, n\}$, the determinant in this form is a polynomial in n^2 variables, and $\det(\mathbf{A})$ is continuous. Consequently, $f(\mathbf{y})$ is also continuous. $\Omega(\mathbf{S}_2)$ consists of linear matrix inequalities, hence it is a continuous point-to-set map. Furthermore, $\Omega(\mathbf{S}_2)$ is uniformly compact because for any sequence $\mathbf{S}_2^{(i)}$ in the neighborhood of $\mathbf{S}_2$, i.e., the metric distance $d(\mathbf{S}_2^{(i)}, \mathbf{S}_2) = \text{tr}\left((\mathbf{S}_2^{(i)} - \mathbf{S}_2)(\mathbf{S}_2^{(i)} - \mathbf{S}_2)^T\right) \leq \delta^2$ for some finite $\delta > 0$, one can find $k_i = \max \lambda(\mathbf{S}_2^{(i)})$ where $\lambda(\mathbf{S}_2^{(i)})$ is an eigenvalue of matrix $\mathbf{S}_2^{(i)}$ such that

$$\Omega(\mathbf{S}_2^{(i)}) \subseteq \mathcal{Y} = \{\text{row concatenation of } (\mathbf{Q}_1, \mathbf{Q}_2) : \mathbf{Q}_1, \mathbf{Q}_2 \succeq \mathbf{0},\ \text{tr}(\mathbf{Q}_1 + \mathbf{Q}_2) \leq k\} \tag{10.130}$$

where $k = \max_i k_i \leq P + \delta$, where P is the power constraint imposed on $\mathcal{S}_{PE}$. Since $\mathcal{Y}$ is compact and contains $\bigcup_i \Omega(\mathbf{S}_2^{(i)})$, $\Omega(\mathbf{S}_2)$ is uniformly compact. Hence, the requirements of [140, Theorem 7] are satisfied and $C(\mathbf{S}_1, \mathbf{S}_2, \mathbf{H}, \mathbf{G})$ is continuous with respect to $\mathbf{S}_2$.

CHAPTER 11

Conclusions

In this dissertation, we used the information-theoretic approach to characterize the fundamental limits of information retrieval and communication rates under privacy and/or security constraints for the next generation networks.

In Chapters 2-6, we focused on the private information retrieval (PIR) problem while taking into account various practical constraints.

In Chapter 2, we considered the PIR problem over MDS-coded (CPIR) and non-colluding databases. We employed information-theoretic arguments to derive the optimal retrieval rate for the desired message for any given MDS storage code. Our result shows that the optimal retrieval cost is independent of the explicit structure of the storage code, and the number of databases, but depends only on the code rate and the number of messages. Interestingly, the result implies that there is no gain to be obtained from a joint design of the MDS storage code and the retrieval procedure. The result also establishes a fundamental tradeoff between the code rate and the PIR capacity for MDS codes.

In Chapter 3, we introduced the multi-message private information retrieval

(MPIR) problem. We determined the exact sum capacity for this problem when the number of desired messages is at least half of the number of total stored messages and when the total number of messages is an integer multiple of the number of desired messages. For the remaining cases, we derived lower and upper bounds. We observed numerically that the gap between the lower and upper bounds decreases monotonically in the number of databases, and the worst case gap is 0.0082. The result implies that joint retrieval of the desired messages strictly outperforms repeating the single-message capacity achieving scheme for each message.

In Chapter 4, we investigated the PIR problem in the presence of Byzantine and colluding databases (BPIR). We determined the exact capacity of the BPIR problem. The capacity expression shows the severe degradation in the retrieval rate in the presence of Byzantine databases. The expression shows that in order to correct the errors introduced by the adversarial databases, the system needs to have twice the number of Byzantine databases as redundant storage nodes. The retrieval rate is further penalized by a multiplicative factor, which reflects the ignorance of the user to the identity of the honest databases. We extended the optimal scheme for the RPIR problem to permit error correction of any error pattern introduced by the Byzantine databases. For the converse, we adapted the cut-set bound, which was originally derived for the network coding problem against adversarial nodes, for the PIR setting.

In Chapter 5, we introduced the PIR problem under asymmetric traffic constraints. We investigated the fundamental limits of this problem by developing a novel upper bound. The upper bound generalizes the converse proof for the case

of classical PIR, which inherently utilizes database symmetry. The upper bound is a piece-wise affine function in the traffic constraints. The upper bound implies a strict capacity loss due to the asymmetric traffic constraints for certain cases. We developed explicit achievable schemes for specific corner points, and achieved the remaining points by time-sharing. We described the achievable scheme by means of a system of difference equations. We proved that the upper bound and the lower bound exactly match for the cases of 2 and 3 messages for any number of databases.

In Chapter 6, we introduced noisy PIR with orthogonal links (NPIR), and PIR from multiple access channels (MAC-PIR). We focused on the issue of separability of the channel coding and the retrieval scheme. For the NPIR problem, we proved that the channel coding and the retrieval scheme are almost separable in the sense that every database implements its own channel coding independently of the other databases. The problem is coupled only through agreeing on a suitable traffic ratio vector to maximize the retrieval rate. However, these conclusions are not valid for the MAC-PIR problem. We showed two examples, namely: PIR from additive MAC and PIR from logical conjunction/disjunction MAC, where the channel coding and retrieval schemes are indeed inseparable unlike the NPIR problem. In both cases, we showed that we can attain the full capacity with no loss due to privacy.

In Chapter 7, we investigated joint security and privacy constraints by considering the PIR problem from wiretap channel II (PIR-WTC-II). We showed that the problem is a concrete example of the PIR problem under asymmetric traffic constraints. We obtained a general upper bound that takes the form of a max-min optimization problem. The inner minimization problem derives the tightest upper

bound for the retrieval rate for an arbitrary traffic ratio vector, while the outer maximization problem optimizes over traffic ratio vector. The security constraint is satisfied by (information-theoretically) encrypting each returned answering string by an artificial noise vector. To obtain the artificial noise vector, each database generates a secret key and encodes it into artificial noise by an MDS code. The upper and lower bounds match for the cases of 2 and 3 messages for any number of databases, and for any eavesdropping capability vector.

In Chapters 8-10, we focused on the security problem in multi-user networks.

In Chapter 8, we determined the exact s.d.o.f. region of a two-user $M \times N$ MIMO ICCM. For the converse proof, we combined three distinct upper bounds: the cooperative upper bound which treats ICCM as a BCCM; the upper bounding technique that uses vectorized versions of secrecy penalty and role of a helper lemmas; and the IC upper bound without any secrecy constraints. For achievability, we showed that the s.d.o.f. region is a four-vertex polytope. To that end, we developed a novel achievable scheme for the basic 2×2 MIMO ICCM, which is central for achievable schemes for general M and N. The 2×2 scheme combines spatial alignment for secrecy and asymptotic real interference alignment for decodability. We showed the achievability of the other non-trivial polytope corner points by forcing one of the users to act as a cooperative jammer (helper) that jams its own receiver. Interestingly, we showed that the s.d.o.f. region starts as a square region, then it takes the shape of an irregular polytope until it returns back to a square region when the number of transmit antennas is at least twice the number of receiving antennas. We showed that if the ICCM channel is time-varying, the achievable schemes can

be simplified by using vector space alignment instead of real interference alignment that is necessary for static channels.

In Chapter 9, we introduced three new channel models: broadcast channel with combating helpers, interference channel with selfish users, and multiple access wiretap channel with deviating users. We aimed at studying the effects of selfishness and malicious behaviour on secrecy in networks. For BCCM with combating helpers, we showed that the malicious intentions of the helpers are neutralized and the full s.d.o.f. is retained. On the contrary, for the ICCM with selfish users, we showed that selfishness precludes secure communication and no s.d.o.f. is achieved. For the MAC-WTC with deviating users, we considered two kinds of deviation: when some of the users stop transmitting cooperative jamming signals, and when a user starts sending intentional jamming signals. For the first scenario, we investigated possible responses of the remaining users to counteract such deviation. For the second scenario, we showed that although a deviating user can drive the sum s.d.o.f. to zero, the remaining users can exploit the jamming signals as cooperative jamming signals against the eavesdropper and achieve an optimum s.d.o.f.

In Chapter 10, we considered the MIMO wiretap channel with the usual transmitter-side maximum power constraint and an additional receiver-side minimum power constraint. For the converse, we first proved that the problem is equivalent to solving a secrecy capacity problem with a double-sided correlation matrix constraint on the channel input. We then extended the channel enhancement technique to our setting. For the achievability, we proposed two optimum schemes that achieve the converse rate: Gaussian signalling with a fixed mean and

Gaussian signalling with Gaussian artificial noise. This is the first instance of a problem where transmission with a mean or channel prefixing are strictly necessary for a MIMO wiretap channel under power constraints, showing that receiver-side power constraints may change the solution significantly and may introduce non-trivial tradeoffs. We then extended our setting to the cases of minimum power constraints at both receivers in a wiretap channel; maximum receiver-side power constraints at both receivers in a wiretap channel; minimum receiver-side power constraints in a broadcast channel; and minimum receiver-side power constraints in a broadcast channel with confidential messages.

The contents of Chapter 2 are published in [116, 117], Chapter 3 in [123, 141], Chapter 4 in [120, 142], Chapter 5 in [125, 143], Chapter 6 in [144–146], Chapter 7 in [124, 147], Chapter 8 in [148, 149], Chapter 9 in [150, 151], and Chapter 10 in [152, 153].

Bibliography

[1] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. Private information retrieval. *Journal of the ACM*, 45(6):965–981, November 1998.

[2] W. Gasarch. A survey on private information retrieval. In *Bulletin of the EATCS*, volume 82, pages 72–107, 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, April 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. Blind interference alignment for private information retrieval. In *IEEE ISIT*, July 2016.

[12] H. Sun and S. A. Jafar. The capacity of private information retrieval. *IEEE Trans. on Info. Theory*, 63(7):4075–4088, July 2017.

[13] S. 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. The capacity of robust private information retrieval with colluding databases. *IEEE Trans. on Info. Theory*, 64(4):2361–2370, April 2018.

[15] 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.

[16] R. Tajeddine and S. E. Rouayheb. Robust private information retrieval on coded data. In *IEEE ISIT*, June 2017.

[17] H. Sun and S. Jafar. The capacity of symmetric private information retrieval. 2016. Available at arXiv:1606.08828.

[18] H. Sun and S. A. Jafar. Optimal download cost of private information retrieval for arbitrary message length. *IEEE Trans. on Info. Forensics and Security*, 12(12):2920–2932, December 2017.

[19] H. Sun and S. A. Jafar. Multiround private information retrieval: Capacity and storage overhead. *IEEE Trans. on Info. Theory*, 64(8):5743–5754, August 2018.

[20] Q. Wang and M. Skoglund. Symmetric private information retrieval for MDS coded distributed storage. In *IEEE ICC*, May 2017.

[21] Q. Wang and M. Skoglund. Secure symmetric private information retrieval from colluding databases with adversaries. In *IEEE Allerton*, October 2017.

[22] 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.

[23] 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.

[24] Y. Zhang and G. Ge. A general private information retrieval scheme for MDS coded databases with colluding servers. 2017. Available at arXiv: 1704.06785.

[25] Y. Zhang and G. Ge. Multi-file private information retrieval from MDS coded databases with colluding servers. 2017. Available at arXiv: 1705.03186.

[26] 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. In *IEEE ISIT*, June 2018.

[27] Q. Wang and M. Skoglund. Linear symmetric private information retrieval for MDS coded distributed storage with colluding servers. 2017. Available at arXiv:1708.05673.

[28] R. Tandon. The capacity of cache aided private information retrieval. In *IEEE Allerton*, October 2017.

[29] Y.-P. Wei, K. Banawan, and S. Ulukus. Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching. 2017. Available at arXiv:1709.01056.

[30] S. Kadhe, B. Garcia, A. Heidarzadeh, S. El Rouayheb, and A. Sprintson. Private information retrieval with side information. 2017. Available at arXiv:1709.00112.

[31] Z. Chen, Z. Wang, and S. Jafar. The capacity of private information retrieval with private side information. 2017. Available at arXiv:1709.03022.

[32] 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.

[33] Y.-P. Wei, K. Banawan, and S. Ulukus. Cache-aided private information retrieval with partially known uncoded prefetching: Fundamental limits. *IEEE Journal on Selected Areas in Communications*, 36(6):1126–1139, June 2018.

[34] 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.

[35] S. Li and M. Gastpar. Single-server multi-message private information retrieval with side information. 2018. Available at arXiv:1808.05797.

[36] H. Sun and S. A. Jafar. The capacity of private computation. 2017. Available at arXiv:1710.11098.

[37] M. Mirmohseni and M. A. Maddah-Ali. Private function retrieval. 2017. Available at arXiv:1711.04677.

[38] Z. Chen, Z. Wang, and S. Jafar. The asymptotic capacity of private search. In *IEEE ISIT*, June 2018.

[39] 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.

[40] M. Attia, D. Kumar, and R. Tandon. The capacity of private information retrieval from uncoded storage constrained databases. 2018. Available at arXiv:1805.04104.

[41] Q. Wang and M. Skoglund. Secure private information retrieval from colluding databases with eavesdroppers. 2017. Available at arXiv: 1710.01190.

[42] Q. Wang, H. Sun, and M. Skoglund. The capacity of private information retrieval with eavesdroppers. 2018. Available at arXiv:1804.10189.

[43] H. Yang, W. Shin, and J. Lee. Private information retrieval for secure distributed storage systems. *IEEE Trans. on Info. Forensics and Security*, 13(12):2953–2964, December 2018.

[44] Z. Jia, H. Sun, and S. Jafar. Cross subspace alignment and the asymptotic capacity of X-secure T-private information retrieval. 2018. Available at arXiv:1808.07457.

[45] A. Dimakis, K. Ramchandran, Y. Wu, and C. Suh. A survey on network codes for distributed storage. *Proceedings of the IEEE*, 99(3):476–489, March 2011.

[46] A. Fazeli, A. Vardy, and E. Yaakobi. PIR with low storage overhead: coding instead of replication. 2015. Available at arXiv:1505.06241.

[47] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012.

[48] R. Henry, Y. Huang, and I. Goldberg. One (block) size fits all: PIR and SPIR with variable-length records via multi-block queries. In *NDSS*, 2013.

[49] D. Demmler, A. Herzberg, and T. Schneider. RAID-PIR: Practical multi-server PIR. In *ACM Workshop on Cloud Computing Security*, pages 45–56, 2014.

[50] L. Wang, T. K. Kuppusamy, Y. Liu, and J. Cappos. A fast multi-server, multi-block private information retrieval protocol. In *IEEE Globecom*, December 2015.

[51] J. Groth, A. Kiayias, and H. Lipmaa. Multi-query computationally-private information retrieval with constant communication rate. In *International Workshop on Public Key Cryptography*. Springer, May 2010.

[52] Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai. Batch codes and their applications. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 262–271. ACM, 2004.

[53] W. Lueks and I. Goldberg. Sublinear scaling for multi-client private information retrieval. In *International Conference on Financial Cryptography and Data Security*, pages 168–186. Springer, 2015.

[54] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing*. Pearson Higher Education, 2010.

[55] A. Beimel and Y. Stahl. Robust information-theoretic private information retrieval. In *International Conference on Security in Communication Networks*, pages 326–341. Springer, 2002.

[56] C. Devet, I. Goldberg, and N. Heninger. Optimally robust private information retrieval. In *USENIX Security Symposium*, August 2012.

[57] E. Y. Yang, J. Xu, and K. H. Bennett. Private information retrieval in the presence of malicious failures. In *Proceedings 26th Annual International Computer Software and Applications*, August 2002.

[58] A. Shamir. How to share a secret. *Communications of the ACM*, 22(11):612–613, 1979.

[59] V. Guruswami and M. Sudan. Improved decoding of Reed-Solomon and algebraic-geometric codes. In *39th Annual Symposium on Foundations of Computer Science*, pages 28–37, November 1998.

[60] O. Kosut, L. Tong, and D. N. C. Tse. Polytope codes against adversaries in networks. *IEEE Trans. on Info. Theory*, 60(6):3308–3344, June 2014.

[61] S. Pawar, S. El Rouayheb, and K. Ramchandran. Securing dynamic distributed storage systems against eavesdropping and adversarial attacks. *IEEE Trans. on Info.Theory*, 57(10):6734–6753, October 2011.

[62] R. Singleton. Maximum distance Q-nary codes. *IEEE Trans. on Info. Theory*, 10(2):116–118, April 1964.

[63] J. Garay, R. Gennaro, C. Jutla, and T. Rabin. Secure distributed storage and retrieval. *Theoretical Computer Science*, 243(1):363 – 389, July 2000.

[64] L. H. Ozarow and A. D. Wyner. Wire-tap channel II. *AT&T Bell Laboratories Technical Journal*, 63(10):2135–2157, December 1984.

[65] Y. Liang, H. V. Poor, and S. Shamai. Information theoretic security. *Foundations and Trends in Communications and Information Theory*, 5(4–5):355–580, 2009.

[66] A. D. Wyner. The wire-tap channel. *The Bell System Technical Journal*, 54(8):1355–1387, October 1975.

[67] I. Csiszar and J. Korner. Broadcast channels with confidential messages. *IEEE Trans. on Info. Theory*, 24(3):339–348, May 1978.

[68] S. Leung-Yan-Cheong and M. Hellman. The Gaussian wire-tap channel. *IEEE Trans. on Info. Theory*, 24(4):451–456, July 1978.

[69] E. Tekin and A. Yener. The Gaussian multiple access wire-tap channel. *IEEE Trans. on Info. Theory*, 54(12):5747–5755, December 2008.

[70] R. Liu, I. Maric, P. Spasojevic, and R. D. Yates. Discrete memoryless interference and broadcast channels with confidential messages: Secrecy rate regions. *IEEE Trans. on Info. Theory*, 54(6):2493–2507, June 2008.

[71] H. F. Chong and Y. C. Liang. Secrecy capacity region of a class of two-user Gaussian MIMO BC with degraded message sets. In *IEEE ISIT*, July 2013.

[72] R. Liu, T. Liu, H. V. Poor, and S. Shamai. Multiple-input multiple-output Gaussian broadcast channels with confidential messages. *IEEE Trans. on Info. Theory*, 56(9):4215–4227, September 2010.

[73] A. Khina, Y. Kochman, and A. Khisti. The confidential MIMO broadcast capacity: A simple derivation. In *IEEE ISIT*, June 2015.

[74] Z. Goldfeld and H. Permuter. MIMO Gaussian broadcast channels with common, private and confidential messages. 2016. Available at arXiv:1608.06057.

[75] A. Khisti, A. Tchamkerten, and G. W. Wornell. Secure broadcasting over fading channels. *IEEE Trans. on Info. Theory*, 54(6):2453–2469, June 2008.

[76] E. Ekrem and S. Ulukus. The secrecy capacity region of the Gaussian MIMO multi-receiver wiretap channel. *IEEE Trans. on Info. Theory*, 57(4):2083–2114, April 2011.

[77] L. Lai and H. El Gamal. The relay-eavesdropper channel: Cooperation for secrecy. *IEEE Trans. on Info. Theory*, 54(9):4005–4019, September 2008.

[78] J. Xie and S. Ulukus. Secure degrees of freedom of one-hop wireless networks. *IEEE Trans. on Info. Theory*, 60(6):3359–3378, June 2014.

[79] M. Nafea and A. Yener. Secure degrees of freedom for the MIMO wire-tap channel with a multi-antenna cooperative jammer. *IEEE Trans. on Info. Theory*, 63(11):7420–7441, November 2017.

[80] G. Bagherikaram, A. S. Motahari, and A. K. Khandani. On the secure DoF of the single-antenna MAC. In *IEEE ISIT*, June 2010.

[81] X. He, A. Khisti, and A. Yener. MIMO multiple access channel with an arbitrarily varying eavesdropper: Secrecy degrees of freedom. *IEEE Trans. on Info. Theory*, 59(8):4733–4745, August 2013.

[82] J. Xie and S. Ulukus. Secure degrees of freedom regions of multiple access and interference channels: The polytope structure. *IEEE Trans. on Info. Theory*, 62(4):2044–2069, April 2016.

[83] O. O. Koyluoglu, H. El Gamal, L. Lai, and H. V. Poor. On the secure degrees of freedom in the K-user Gaussian interference channel. In *IEEE ISIT*, July 2008.

[84] X. He and A. Yener. K-user interference channels: Achievable secrecy rate and degrees of freedom. In *IEEE ITW*, June 2009.

[85] X. He and A. Yener. Secure degrees of freedom for Gaussian channels with interference: Structured codes outperform Gaussian signaling. In *IEEE Globecomm*, November 2009.

[86] J. Xie and S. Ulukus. Secure degrees of freedom of K-user Gaussian interference channels: A unified view. *IEEE Trans. on Info. Theory*, 61(5):2647–2661, May 2015.

[87] T. Gou and S. A. Jafar. On the secure degrees of freedom of wireless X networks. In *IEEE Allerton*, September 2008.

[88] Z. Wang, M. Xiao, M. Skoglund, and H. V. Poor. Secure degrees of freedom of wireless X networks using artificial noise alignment. *IEEE Trans. on Comm.*, 63(7):2632–2646, July 2015.

[89] T. T. Kim and H. V. Poor. On the secure degrees of freedom of relaying with half-duplex feedback. *IEEE Trans. on Info.Theory*, 57(1):291–302, January 2011.

[90] A. Khisti. Interference alignment for the multiantenna compound wiretap channel. *IEEE Trans. on Info. Theory*, 57(5):2976–2993, May 2011.

[91] S. H. Lee, W. Zhao, and A. Khisti. Secure degrees of freedom of the Gaussian diamond-wiretap channel. *IEEE Trans. on Info. Theory*, 63(1):496–508, January 2017.

[92] R. Tandon, S. Mohajer, H. V. Poor, and S. Shamai. Degrees of freedom region of the MIMO interference channel with output feedback and delayed CSIT. *IEEE Trans. on Info. Theory*, 59(3):1444–1457, March 2013.

[93] S. Yang, M. Kobayashi, P. Piantanida, and S. Shamai. Secrecy degrees of freedom of MIMO broadcast channels with delayed CSIT. *IEEE Trans. on Info. Theory*, 59(9):5244–5256, September 2013.

[94] A. Zaidi, Z. H. Awan, S. Shamai, and L. Vandendorpe. Secure degrees of freedom of MIMO X-channels with output feedback and delayed CSIT. *IEEE Trans. on Info. Forensics and Security*, 8(11):1760–1774, November 2013.

[95] P. Mukherjee, J. Xie, and S. Ulukus. Secure degrees of freedom of one-hop wireless networks with no eavesdropper CSIT. *IEEE Trans. on Info. Theory*, 63(3):1898–1922, March 2017.

[96] P. Mukherjee, R. Tandon, and S. Ulukus. Secure degrees of freedom region of the two-user MISO broadcast channel with alternating CSIT. *IEEE Trans. on Info. Theory*, 63(6):3823–3853, June 2017.

[97] S. Shafiee, N. Liu, and S. Ulukus. Towards the secrecy capacity of the Gaussian MIMO wire-tap channel: The 2-2-1 channel. *IEEE Trans. on Info. Theory*, 55(9):4033–4039, September 2009.

[98] A. Khisti and G. W. Wornell. Secure transmission with multiple antennas – Part II: The MIMOME wiretap channel. *IEEE Trans. on Info. Theory*, 56(11):5515–5532, November 2010.

[99] F. Oggier and B. Hassibi. The secrecy capacity of the MIMO wiretap channel. *IEEE Trans. on Inform. Theory*, 57(8):4961–4972, August 2011.

[100] T. Liu and S. Shamai. A note on the secrecy capacity of the multiple-antenna wiretap channel. *IEEE Trans. on Info. Theory*, 55(6):2547–2553, June 2009.

[101] A. S. Motahari, S. Oveis-Gharan, M. A. Maddah-Ali, and A. K. Khandani. Real interference alignment: Exploiting the potential of single antenna systems. *IEEE Trans. on Info. Theory*, 60(8):4799–4810, August 2014.

[102] H. Weingarten, Y. Steinberg, and S. Shamai. The capacity region of the Gaussian multiple-input multiple-output broadcast channel. *IEEE Trans. on Info. Theory*, 52(9):3936–3964, September 2006.

[103] R. H. Etkin, D. N. C. Tse, and H. Wang. Gaussian interference channel capacity to within one bit. *IEEE Trans. on Info. Theory*, 54(12):5534–5562, December 2008.

[104] V. R. Cadambe and S. A. Jafar. Interference alignment and degrees of freedom of the K-user interference channel. *IEEE Trans. on Info. Theory*, 54(8):3425–3441, August 2008.

[105] S. A. Jafar and M. J. Fakhereddin. Degrees of freedom for the MIMO interference channel. *IEEE Trans. on Info. Theory*, 53(7):2637–2642, July 2007.

[106] S. Tadelis. *Game Theory: An Introduction*. Princeton Univ. Press, 2013.

[107] M. Gastpar. On capacity under receive and spatial spectrum-sharing constraints. *IEEE Trans. on Info. Theory*, 53(2):471–487, February 2007.

[108] L. R. Varshney. Transporting information and energy simultaneously. In *IEEE ISIT*, July 2008.

[109] J. G. Smith. The information capacity of amplitude and variance-constrained scalar Gaussian channels. *Information and Control*, 18:203–219, April 1971.

[110] S. Goel and R. Negi. Guaranteeing secrecy using artificial noise. *IEEE Trans. on Wireless Comm.*, 7(6):2180–2189, June 2008.

[111] E. Tekin and A. Yener. The general Gaussian multiple-access and two-way wiretap channels: Achievable rates and cooperative jamming. *IEEE Trans. on Info. Theory*, 54(6):2735–2751, June 2008.

[112] L. Liu, R. Zhang, and K-C. Chua. Secrecy wireless information and power transfer with MISO beamforming. *IEEE Trans. on Signal Proc.*, 62(7):1850–1863, April 2014.

[113] D. W. K. Ng, E. S. Lo, and R. Schober. Robust beamforming for secure communication in systems with wireless information and power transfer. *IEEE Trans. on Wireless Comm.*, 13(8):4599–4615, August 2014.

[114] J. Zhang, C. Yuen, C.-K. Wen, S. Jin, K.-K. Wong, and H. Zhu. Achievable ergodic secrecy rate for MIMO SWIPT wiretap channels. In *IEEE ICC*, June 2015.

[115] J. Zhang, C. Yuen, C.-K. Wen, S. Jin, K.-K. Wong, and H. Zhu. Large system secrecy rate analysis for SWIPT MIMO wiretap channels. *IEEE Trans. on Info. Forensics and Security*, 11(1):74–85, January 2016.

[116] K. Banawan and S. Ulukus. Private information retrieval from coded databases. In *IEEE ICC*, May 2017.

[117] K. Banawan and S. Ulukus. The capacity of private information retrieval from coded databases. *IEEE Trans. on Info. Theory*, 64(3):1945–1956, March 2018.

[118] I. S. Reed and G. Solomon. Polynomial codes over certain finite fields. *SIAM*, 8(2):300–304, June 1960.

[119] S. B. Wicker and V. K. Bhargava. *Reed-Solomon Codes and their Applications*. John Wiley & Sons, 1999.

[120] K. Banawan and S. Ulukus. The capacity of private information retrieval from Byzantine and colluding databases. *IEEE Trans. on Info. Theory*. To appear. Also available at arXiv:1706.01442.

[121] C. Feyling. Punctured maximum distance separable codes. *Electronics Letters*, 29(5):470–471, March 1993.

[122] R. Roth. *Introduction to Coding Theory*. Cambridge University Press, 2006.

[123] K. Banawan and S. Ulukus. Multi-message private information retrieval: Capacity results and near-optimal schemes. *IEEE Trans. on Info. Theory*, 64(10):6842–6862, October 2018.

[124] K. Banawan and S. Ulukus. Private information retrieval through wiretap channel II: Privacy meets security. *IEEE Trans. on Info. Theory*. Submitted January 2018. Also available at arXiv:1801.06171.

[125] K. Banawan and S. Ulukus. Asymmetry hurts: Private information retrieval under asymmetric-traffic constraints. *IEEE Trans. on Info. Theory*. Submitted January 2018. Also available at arXiv:1801.03079.

[126] Z. Wang, R. F. Schaefer, M. Skoglund, H. V. Poor, and M. Xiao. Strong secrecy for interference channels from channel resolvability. In *IEEE Asilomar*, November 2015.

[127] M. Nafea and A. Yener. New models for interference and broadcast channels with confidential messages. In *IEEE ISIT*, June 2017.

[128] M. Bloch and J. Barros. *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge University Press, 2011.

[129] A. El Gamal and Y.-H. Kim. *Network Information Theory*. Cambridge university press, 2011.

[130] R. Gallager. *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[131] V. Bryant. *Metric Spaces: Iteration and Application*. Cambridge University Press, 1985.

[132] H. L. Royden and P. Fitzpatrick. *Real Analysis*. Prentice Hall, 2010.

[133] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2012.

[134] A. R. Meenakshi and C. Rajian. On a product of positive semidefinite matrices. *Linear Algebra and its Applications*, 295(1):3–6, July 1999.

[135] S. Luo, J. Xu, T. J. Lim, and R. Zhang. Capacity region of MISO broadcast channel with SWIPT. In *IEEE ICC*, June 2015.

[136] S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[137] J. Li and A. Petropulu. Optimal input covariance for achieving secrecy capacity in Gaussian MIMO wiretap channels. In *IEEE ICASSP*, March 2010.

[138] T. Wang and L. Vandendorpe. Successive convex approximation based methods for dynamic spectrum management. In *IEEE ICC*, pages 4061–4065, June 2012.

[139] T. Lipp and S. P. Boyd. Variations and extensions of the convex-concave procedure, 2014. Available at `http://web.stanford.edu/~boyd/papers/pdf/cvx_ccv.pdf`.

[140] W. Hogan. Point-to-set maps in mathematical programming. *SIAM Review*, 15(3):591–603, July 1973.

[141] K. Banawan and S. Ulukus. Multi-message private information retrieval. In *IEEE ISIT*, June 2017.

[142] K. Banawan and S. Ulukus. Private information retrieval from Byzantine and colluding databases. In *IEEE Allerton*, October 2017.

[143] K. Banawan and S. Ulukus. Private information retrieval under asymmetric traffic constraints. In *IEEE ISIT*, June 2018.

[144] K. Banawan and S. Ulukus. Noisy private information retrieval. In *IEEE Asilomar*, October 2018.

[145] K. Banawan and S. Ulukus. Private information retrieval from multiple access channels. In *IEEE ITW*, November 2018.

[146] K. Banawan and S. Ulukus. Noisy private information retrieval: Separability of channel coding and information retrieval. *IEEE Trans. on Info. Theory.* Submitted July 2018. Available at arXiv: 1807.05997.

[147] K. Banawan and S. Ulukus. Private information retrieval through wiretap channel II. In *IEEE ISIT*, June 2018.

[148] K. Banawan and S. Ulukus. Secure degrees of freedom of the Gaussian MIMO interference channel. In *IEEE Asilomar*, November 2015.

[149] K. Banawan and S. Ulukus. Secure degrees of freedom region of static and time-varying Gaussian MIMO interference channel. *IEEE Trans. on Info. Theory.* To appear.

[150] K. Banawan and S. Ulukus. Secrecy in broadcast channel with combating helpers and interference channel with selfish users. In *IEEE ISIT*, July 2016.

[151] K. Banawan and S. Ulukus. Achievable secrecy rates in the multiple access wiretap channel with deviating users. In *IEEE ISIT*, July 2016.

[152] K. Banawan and S. Ulukus. Gaussian MIMO wiretap channel under receiver side power constraints. In *IEEE Allerton*, September 2014.

[153] K. Banawan and S. Ulukus. MIMO wiretap channel under receiver-side power constraints with applications to wireless power transfer and cognitive radio. *IEEE Trans.on Comm.*, 64(9):3872–3885, September 2016.