**Title**
On the Capacity of Weakly-Private Information Retrieval



**Author**
Jia, Zhuqing

**Publication Date**
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

# On the Capacity of Weakly-Private Information Retrieval

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Electrical Engineering

by

Zhuqing Jia

Thesis Committee:
Professor Syed Jafar, Chair
Professor Hamid Jafarkhani
Assistant Professor Zhiying Wang

2019

# TABLE OF CONTENTS

# LIST OF FIGURES

Page

# LIST OF TABLES

# ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my committee chair, Professor Syed Jafar, for his continuous support of my research and the preparation of the thesis. His motivation and enthusiasm in regard to research and scholarship continually encourage me to advance on the road of academic. Without his insights and immense knowledge, this work would not have been possible.

I would like to thank my committee members, Professor Hamid Jafarkhani and Professor Zhiying Wang, for their insightful comments and valuable questions that shaped my final thesis.

Last but not the least, I would like to thank my family, for supporting me spiritually throughout writing this thesis and my my life in general.

# ABSTRACT OF THE THESIS

On the Capacity of Weakly-Private Information Retrieval

By

Zhuqing Jia

Master of Science in Electrical Engineering

University of California, Irvine, 2019

Professor Syed Jafar, Chair

The problem of weakly-private information retrieval, is a variant of private information retrieval, where the user wants to retrieve 1 out of $K$ messages from a distributed storage system with $N$ servers that stores all $K$ messages, and is willing to leak some information of the identity of the desired message. In this work, we study the problem of weakly-private information retrieval. A novel information leakage metric is proposed, and the capacity for the setting of $N = 2$ servers, and arbitrary number of messages $K$ is characterized. In particular, in the capacity achieving scheme, designing the distribution of the non-uniformly distributed noise $\mathbf{Z}$ turns out to be the key to achieve the capacity.

# Chapter 1

# Introduction

As we are entering the era of big data, the design of distributed storage systems (DSS) is of growing interest. Motivated by the importance of privacy in DSS, the problem of private information retrieval (PIR) was first proposed in [1]. The problem of PIR is to retrieve a desired message from a database of $K$ messages stored at $N$ distributed servers, without revealing any information about the identity of desired message to any server. The goal is to find the most efficient approach. Initially, the problem of PIR was studied in the computer science perspective, and the cost of PIR was considered as the total amount of communication (upload and download). However, in recent years, the problem of PIR was studied in the information-theoretic perspective, where upload cost are ignored, since it is negligible compared to the download cost. The rate of a PIR scheme is the ratio of the number of bits of the desired message retrieved by the user to the total number of bits downloaded from all servers. The supremum of achievable rates is called the capacity of PIR. This is an active area, and a growing body of literature has produced many capacity results of PIR and its variants.

The problem of PIR was first studied information-theoretically by Sun and Jafar in [2]. The

capacity of PIR was found to be

$$C_{\text{PIR}}(N,K) = \left(1 + 1/N + 1/N^2 + \cdots + 1/N^{K-1}\right). \tag{1.1}$$

This capacity result was then generalized to the $T$-PIR setting, where privacy is guaranteed to any group of up to $T$ colluding servers. In [3], the capacity of $T$-PIR was shown to be

$$C_{\text{TPIR}}(N,K,T) = \begin{cases} \left(1 + T/N + T^2/N^2 + \cdots + T^{K-1}/N^{K-1}\right)^{-1}, & T < N \\ 1/K, & T \geq N. \end{cases} \tag{1.2}$$

Subsequently, the problem of PIR with arbitrary colluding pattern was studied in [4, 5], and the capacity for disjoint colluding sets was found in [5].

In all above-mentioned settings, the storage of DSS is repetition code. Moreover, another interesting setting is PIR with coded storage. The capacity of MDS-PIR, where in each server the MDS-coded version of messages are stored, was proved to be

$$C_{\text{MDS-PIR}} = \left(1 + \left(\frac{M}{N}\right) + \cdots + \left(\frac{M}{N}\right)^{K-1}\right)^{-1} \tag{1.3}$$

in [6], where messages are stored across $N$ servers according to an $(N, M)$ MDS code. However, the natural generalization of MDS-PIR, the problem of MDS-TPIR remains an open problem [7, 8], and an achievable rate was found to be $1 - \frac{M+T-1}{N}$ for any $1 \leq T \leq N - M$.

Another class of PIR variants is PIR with side information. The problem of PIR with private side information (PIR-PSI) was first studied in [9], where the user has $M$ out of $K$ messages available as side-information, and the servers are not aware of the identity of these

$M$ messages. The capacity of such setting is

$$C_{\text{PIR-PSI}}(N,K,M) = \left(1 + 1/N + 1/N^2 + \cdots + 1/N^{K-M-1}\right). \tag{1.4}$$

Following this work, a number of PIR with side-information under variety constraints were studied [10, 11, 12, 13, 14].

Variety models of PIR with security concerns were also studied in an information-theoretical perspective. The problem of symmetric PIR requires the user learns nothing about the messages besides his desired message. In [15, 16, 17], the capacity of symmetric PIR under $T$-private constraint was proved to be

$$C_{\text{TSPIR}}(N,K,T) = 1 - \frac{T}{N}. \tag{1.5}$$

The problem of $T$-PIR with a Byzantine adversary, who may introduce arbitrary errors to the answer strings of any group of up to $B$ servers, was studied in [18], and its capacity was shown to be

$$C_{\text{0-BTPIR}}(N,K,T,B) = \frac{N-2B}{N} C_{\text{TPIR}}(N-2B,K,T) \tag{1.6}$$

under zero-error criteria, and in [19, 20], BTPIR with coded storage were studied. Recently in [21], the capacity of symmetric PIR with a Byzantine adversary under $\epsilon$-error criteria was characterized to be

$$C_{\epsilon-\text{BTSPIR}}(N,K,T,B) = 1 - \frac{T+B}{N}. \tag{1.7}$$

Wiretap models requires data security when communication between user and servers may be listened by eavesdroppers[22, 23]. In [23], the problem of ETPIR was considered, where

the eavesdropper can listen any group of up to $E$ servers. The capacity is shown to be

$$C_{\text{ETPIR}}(N,K,T,E) = \begin{cases} \left(1-\frac{E}{N}\right)\left(1+\frac{T-E}{N-E}+\left(\frac{T-E}{N-E}\right)^2+\cdots+\left(\frac{T-E}{N-E}\right)^{K-1}\right)^{-1}, & E<T \\ 1-\frac{E}{N}, & E\geq T. \end{cases} \tag{1.8}$$

While the most natural generalization of PIR under security constraints is PIR with secure storage, its capacity was characterized until recently in [24]. The problem of $X$-secure $T$-private information retrieval (XSTPIR) is to guarantee data security against collusion among up to $X$ servers and the user's privacy against collusion among up to $T$ servers. In [24], the asymptotic capacity (the capacity in the limit as the number of messages $K \to \infty$) of XSTPIR was proved to be

$$\lim_{K\to\infty} C_{\text{XSTPIR}}(N,K,X,T) = \begin{cases} 1-\left(\frac{X+T}{N}\right), & N>X+T \\ 0, & N\leq X+T. \end{cases} \tag{1.9}$$

Besides, the exact capacity for any $K$ if $X=T=1, N=3$ was also characterized to be

$$C_{\text{XSTPIR}}(N=3,K,X=1,T=1) = \left(\frac{N-X}{N}\right) C_{\text{TPIR}}(N-X,K,T) \tag{1.10}$$

$$= \frac{2}{3}\left(1+\frac{1}{2}+\frac{1}{2^2}+\cdots+\frac{1}{2^{K-1}}\right)^{-1}. \tag{1.11}$$

Moreover, it was also shown that if we relax privacy constraint to $T=1$, we get symmetric security for free, i.e., the user learns nothing beyond its desired message. Finally, the asymptotic capacity of $X$-secure $T$-private computation [25] was also settled.

Recently in [26], another novel PIR model, weakly-private information retrieval (WPIR) was proposed, where the privacy constraint is relaxed such that information-theoretic privacy leakage is allowed. In [26], two different information leakage metrics were studied and two

WPIR schemes were also proposed accordingly. However, no capacity results was given. In this work, we propose a novel information leakage metric for the weakly-private information retrieval problem, and characterize the capacity for the setting of $N = 2$ servers.

*Notation:* For $n_1, n_2 \in \mathbb{Z}$, $n_1 < n_2$, $[n_1 : n_2]$ denotes the set $\{n_1, n_1+1, \ldots, n_2\}$. For an index set $\mathcal{I} = \{i_1, i_2, \ldots, i_n\}$, $X_{\mathcal{I}}$ denotes the set $\{X_{i_1}, X_{i_2}, \ldots, X_{i_n}\}$.

# Chapter 2

# Problem Statement

Consider a DSS with $N$ servers. We have a total of $K$ independent messages, $W_1, W_2, \cdots, W_K$, that are stored in the DSS, and each message consists of $L$ random symbols from the finite field $\mathbb{F}_q$.

$$H(W_1) = H(W_2) = \cdots = H(W_K) = L, \tag{2.1}$$

$$H(W_1, W_2, \ldots, W_K) = KL, \tag{2.2}$$

in $q$-ary units. The information stored at the $n^{th}$ server is denoted by $S_n, n \in [1 : N]$. To ensure information retrieval, note that the set of messages $W_1, \cdots, W_K$ must be a function of $S_{[1:N]}$.

$$H(W_1, \cdots, W_K \mid S_{[1:N]}) = 0. \tag{2.3}$$

The user generates a desired message index $\theta$ privately and uniformly from $[1 : K]$. In order to retrieve $W_\theta$ privately, the user generates $N$ queries, $Q_1^{[\theta]}, Q_2^{[\theta]}, \ldots, Q_N^{[\theta]}$. The query $Q_n^{[\theta]}$ is

sent to the $n^{th}$ server.

For compact notation, let us define

$$\mathcal{Q} = \{\mathcal{Q}_n^{[k]} : k \in [1:K], n \in [1:N]\}. \tag{2.4}$$

The user has no prior knowledge of the information stored at the servers, i.e.,

$$I(S_{[1:N]}; Q_{[1:N]}^{[\theta]}, \theta) = 0. \tag{2.5}$$

Now consider the privacy constraint. In this work, we use an information leakage metric derived from the converse of PIR problem. The $W$-privacy constraint, given the information leakage metric $W$, is defined as

$$\begin{aligned}
WL &= H(A_1^{[2]}|W_1, \mathcal{Q}) - H(A_1^{[1]}|W_1, \mathcal{Q}) && (2.6)\\
&= H(A_1^{[3]}|W_1, W_2, \mathcal{Q}) - H(A_1^{[2]}|W_1, W_2, \mathcal{Q}) && (2.7)\\
&= \dots && (2.8)\\
&= H(A_1^{[K]}|W_1, W_2, \dots, W_{K-1}\mathcal{Q}) - H(A_1^{[K-1]}|W_1, W_2, \dots, W_{K-1}, \mathcal{Q}) && (2.9)\\
&= \dots && (2.10)\\
&= H(A_N^{[2]}|W_1, \mathcal{Q}) - H(A_N^{[1]}|W_1, \mathcal{Q}) && (2.11)\\
&= H(A_N^{[3]}|W_1, W_2, \mathcal{Q}) - H(A_N^{[2]}|W_1, W_2, \mathcal{Q}) && (2.12)\\
&= \dots && (2.13)\\
&= H(A_N^{[K]}|W_1, W_2, \dots, W_{K-1}\mathcal{Q}) - H(A_N^{[K-1]}|W_1, W_2, \dots, W_{K-1}, \mathcal{Q}) && (2.14)
\end{aligned}$$

Here we note that $W \in \mathbb{R}, 0 \leq W \leq \frac{1}{N}$. The lower bound of $W$ is easy to see, since $W = 0$ indicates the perfect privacy. Now let us explain the upper bound of $W$. Consider an information retrieval scheme with no privacy constraint completely, where the user randomly,

privately choose 1 out of $N$ servers, and retrieve the desired message from the chosen server. It is easy to see that for the given scheme, $W = \frac{1}{N}$, and the rate achieved is trivially 1. Since the information retrieval rate can not exceed 1, $\frac{1}{N}$ is the upperbound of $W$.

Upon receiving the query $Q_n^{[\theta]}$, the $n^{th}$ server generates an answering string $A_n^{[\theta]}$, as a function of the query $Q_n^{[\theta]}$ and its stored information $S_n$.

$$H(A_n^{[\theta]}|Q_n^{[\theta]}, S_n) = 0. \tag{2.15}$$

From all the answers the user must be able to recover the desired message $W_\theta$,

$$\text{[Correctness]} \qquad H(W_\theta|A_{[1:N]}^{[\theta]}, Q_{[1:N]}^{[\theta]}, \theta) = 0. \tag{2.16}$$

The rate of an WPIR scheme characterizes how many bits of desired message are retrieved per downloaded bit, (equivalently, how many $q$-ary symbols of desired message are retrieved per downloaded $q$-ary symbol),

$$R = \frac{L}{D}, \tag{2.17}$$

where $D$ is the expected value (with respect to the random queries) of the number of $q$-ary symbols downloaded by the user from all servers. The capacity of WPIR, denoted $C$, is the supremum of achievable rates over all $L$.

# Chapter 3

# Main Result

For the sake of simplicity, let us define

$$S(K) = \frac{1 - \frac{1}{2^K}}{1 - \frac{1}{2}}. \tag{3.1}$$

The main result of this work, is the characterization of the capacity of WPIR problem with $N = 2$ servers and arbitrary number of messages $K$ under the information leakage metric defined in (2.14), which is presented in the following theorem.

**Theorem 3.1.** *The capacity of WPIR with $N = 2$ servers and arbitrary number of messages $K$ under the information leakage metric defined in (2.14) is*

$$C = \frac{1}{S(K) - S(K-1)W}. \tag{3.2}$$

The capacity of Weakly-private information retrieval (WPIR), as a function of the information leakage metric $W$ defined in (2.14), with $N = 2$ servers and different number of messages $K$, is presented in the following figure.

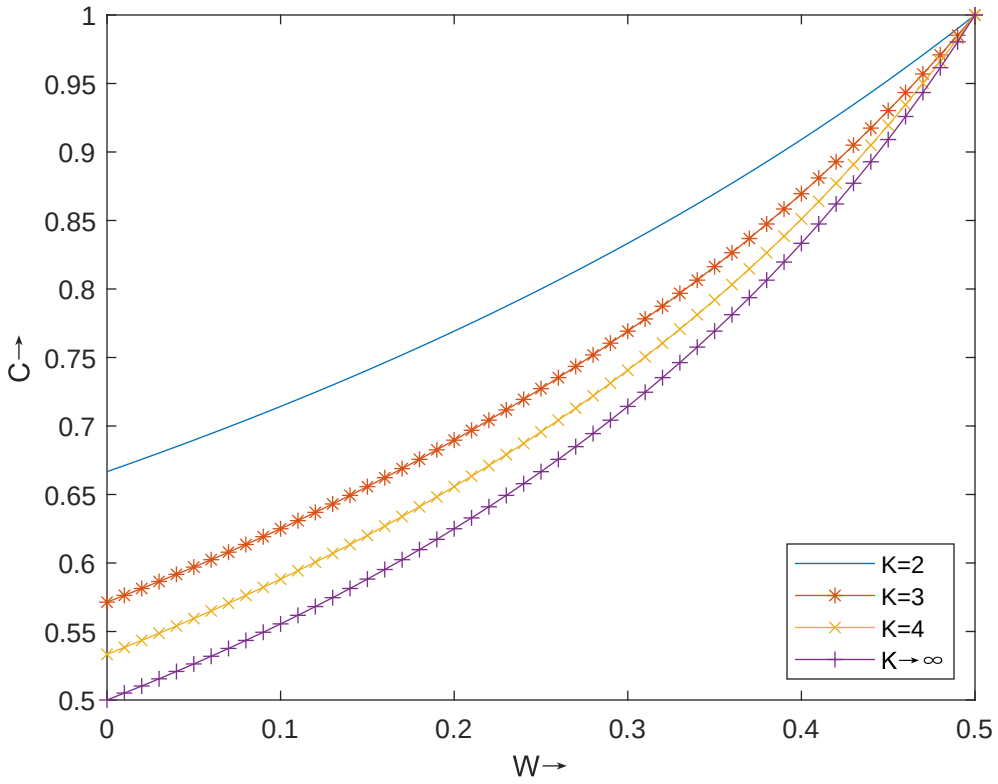


Figure 3.1: The capacity of WPIR as a function of proposed $W$ with $N = 2$ servers and different $K$.

The proof of achievability is presented in Chapter 4, while the proof of converse is presented in Chapter 5.

# Chapter 4

# Achievability

Throughout the scheme, we will assume that each message consist of $L = 1$ symbol from $\mathbb{F}_2$, i.e., one binary symbol. Besides, let us define $1 \times K$ vector

$$\mathbf{W} = [W_1, W_2, \ldots, W_K]. \tag{4.1}$$

Also, we define $K \times 1$ vector $\mathbf{Q}^{[\theta]}$ to be the $\theta$-th column of a $K \times K$ identity matrix. $\mathbf{Z}$ is a $K \times 1$ random vector from all vectors over $\mathbb{F}_2^K$, according to some distribution. A WPIR scheme, for $N = 2$, presented in terms of storage of each server, query to each server and answer symbol downloaded from each server, is as follows

| | DB1 | DB2 |
|---|---|---|
| Storage | $\mathbf{W}$ | $\mathbf{W}$ |
| Query | $\mathbf{Z}$ | $\mathbf{Q}^{[\theta]} + \mathbf{Z}$ |
| Answer | $\mathbf{WZ}$ | $\mathbf{WQ}^{[\theta]} + \mathbf{WZ}$ |

Table 4.1: Construction of storage, query and answer of $N = 2$ servers.

Correctness of the scheme is proved as follows. Upon receiving $N = 2$ answer symbols, the

user finds summation of the two symbols.

$$\mathbf{WZ} + \mathbf{WQ}^{[\theta]} + \mathbf{WZ} \tag{4.2}$$

$$= \mathbf{WQ}^{[\theta]}. \tag{4.3}$$

The interference term $\mathbf{WZ}$ cancels out because all symbols are binary. Due to the definition of $\mathbf{W}$ and $\mathbf{Q}^{[\theta]}$, we have

$$\mathbf{WQ}^{[\theta]} = W_\theta \tag{4.4}$$

is the desired message symbol, this completes the proof of correctness.

One may notice that the construction of the scheme resembles that in [1], however, we highlight that this scheme achieves the PIR capacity when $\mathbf{Z}$ is a uniformly distributed random vector from all vectors over $\mathbb{F}_2^K$ and also independent from $\mathbf{Q}^{[\theta]}$, in other words, we focus on $W = 0$ for now. First, let us see why perfect privacy is guaranteed. For the first database, we have

$$I(\mathbf{Z}; \mathbf{Q}^{[\theta]}) = 0. \tag{4.5}$$

Similarly, for the second database, we have

$$I(\mathbf{Q}^{[\theta]} + \mathbf{Z}; \mathbf{Q}^{[\theta]}) \tag{4.6}$$

$$= I(\mathbf{Z}; \mathbf{Q}^{[\theta]}) = 0 \tag{4.7}$$

This guarantees the perfect privacy. Now let us consider the rate that the scheme achieves. Since $\mathbf{Z}$ and $\mathbf{Q}^{[\theta]} + \mathbf{Z}$ are uniformly distributed random vectors from all vectors over $\mathbb{F}_2^K$, with probability $\frac{1}{2^K}$, the query vector is the all zero vector, thus no download is needed whenever

this happens. Therefore, we can calculate that the rate achieved is

$$R = \frac{1}{1-\frac{1}{2^K}}\left(1-\frac{1}{2}\right) \tag{4.8}$$

$$= \frac{1-\frac{1}{2}}{1-\frac{1}{2^K}} \tag{4.9}$$

$$= \left(1+\frac{1}{2}+\frac{1}{4}+\cdots+\frac{1}{2^{K-1}}\right)^{-1} \tag{4.10}$$

which matches the capacity of this setting ($W = 0$).

From the construction of the scheme, we can see the role of the field size plays in the achieved rate. With higher probability that the query vector is the all zero vector, the average download is reduced. Therefore, the intuition behind the construction of WPIR scheme is to increase the probability that the query vector is the all zero vector, by carefully designing the distribution of $\mathbf{Z}$. Let us start from the setting where $K = 2$.

## 4.1 K=2 Messages

As we stated before, the key insight to design the capacity achieving WPIR scheme is to carefully design the distribution of $\mathbf{Z}$. When $K = 2$, the distribution of $\mathbf{Z}$, conditioning on desired message index $\theta$, is shown in the following table.

| $\mathbf{Z}$ | $\theta = 1$ | $\theta = 2$ |
|---|---|---|
| $[0,0]^T$ | $\frac{1}{4}+\frac{1}{2}W$ | $\frac{1}{4}+\frac{1}{2}W$ |
| $[0,1]^T$ | $\frac{1}{4}-\frac{1}{2}W$ | $\frac{1}{4}+\frac{1}{2}W$ |
| $[1,0]^T$ | $\frac{1}{4}+\frac{1}{2}W$ | $\frac{1}{4}-\frac{1}{2}W$ |
| $[1,1]^T$ | $\frac{1}{4}-\frac{1}{2}W$ | $\frac{1}{4}-\frac{1}{2}W$ |

Table 4.2: Distribution of $\mathbf{Z}$, for $N = 2, K = 2$ setting.

where $W$ is the information leakage metric, $W \in \mathbb{R}, 0 \leq W \leq \frac{1}{2}$. Recall that $W$ is defined as

$$WL = H(A_1^{[2]}|W_1, \mathcal{Q}) - H(A_1^{[1]}|W_1, \mathcal{Q}) \tag{4.11}$$

$$= H(A_2^{[2]}|W_1, \mathcal{Q}) - H(A_2^{[1]}|W_1, \mathcal{Q}). \tag{4.12}$$

This is a valid distribution, which can be easily verified. We also note that by symmetry of the distribution, $\mathbf{Z}$ and $\mathbf{Q}^{[\theta]} + \mathbf{Z}$ have the same distribution conditioning on $\theta$. Let us see why $W$-privacy is achieved. In order to calculate $W$, consider the distribution of $A_1^{[1]}, A_2^{[1]}, A_1^{[2]}$ and $A_2^{[2]}$.

| $P$ | $A_1^{[1]}$ | $A_2^{[1]}$ | $A_1^{[2]}$ | $A_2^{[2]}$ |
|---|---|---|---|---|
| $\frac{1}{4} + \frac{1}{2}W$ | $0, W_1$ | $0, W_1$ | $0, W_2$ | $0, W_2$ |
| $\frac{1}{4} - \frac{1}{2}W$ | $W_2, W_1 + W_2$ | $W_2, W_1 + W_2$ | $W_1, W_1 + W_2$ | $W_1, W_1 + W_2$ |

Table 4.3: Distribution of $A_1^{[1]}, A_2^{[1]}, A_1^{[2]}$ and $A_2^{[2]}$, for $N = 2, K = 2$ setting.

Therefore,

$$H(A_1^{[1]}|W_1, \mathcal{Q}) = 0 \times \left(\frac{1}{2} + W\right) + H(W_2)\left(\frac{1}{2} - W\right) \tag{4.13}$$

$$= L\left(\frac{1}{2} - W\right). \tag{4.14}$$

Similarly,

$$H(A_1^{[2]}|W_1, \mathcal{Q}) = \frac{1}{2}L. \tag{4.15}$$

Thus we obtain

$$H(A_1^{[2]}|W_1, \mathcal{Q}) - H(A_1^{[1]}|W_1, \mathcal{Q}) = WL. \tag{4.16}$$

$H(A_2^{[2]}|W_1, \mathcal{Q}) - H(A_2^{[1]}|W_1, \mathcal{Q}) = WL$ can be also similarly verified. This guarantees the

$W$-privacy. Now let us consider the rate achieved by the scheme. When $\theta = 1$, $\mathbf{Z} = [0, 0]^T$ or $\mathbf{Z} = [1, 0]^T$, and when $\theta = 2$, $\mathbf{Z} = [0, 0]^T$ or $\mathbf{Z} = [0, 1]^T$ the user only need to download 1 symbol from $N = 2$ servers. Since $\theta$ is uniformly distributed from $[1 : K]$, the average download $\bar{D}$ is

$$\bar{D} = \frac{1}{2} \times 4 \times L \left( \frac{1}{4} + \frac{1}{2} W \right) + \frac{1}{2} \times 2 \times 4 \times L \left( \frac{1}{4} - \frac{1}{2} W \right) \tag{4.17}$$

$$= L \left( \frac{3}{2} - W \right). \tag{4.18}$$

Thus the rate achieved is

$$R = \frac{L}{\bar{D}} = \frac{1}{\frac{3}{2} - W}. \tag{4.19}$$

which matches the capacity for this setting.

*Discussion:* As a sanity check, let us consider corner cases, i.e., when $W = 0$ and $W = \frac{1}{2}$. When $W = 0$, no information leakage is allowed, therefore, perfect privacy is guaranteed. It is easy to see that when $W = 0$, $\mathbf{Z}$ is a uniformly distributed random vector from all vectors over $\mathbb{F}_2^2$ and also independent from $\mathbf{Q}^{[\theta]}$, thus guarantees perfect privacy. The rate achieved is $R = 2/3$, which matches the capacity of this setting. On the other hand, when $W = \frac{1}{2}$, i.e., no privacy constraint at all, the scheme is equivalent to randomly select one server from $N = 2$ servers, then download the desired symbol directly from that server. And the achieved rate is trivially 1.

## 4.2 K=3 Messages

When $K = 3$, the distribution of $\mathbf{Z}$, conditioning on desired message index $\theta$, is shown in the following table.

| $\mathbf{Z}$ | $\theta = 1$ | $\theta = 2$ | $\theta = 3$ |
|---|---|---|---|
| $[0,0,0]^T$ | $\frac{1}{8} + \frac{3}{4}W$ | $\frac{1}{8} + \frac{3}{4}W$ | $\frac{1}{8} + \frac{3}{4}W$ |
| $[0,0,1]^T$ | $\frac{1}{8} - \frac{1}{4}W$ | $\frac{1}{8} - \frac{1}{4}W$ | $\frac{1}{8} + \frac{3}{4}W$ |
| $[0,1,0]^T$ | $\frac{1}{8} - \frac{1}{4}W$ | $\frac{1}{8} + \frac{3}{4}W$ | $\frac{1}{8} - \frac{1}{4}W$ |
| $[0,1,1]^T$ | $\frac{1}{8} - \frac{1}{4}W$ | $\frac{1}{8} - \frac{1}{4}W$ | $\frac{1}{8} - \frac{1}{4}W$ |
| $[1,0,0]^T$ | $\frac{1}{8} + \frac{3}{4}W$ | $\frac{1}{8} - \frac{1}{4}W$ | $\frac{1}{8} - \frac{1}{4}W$ |
| $[1,0,1]^T$ | $\frac{1}{8} - \frac{1}{4}W$ | $\frac{1}{8} - \frac{1}{4}W$ | $\frac{1}{8} - \frac{1}{4}W$ |
| $[1,1,0]^T$ | $\frac{1}{8} - \frac{1}{4}W$ | $\frac{1}{8} - \frac{1}{4}W$ | $\frac{1}{8} - \frac{1}{4}W$ |
| $[1,1,1]^T$ | $\frac{1}{8} - \frac{1}{4}W$ | $\frac{1}{8} - \frac{1}{4}W$ | $\frac{1}{8} - \frac{1}{4}W$ |

Table 4.4: Distribution of $\mathbf{Z}$, for $N = 2, K = 3$ setting.

where $W$ is the information leakage metric, $W \in \mathbb{R}, 0 \leq W \leq \frac{1}{2}$. In this setting, $W$ is defined as

$$WL = H(A_1^{[2]}|W_1, \mathcal{Q}) - H(A_1^{[1]}|W_1, \mathcal{Q}) \tag{4.20}$$

$$= H(A_1^{[3]}|W_1, W_2, \mathcal{Q}) - H(A_1^{[2]}|W_1, W_2, \mathcal{Q}) \tag{4.21}$$

$$= H(A_2^{[2]}|W_1, \mathcal{Q}) - H(A_2^{[1]}|W_1, \mathcal{Q}) \tag{4.22}$$

$$= H(A_2^{[3]}|W_1, W_2, \mathcal{Q}) - H(A_2^{[2]}|W_1, W_2, \mathcal{Q}). \tag{4.23}$$

Considering the distribution of answer symbols, and due to symmetry of the distribution, we obtain

$$H(A_1^{[2]}|W_1, \mathcal{Q}) - H(A_1^{[1]}|W_1, \mathcal{Q}) \tag{4.24}$$

$$= L\left(\frac{1}{8} + \frac{3}{4}W - \frac{1}{8} + \frac{1}{4}W\right) \tag{4.25}$$

$$= WL. \tag{4.26}$$

Similarly,

$$H(A_1^{[2]}|W_1,\mathcal{Q}) - H(A_1^{[1]}|W_1,\mathcal{Q}) \tag{4.27}$$
$$= H(A_1^{[3]}|W_1,W_2,\mathcal{Q}) - H(A_1^{[3]}|W_1,W_2,\mathcal{Q}) \tag{4.28}$$
$$= H(A_2^{[2]}|W_1,\mathcal{Q}) - H(A_2^{[1]}|W_1,\mathcal{Q}) \tag{4.29}$$
$$= H(A_2^{[3]}|W_1,W_2,\mathcal{Q}) - H(A_2^{[3]}|W_1,W_2,\mathcal{Q}) \tag{4.30}$$
$$= WL. \tag{4.31}$$

Thus the $W$-privacy is guaranteed. Now let us calculate the average download of the scheme.

$$\bar{D} = \frac{1}{3} \times 6L \times \left(\frac{1}{8} + \frac{3}{4}W\right) + \frac{1}{3} \times 36L \times \left(\frac{1}{8} - \frac{1}{4}W\right) \tag{4.32}$$
$$= L\left(\frac{7}{4} - \frac{3}{2}W\right) \tag{4.33}$$

The rate achieved is

$$R = \frac{L}{\bar{D}} = \frac{1}{\frac{7}{4} - \frac{3}{2}W} \tag{4.34}$$

which matches the capacity of this setting.

## 4.3 Arbitrary Number of Messages $K$

For arbitrary number of messages $K$, the distribution of $\mathbf{Z}$ conditioning on $\theta$ is as follows

$$P(\mathbf{Z}|\theta) = \begin{cases} 1 - \frac{1}{2}(S(K) - S(K-1)W), & \mathbf{Z} = \mathbf{0} \\ 1 - \frac{1}{2}(S(K) - S(K-1)W), & \mathbf{Z} = \mathbf{Q}^{[\theta]} \\ \frac{1}{2^K - 2}(S(K) - S(K-1)W - 1), & otherwise. \end{cases} \tag{4.35}$$

To verify the distribution is valid, we note that for each realization of $\theta$, among $2^K$ possible realizations of $\mathbf{Z}$, we have

$$2 \times \left(1 - \frac{1}{2}(S(K) - S(K-1)W)\right) + (2^K - 2)\frac{1}{2^K - 2}(S(K) - S(K-1)W - 1) \tag{4.36}$$

$$= 2 - (S(K) - S(K-1)W) + (S(K) - S(K-1))W - 1 \tag{4.37}$$

$$= 1. \tag{4.38}$$

Thus this distribution is valid. Now let us see why $W$-privacy holds. For arbitrary $K$, the information leakage metric $W$ is defined as

$$WL = H(A_1^{[2]}|W_1, \mathcal{Q}) - H(A_1^{[1]}|W_1, \mathcal{Q}) \tag{4.39}$$

$$= H(A_1^{[3]}|W_1, W_2, \mathcal{Q}) - H(A_1^{[2]}|W_1, W_2, \mathcal{Q}) \tag{4.40}$$

$$= \ldots \tag{4.41}$$

$$= H(A_1^{[K]}|W_1, W_2, \ldots, W_{K-1}\mathcal{Q}) - H(A_1^{[K-1]}|W_1, W_2, \ldots, W_{K-1}, \mathcal{Q}) \tag{4.42}$$

$$= H(A_2^{[2]}|W_1, \mathcal{Q}) - H(A_2^{[1]}|W_1, \mathcal{Q}) \tag{4.43}$$

$$= H(A_2^{[3]}|W_1, W_2, \mathcal{Q}) - H(A_2^{[2]}|W_1, W_2, \mathcal{Q}) \tag{4.44}$$

$$= \ldots \tag{4.45}$$

$$= H(A_2^{[K]}|W_1, W_2, \ldots, W_{K-1}\mathcal{Q}) - H(A_2^{[K-1]}|W_1, W_2, \ldots, W_{K-1}, \mathcal{Q}) \tag{4.46}$$

Due to symmetry of the distribution, we have

$$
\begin{aligned}
& H(A_1^{[2]}|W_1, \mathcal{Q}) - H(A_1^{[1]}|W_1, \mathcal{Q}) && (4.47)\\
&= H(A_1^{[3]}|W_1, W_2, \mathcal{Q}) - H(A_1^{[2]}|W_1, W_2, \mathcal{Q}) && (4.48)\\
&= \dots && (4.49)\\
&= H(A_1^{[K]}|W_1, W_2, \dots, W_{K-1}\mathcal{Q}) - H(A_1^{[K-1]}|W_1, W_2, \dots, W_{K-1}, \mathcal{Q}) && (4.50)\\
&= H(A_2^{[2]}|W_1, \mathcal{Q}) - H(A_2^{[1]}|W_1, \mathcal{Q}) && (4.51)\\
&= H(A_2^{[3]}|W_1, W_2, \mathcal{Q}) - H(A_2^{[2]}|W_1, W_2, \mathcal{Q}) && (4.52)\\
&= \dots && (4.53)\\
&= H(A_2^{[K]}|W_1, W_2, \dots, W_{K-1}\mathcal{Q}) - H(A_2^{[K-1]}|W_1, W_2, \dots, W_{K-1}, \mathcal{Q}) && (4.54)\\
&= L\left(1 - \frac{1}{2}(S(K) - S(K-1)W) - \frac{1}{2^K - 2}(S(K) - S(K-1)W - 1)\right) && (4.55)\\
&= L\left(\frac{2^K - 1}{2^K - 2} + \frac{2^{K-1}}{2^K - 2}S(K-1)W - \frac{2^{K-1}}{2^K - 2}S(K)\right) && (4.56)\\
&= L\left(\frac{2^K - 1}{2^K - 2} + \frac{2^{K-1}}{2^K - 2}S(K-1)W - \frac{2^K - 1}{2^K - 2}\right) && (4.57)\\
&= L\left(\frac{2^{K-1}}{2^K - 2}S(K-1)W\right) && (4.58)\\
&= WL. && (4.59)
\end{aligned}
$$

Equation (4.54) is justified as follows. Since $A_n^{[\theta]}$ is a linear combination of $W_{[1:K]}$, given a subset of $W_{[1:K]}$, $H(A_n^{[\theta]})$ equals either 0 or $L$. Therefore,

$$H(A_n^{[\theta]}|W_{\mathcal{K}}, \mathcal{Q}) = (1 - P(A_n^{[\theta]} = 0|W_{\mathcal{K}}))L. \tag{4.60}$$

From the construction of distribution of $\mathbf{Z}$, the following equation always holds.

$$
\begin{aligned}
& P(A_n^{[\theta]} = 0|W_{\mathcal{K}}) - P(A_n^{[\theta']} = 0|W_{\mathcal{K}}) && (4.61)\\
&= 1 - \frac{1}{2}(S(K) - S(K-1)W) - \frac{1}{2^K - 2}(S(K) - S(K-1)W - 1. && (4.62)
\end{aligned}
$$

where $\theta \in \mathcal{K}, \theta' \notin \mathcal{K}$. Thus the equation (4.54) holds. This completes the proof of $W$-privacy. Finally, let us calculate the average download of the scheme.

$$\bar{D} = 2 \times \left(1 - \frac{1}{2}(S(K) - S(K-1)W)\right) \tag{4.63}$$

$$+ 2(2^K - 2)\left(\frac{1}{2^K - 2}(S(K) - S(K-1)W - 1)\right)$$

$$= S(K) - S(K-1)W. \tag{4.64}$$

which matches the capacity of this setting. This completes the construction of capacity achieving WPIR scheme for $N = 2$, arbitrary number of messages $K$ setting.

*Discussion:* The construction of the capacity achieving scheme exploits the fact that when the query is the all zero vector, the user don't have to download anything from the server, thus the rate is increased. This is done by designing the distribution of $\mathbf{Z}$ and increase the probability that the query is the all zero vector. Besides, the distribution is symmetric, such that the query vectors for two servers, $\mathbf{Z}$ and $\mathbf{Q}^{[\theta]} + \mathbf{Z}$ are identically distributed. Consequently, under the information leakage metric used in this work, for the setting of $N = 2$ and arbitrary number of messages $K$, the capacity is achieved. However, we note that the capacity of WPIR problem with the same setting under other information leakage metrics remains open.

# Chapter 5

# Converse

In this chapter, we prove the converse of Theorem 3.1. Let us start from the setting where $N = 2$, $K = 2$.

## 5.1 $K = 2$ Messages

The download satisfies

$$\begin{aligned} D &\geq H(A_1^{[1]}, A_2^{[1]}|\mathcal{Q}) && (5.1)\\ &= H(A_1^{[1]}, A_2^{[1]}, W_1|\mathcal{Q}) && (5.2)\\ &= H(W_1|\mathcal{Q}) + H(A_1^{[1]}, A_2^{[1]}|W_1, \mathcal{Q}) && (5.3)\\ &= L + H(A_1^{[1]}, A_2^{[1]}|W_1, \mathcal{Q}). && (5.4) \end{aligned}$$

(5.2) follows from the correctness constraint, i.e., from $A_1^{[1]}, A_2^{[1]}$ we can decode $W_1$. Therefore, we have

$$D \geq L + H(A_1^{[1]}|W_1, \mathcal{Q}) \tag{5.5}$$
$$= L + H(A_1^{[2]}|W_1, \mathcal{Q}) - WL. \tag{5.6}$$

(5.6) follows from the definition of $W$-privacy. Similarly, we have

$$D \geq L + H(A_2^{[1]}|W_1, \mathcal{Q}) \tag{5.7}$$
$$= L + H(A_2^{[2]}|W_1, \mathcal{Q}) - WL. \tag{5.8}$$

Averaging the two inequalities, we have

$$D \geq L + \frac{1}{2}H(A_1^{[2]}|W_1, \mathcal{Q}) + \frac{1}{2}H(A_2^{[1]}|W_1, \mathcal{Q}) - WL \tag{5.9}$$
$$\geq L + \frac{1}{2}H(A_1^{[2]}, A_2^{[2]}|W_1, \mathcal{Q}) - WL \tag{5.10}$$
$$= L + \frac{1}{2}H(A_1^{[2]}, A_2^{[2]}, W_2|W_1, \mathcal{Q}) - WL \tag{5.11}$$
$$\geq L\left(\frac{3}{2} - W\right). \tag{5.12}$$

In (5.11) we used the fact that from $A_1^{[2]}, A_2^{[2]}$ we can decode $W_2$. Thus we have

$$R \leq \frac{L}{D} = \frac{1}{\frac{3}{2} - W}. \tag{5.13}$$

This completes converse proof for this setting.

## 5.2 $K = 3$ Messages

From (5.11), we have

$$D \geq L + \frac{1}{2}H(A_1^{[2]}, A_2^{[2]}, W_2|W_1, \mathcal{Q}) - WL \tag{5.14}$$
$$= L + \frac{1}{2}(H(W_2|W_1, \mathcal{Q}) + H(A_1^{[2]}, A_2^{[2]}|W_1, W_2, \mathcal{Q})) - WL \tag{5.15}$$
$$= \frac{3}{2}L + \frac{1}{2}H(A_1^{[2]}, A_2^{[2]}|W_1, W_2, \mathcal{Q}) - WL \tag{5.16}$$

Therefore, we have

$$D \geq \frac{3}{2}L + \frac{1}{2}H(A_1^{[2]}|W_1, W, \mathcal{Q}) - WL \tag{5.17}$$
$$= \frac{3}{2}L + \frac{1}{2}H(A_1^{[3]}|W_1, W_2, \mathcal{Q}) - \frac{3}{2}WL \tag{5.18}$$

where in (5.18), we used the definition of $W$-privacy. Similarly, we have

$$D \geq \frac{3}{2}L + \frac{1}{2}H(A_2^{[2]}|W_1, W_2, \mathcal{Q}) - WL \tag{5.19}$$
$$= \frac{3}{2}L + \frac{1}{2}H(A_2^{[3]}|W_1, W_2, \mathcal{Q}) - \frac{3}{2}WL \tag{5.20}$$

Averaging the two inequalities, we have

$$D \geq \frac{3}{2}L + \frac{1}{4}H(A_1^{[3]}|W_1, W_2, \mathcal{Q}) + \frac{1}{4}H(A_2^{[3]}|W_1, W_2, \mathcal{Q}) - \frac{3}{2}L \tag{5.21}$$
$$\geq \frac{3}{2}L + \frac{1}{4}H(A_1^{[3]}, A_2^{[3]}|W_1, W_2, \mathcal{Q}) - \frac{3}{2}L \tag{5.22}$$
$$= \frac{3}{2}L + \frac{1}{4}H(A_1^{[3]}, A_2^{[3]}, W_3|W_1, W_2, \mathcal{Q}) - \frac{3}{2}L \tag{5.23}$$
$$\geq L\left(\frac{7}{4} + \frac{3}{2}W\right). \tag{5.24}$$

(5.23) follows from the decodability of message $W_3$. Thus we have

$$R \leq \frac{L}{D} = \frac{1}{\frac{7}{4} - \frac{3}{2}W}. \tag{5.25}$$

This completes converse proof for this setting.

## 5.3 Arbitrary Number of Messages $K$

The proof of converse for arbitrary number of messages $K$ is based on an induction argument. We repeatedly use the correctness constraint and the definition of $W$-privacy. Starting from (5.23), we have

$$\begin{aligned} D &\geq \frac{3}{2}L + \frac{1}{4}H(A_1^{[3]}, A_2^{[3]}, W_3|W_1, W_2, \mathcal{Q}) - \frac{3}{2}WL & (5.26)\\ &= \frac{3}{2}L + \frac{1}{4}H(W_3|W_1, W_2, \mathcal{Q}) + H(A_1^{[3]}, A_2^{[3]}|W_1, W_2, W_3, \mathcal{Q}) - \frac{3}{2}WL & (5.27)\\ &\geq \frac{7}{4}L + \frac{1}{4}H(A_1^{[3]}, A_2^{[3]}|W_1, W_2, W_3, \mathcal{Q}) - \frac{3}{2}WL. & (5.28) \end{aligned}$$

Thus we have

$$\begin{aligned} D &\geq \frac{7}{4}L + \frac{1}{4}H(A_1^{[3]}|W_1, W_2, W_3, \mathcal{Q}) - \frac{3}{2}WL & (5.29)\\ &= \frac{7}{4}L + \frac{1}{4}H(A_1^{[4]}|W_1, W_2, W_3, \mathcal{Q}) - \frac{7}{4}WL. & (5.30) \end{aligned}$$

Similarly, we have

$$D \geq \frac{7}{4}L + \frac{1}{4}H(A_2^{[4]}|W_1, W_2, W_3, \mathcal{Q}) - \frac{7}{4}WL. \tag{5.31}$$

Averaging the two inequalities, we have

$$D \geq \frac{7}{4}L + \frac{1}{8}H(A_1^{[4]}|W_1, W_2, W_3, \mathcal{Q}) + \frac{1}{8}H(A_2^{[4]}|W_1, W_2, W_3, \mathcal{Q}) - \frac{7}{4}WL \tag{5.32}$$

$$\geq \frac{7}{4}L + \frac{1}{8}H(A_1^{[4]}, A_2^{[4]}|W_1, W_2, W_3, \mathcal{Q}) - \frac{7}{4}WL \tag{5.33}$$

Now we continue from (5.33), and repeat the procedure $(K-4)$ times, we have

$$D \geq S(K-1)L + \frac{1}{2^{K-1}}H(A_1^{[K]}, A_2^{[K]}|W_1, W_2, W_3, \ldots, W_{K-1}\mathcal{Q}) - S(K-1)WL \tag{5.34}$$

$$= S(K-1)L + \frac{1}{2^{K-1}}H(A_1^{[K]}, A_2^{[K]}, W_K|W_1, W_2, W_3, \ldots, W_{K-1}\mathcal{Q}) - S(K-1)WL \tag{5.35}$$

$$\geq S(K-1)L + \frac{1}{2^{K-1}}L - S(K-1)WL \tag{5.36}$$

$$= L(S(K) - S(K-1)W). \tag{5.37}$$

Thus the outer bound of the rate is

$$R \leq \frac{1}{S(K) - S(K-1)W}. \tag{5.38}$$

This completes the proof of converse for $N = 2$ and arbitrary $K$ setting.

# Chapter 6

# Discussion

The capacity of WPIR with $N = 2$ servers and arbitrary number of messages $K$ under a novel information leakage metric is characterized in this work. The following observations are noticeable. First, for a fixed number of messages $K$, the lower bound of expected download $D$, is a linear function of $W$. Under the asymptotic setting, i.e., $K \to \infty$, we have

$$\lim_{k\to\infty} C = \frac{1}{2-2W}. \tag{6.1}$$

Thus with $W$-privacy, the rate increases, even the number of messages $K$ is large. Besides, the capacity achieving scheme, exploits the fact that all symbols are from $\mathbb{F}_2$. Symmetric distributions of the noise vector $\mathbf{Z}$, that is used to protect the query vector, are designed according to different $K$. However, the generalization of the scheme to arbitrary number of servers $N$ is not straightforward. For example, consider the setting where $N = 3$. To achieve the asymptotic capacity, for each message, we need $L = 2$ symbols from $\mathbb{F}_2$. Therefore, for arbitrary $K$, the scheme achieves the rate

$$R = \frac{1}{1-\frac{1}{2^{2K}}}\left(1-\frac{1}{3}\right). \tag{6.2}$$

However, the capacity for this setting [2] is

$$C = \frac{1 - \frac{1}{3}}{1 - \frac{1}{3^K}}. \tag{6.3}$$

Thus the rate achieved does not match the capacity. Therefore, the capacity of WPIR for arbitrary $N, K$ under the information leakage metric defined in (2.14) merit investigation for the future work. Besides, in practical, other information leakage metrics may also be considered, such as the worst-case information leakage [26]. The capacity of WPIR under such information leakage metrics, even for $N = 2, K = 2$ and asymptotic settings, remains open. The relation between $W$-privacy and upload cost, computation cost, storage overhead, are other interesting problems that deserve study.

# Bibliography


[1] Benny Chor, Eyal Kushilevitz, Oded Goldreich, and Madhu Sudan. Private Information Retrieval. *Journal of the ACM (JACM)*, 45(6):965–981, 1998.

[2] Hua Sun and Syed Ali Jafar. The Capacity of Private Information Retrieval. *IEEE Transactions on Information Theory*, 63(7):4075–4088, July 2017.

[3] Hua Sun and Syed Ali Jafar. The Capacity of Robust Private Information Retrieval with Colluding Databases. *IEEE Transactions on Information Theory*, 64(4):2361–2370, April 2018.

[4] Razane Tajeddine, Oliver W Gnilke, David Karpuk, Ragnar Freij-Hollanti, Camilla Hollanti, and Salim El Rouayheb. Private information retrieval schemes for codec data with arbitrary collusion patterns. *IEEE International Symposium on Information Theory (ISIT)*, pages 1908–1912, 2017.

[5] Zhuqing Jia, Hua Sun, and Syed Jafar. The capacity of private information retrieval with disjoint colluding sets. In *IEEE GLOBECOM*, 2017.

[6] Karim Banawan and Sennur Ulukus. The Capacity of Private Information Retrieval from Coded Databases. *IEEE Transactions on Information Theory*, 64(3):1945–1956, 2018.

[7] Ragnar Freij-Hollanti, Oliver Gnilke, Camilla Hollanti, and David Karpuk. Private Information Retrieval from Coded Databases with Colluding Servers. *SIAM Journal on Applied Algebra and Geometry*, 1(1):647–664, 2017.

[8] Hua Sun and Syed 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.

[9] Zhen Chen, Zhiying Wang, and Syed Jafar. The capacity of private information retrieval with private side information. *arXiv preprint arXiv:1709.03022*, 2017.

[10] Ravi Tandon. The capacity of cache aided private information retrieval. *arXiv preprint arXiv:1706.07035*, 2017.

[11] Yi-Peng Wei, Karim Banawan, and Sennur Ulukus. Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching. *arXiv preprint arXiv:1709.01056*, 2017.

[12] Yi-Peng Wei, Karim Banawan, and Sennur Ulukus. The capacity of private information retrieval with partially known private side information. *arXiv preprint arXiv:1710.00809*, 2017.

[13] Seyed Pooya Shariatpanahi, Mahdi Jafari Siavoshani, and Mohammad Ali Maddah-Ali. Multi-message private information retrieval with private side information. *arXiv preprint arXiv:1805.11892*, 2018.

[14] Su Li and Michael Gastpar. Single-server multi-message private information retrieval with side information. *arXiv preprint arXiv:1808.05797*, 2018.

[15] Hua Sun and Syed A. Jafar. The capacity of symmetric private information retrieval. *IEEE Transactions on Information Theory*, 2018.

[16] Qiwen Wang and Mikael Skoglund. Linear symmetric private information retrieval for MDS coded distributed storage with colluding servers. *arXiv preprint arXiv:1708.05673*, 2017.

[17] Qiwen Wang and Mikael Skoglund. Secure symmetric private information retrieval from colluding databases with adversaries. *arXiv preprint arXiv:1707.02152*, 2017.

[18] Karim Banawan and Sennur Ulukus. The capacity of private information retrieval from byzantine and colluding databases. *arXiv preprint arXiv:1706.01442*, 2017.

[19] Yiwei Zhang and Gennian Ge. Private information retrieval from MDS coded databases with colluding servers under several variant models. *arXiv preprint arXiv:1705.03186*, 2017.

[20] Razane Tajeddine, Oliver W. Gnilke, David Karpuk, Ragnar Freij-Hollanti, and Camilla Hollanti. Private information retrieval from coded storage systems with colluding, byzantine, and unresponsive servers. *arXiv preprint arXiv:1806.08006*, 2018.

[21] Qiwen Wang, Hua Sun, and Mikael Skoglund. The $\epsilon$-error capacity of symmetric pir with byzantine adversaries. *arXiv preprint arXiv:1809.03988*, 2018.

[22] Karim Banawan and Sennur Ulukus. Private information retrieval through wiretap channel ii: Privacy meets security. *arXiv preprint arXiv:1801.06171*, 2018.

[23] Qiwen Wang, Hua Sun, and Mikael Skoglund. The capacity of private information retrieval with eavesdroppers. *arXiv preprint arXiv:1804.10189*, 2018.

[24] Zhuqing Jia, Hua Sun, and Syed A. Jafar. Cross subspace alignment and the asymptotic capacity of $x$-secure $t$-private information retrieval. *arXiv preprint arXiv:1808.07457*, 2018.

[25] Hua Sun and Syed A. Jafar. The capacity of private computation. *arXiv preprint arXiv:1710.11098*, 2017.

[26] Hsuan-Yin Lin, Siddhartha Kumar, Eirik Rosnes, Alexandre Graell i Amat, and Eitan Yaakobi. Weakly-private information retrieval. *arXiv preprint arXiv:1901.06907*, 2019.