Communication Cost of Two-Database Symmetric Private Information Retrieval: A Conditional Disclosure of Multiple Secrets Perspective

Zhusheng Wang Sennur Ulukus
Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742
zhusheng@umd.edu *ulukus@umd.edu*

***Abstract*—We consider the total (upload plus download) communication cost of two-database symmetric private information retrieval (SPIR) through its relationship to conditional disclosure of secrets (CDS). In SPIR, a user wishes to retrieve a message out of K messages from N non-colluding and replicated databases without learning anything beyond the retrieved message, while no individual database learns the retrieved message index. In CDS, two parties each holding an individual input and sharing a common secret wish to disclose this secret to an external party in an efficient manner if and only if their inputs satisfy a public deterministic function. As a natural extension of CDS, we introduce conditional disclosure of multiple secrets (CDMS) where two parties share multiple i.i.d. common secrets rather than a single common secret as in CDS. We show that a special configuration of CDMS is equivalent to two-database SPIR. Inspired by this equivalence, we design download cost efficient SPIR schemes using bipartite graph representation of CDS and CDMS, and determine the exact minimum total communication cost of $N = 2$ database SPIR for $K = 3$ messages when the message length is 1.**

I. Introduction

As initially introduced in [1], symmetric private information retrieval (SPIR) refers to the problem where a user downloads a message out of K possible messages stored in N non-colluding and replicated databases in such a way that, not only no individual database can know which message the user has just downloaded, but also the user learns nothing about the remaining messages stored in the databases. The total communication cost of SPIR consists of two parts: the total number of bits sent from the user to the databases (*upload cost*) denoted by U, and the total number of bits downloaded by the user from the databases (*download cost*) denoted by D. For a message length of L bits, the total communication cost $(U+D)$ of SPIR depends on three basic parameters (N, K, L).

In [2], without any constraints on U and L, the optimal download cost for SPIR is found to be $\frac{NL}{N-1}$, which does not depend on K. In [3], without any constraints on D and L, the optimal upload cost for SPIR is found to be $\log_2(\lceil K^{\frac{1}{N-1}} \rceil)$ which does not depend on L. In addition, [4]–[10] explore the optimal download cost of SPIR under various extended conditions without a consideration on the upload cost (see also many other important variants of PIR and SPIR in [11]–[48]).

This work was supported by ARO Grant W911NF2010142, and NSF Grants CCF 17-13977 and ECCS 18-07348.

To the best of our knowledge, our paper is the first one to investigate the overall (upload and download) communication cost of SPIR with a particular focus on $L = 1$ in an information-theoretic setting. Our focus on $L = 1$ is motivated by two observations: First, as pointed out in [11], when L is allowed to approach infinity, download cost dominates the upload cost, and the consideration of total cost becomes trivial. Second, in some cryptographic applications, e.g., [4], [5], only $L = 1$ may make practical sense.

As a classical cryptographic primitive, conditional disclosure of secrets (CDS) is first introduced in [1] as well to help devise an achievable SPIR scheme. Since CDS itself functions as an essential building block in applications such as secret sharing and attribute based encryption [49]–[51], CDS has also attracted significant attention as a stand-alone computer science problem. Recently, information-theoretic CDS is formulated in [52], [53] to characterize the maximum number of secret bits that can be securely disclosed per communication bit whenever a pre-defined condition is satisfied.

In this paper, we first show the equivalence between a special CDMS configuration and the two-database SPIR. Following this equivalence, we investigate the total communication cost of two-database SPIR through the characteristics of CDS and CDMS. We utilize CDS/CDMS to determine an upload cost, and then proceed to minimize the download cost for the given fixed upload cost. We then consider the feasible upload and download cost achievable region. In the example of $K = 3$ and $L = 1$, we find two optimal corner points for the upload and download cost pair. These two corner points outperform the best-known results in the literature [2], [3] and lead to the optimal total communication cost.

II. Problem Formulation

A. Symmetric Private Information Retrieval

Following the classical SPIR problem statement in [2], we consider $N \geq 2$ non-colluding databases with each individual database storing the replicated set of $K \geq 2$ i.i.d. messages $W_{1:K}$. Moreover, L i.i.d. symbols within each message are uniformly selected from a sufficiently large finite field $\mathbb{F}_q$,

$$H(W_k) = L, \quad \forall k \tag{1}$$

$$H(W_{1:K}) = H(W_1) + \cdots + H(W_K) = KL \tag{2}$$

A random variable $\mathcal{F}$ is used to denote the randomness of the retrieval strategy selection implemented by the user. Due to the user privacy constraint, the realization of $\mathcal{F}$ is only known to the user, and unknown to any of the databases. Due to the database privacy constraint, databases need to share some amount of common randomness $\mathcal{R}$.

The message set $W_{1:K}$ stored in the databases is independent of retrieval strategy randomness $\mathcal{F}$, common randomness $\mathcal{R}$ and user's desired message index θ, which is a random variable uniformly distributed over the set $[K]$,

$$I(W_{1:K}; \theta, \mathcal{F}, \mathcal{R}) = 0 \tag{3}$$

Using the desired message index, the user generates a query for each database according to $\mathcal{F}$. Hence, the queries $Q_n^{[k]}, n \in [N]$ are deterministic functions of $\mathcal{F}$,

$$H(Q_1^{[k]}, \ldots, Q_N^{[k]} | \mathcal{F}) = 0, \quad \forall k \tag{4}$$

After receiving a query from the user, each database should respond with a truthful answer based on the stored message set and common randomness,

$$\text{[deterministic answer]}\ H(A_n^{[k]} | Q_n^{[k]}, W_{1:K}, \mathcal{R}) = 0, \forall n, \forall k \tag{5}$$

After collecting all N answers from the databases, the user should be able to decode the desired message reliably,

$$\text{[reliability]} \quad H(W_k | \mathcal{F}, A_{1:N}^{[k]}) = 0, \quad \forall k \tag{6}$$

Due to the user privacy constraint, the query generated to retrieve the desired message should be statistically indistinguishable from other queries, thus, for all $k, k' \in [K], k' \neq k$,

$$\begin{aligned} \text{[user privacy]} \quad & (Q_n^{[k]}, A_n^{[k]}, W_{1:K}, \mathcal{R}) \\ & \sim \ (Q_n^{[k']}, A_n^{[k']}, W_{1:K}, \mathcal{R}) \end{aligned} \tag{7}$$

Due to the database privacy constraint, the user should learn nothing about $W_{\bar{k}}$ which is the complement of W_k, i.e., $W_{\bar{k}} = \{W_1, \cdots, W_{k-1}, W_{k+1}, \cdots, W_K\}$,

$$\text{[database privacy]} \quad I(W_{\bar{k}}; \mathcal{F}, A_{1:N}^{[k]}) = 0, \quad \forall k \tag{8}$$

An achievable SPIR scheme is a scheme that satisfies the reliability constraint (6), the user privacy constraint (7) and the database privacy constraint (8). In this paper, we focus on the overall communication cost, which is a sum of the number of uploaded bits (named *upload cost* and denoted by U) and the number of downloaded bits (named *download cost* and denoted by D), within the retrieval scheme. As a consequence, the most efficient achievable scheme is the scheme with the lowest total communication cost, i.e., the one that achieves $C^* = \inf(U + D)$ over all achievable SPIR schemes.

B. Conditional Disclosure of a Secret

Two parties Alice and Bob possess their respective inputs X, Y and share a common secret S. Alice and Bob also share an independent randomness $\mathcal{R}$ to assist the secret disclosure of S. With the knowledge of the inputs X, Y but without knowing the common randomness $\mathcal{R}$, another party Carol wishes to learn the secret S under a specific condition by communicating with Alice and Bob simultaneously. Generally, this condition is described as a deterministic public function. Specifically, given a globally public function f, the secret S is disclosed to Carol if and only if $f(X, Y) = 1$ is true. By contrast, if $f(X, Y)$ is not equal to 1, no information about the secret S should be revealed to Carol. To that end, Alice sends a signal A_X and Bob sends another signal B_Y to Carol.

The signals are determined by all the information contained in Alice or Bob before being sent to Carol,

$$\begin{aligned} \text{[deterministic signal]} \quad & H(A_X | X, S, \mathcal{R}) = 0 \\ & H(B_Y | Y, S, \mathcal{R}) = 0 \end{aligned} \tag{9}$$

If the condition is satisfied, Carol is able to decode the secret by using all the information she possesses,

$$\text{[validity]}\ H(S | X, Y, A_X, B_Y) = 0, \ \text{ if } f(X, Y) = 1 \tag{10}$$

Otherwise, if the condition is not satisfied, Carol cannot learn anything about the secret based on all the information she has,

$$\text{[security]}\ I(S; X, Y, A_X, B_Y) = 0, \ \text{ if } f(X, Y) \neq 1 \tag{11}$$

The information-theoretic objective of CDS is to minimize the number of bits contained in A_X and B_Y.

C. Conditional Disclosure of Multiple Secrets

Here, we introduce the concept of CDMS as an extension of CDS. Given the same setting except sharing K i.i.d. common secrets $S_1, \ldots, S_K$ in Alice and Bob, Carol expects to learn partial secrets under some specific conditions (one for each secret) by communicating with Alice and Bob simultaneously. Now, a sequence of functions $f_k, k \in [K]$ are globally public, the constraints in CDMS generalize to the following ones.

The integrated signals are determined by all the information contained in Alice or Bob before being sent to Carol,

$$\begin{aligned} \text{[deterministic signal]} \quad & H(A_X | X, S_{1:K}, \mathcal{R}) = 0 \\ & H(B_Y | Y, S_{1:K}, \mathcal{R}) = 0 \end{aligned} \tag{12}$$

For all $k \in [K]$, if the condition f_k is satisfied, Carol is able to decode the secret S_k,

$$\text{[validity]}\ H(S_k | X, Y, A_X, B_Y) = 0, \ \text{if } f_k(X, Y) = 1 \tag{13}$$

For all $k \in [K]$, if the condition f_k is not satisfied, Carol learns nothing about the secret S_k,

$$\text{[security]}\ I(S_k; X, Y, A_X, B_Y) = 0, \ \text{if } f_k(X, Y) \neq 1 \tag{14}$$

Likewise, the information-theoretic objective of CDMS is to minimize the number of bits contained in A_X and B_Y.

III. Main Results

We design the particular CDMS configuration given below:

1) First, Carol selects a random desired index θ, which is uniformly distributed over $[K]$; θ is independent of the secrets as well as common randomness in Alice and Bob.
2) Second, Carol selects two random vectors X and Y such that no information about θ is leaked in the individual vectors X or Y, i.e., $I(\theta; X) = 0$ and $I(\theta; Y) = 0$.

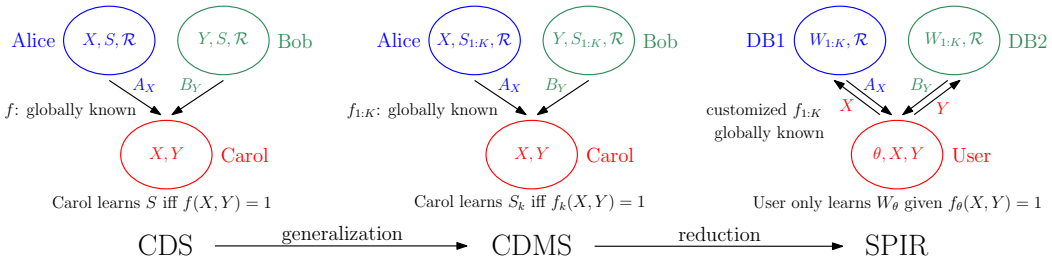

Fig. 1. Relationship among CDS, CDMS and SPIR.

3) Third, Carol sends X to Alice and Y to Bob.
4) Globally known condition functions are set in accordance with the selection of random vectors X and Y, such that, at all times only one condition function f_θ can be 1.

Theorem 1 *CDMS configured as above is equivalent to SPIR with two replicated and non-colluding databases.*

Proof: Within the given configuration, Alice and Bob can be treated as database 1 and database 2, and Carol as the user; the secrets $S_{1:K}$ can be treated as the message set $W_{1:K}$; the random variable θ as the desired message index at the user; the inputs X, Y as the queries $Q_1^{[\theta]}, Q_2^{[\theta]}$; and the signals A_X, B_Y as the answers $A_1^{[\theta]}, A_2^{[\theta]}$. Thus, we have the following conversions, which complete the proof:

1) *Deterministic signal* becomes *deterministic answer*,

$$H(A_1^{[\theta]}|Q_1^{[\theta]}, W_{1:K}, \mathcal{R}) = 0 \tag{15}$$
$$H(A_2^{[\theta]}|Q_2^{[\theta]}, W_{1:K}, \mathcal{R}) = 0 \tag{16}$$

2) From the first two steps in the CDMS configuration, we obtain the *user privacy* for each database,

$$I(\theta; Q_1^{[\theta]}, A_1^{[\theta]}, W_{1:K}, \mathcal{R}) = 0 \tag{17}$$
$$I(\theta; Q_2^{[\theta]}, A_2^{[\theta]}, W_{1:K}, \mathcal{R}) = 0 \tag{18}$$

3) *Validity* becomes *reliability* due to the unique decodable secret S_θ,

$$H(W_\theta|Q_1^{[\theta]}, Q_2^{[\theta]}, A_1^{[\theta]}, A_2^{[\theta]}) = 0 \tag{19}$$

4) *Security* becomes *database privacy* due to the remaining undecodable secrets,

$$I(W_{\tilde{\theta}}; Q_1^{[\theta]}, Q_2^{[\theta]}, A_1^{[\theta]}, A_2^{[\theta]}) = 0 \tag{20}$$

■

We are ready to investigate the total communication cost of two-database SPIR by means of the characteristics of CDS and CDMS. We use the terminologies in [54] for the bipartite graph in CDS/CDMS.

Remark 1 *We can construct an upload cost starting from $2\log_2 K$ in two-database SPIR while satisfying the constraints in the second step of the particular CDMS configuration above. Intuitively, the upload cost of $2\log_2 K$ comes from the needed $\log_2 K$ bits to be sent to each database to represent any one of the K messages. The upload cost $2\log_2 K$ can be achieved by the following setting: X and Y are two uniformly selected symbols from a finite set $\mathbb{S}_K = \{0, 1, \ldots, K-1\}$ such that $X + Y = \theta - 1$ under an assumption that the sum is always calculated over module K. In order to construct a larger upload cost, we can select a larger finite set by utilizing additional dummy messages. As an aside, we note that a larger finite set can be denoted by using multiple symbols from a smaller finite set. This further increases the diversity of upload cost constructions. For example, we can use two symbols from $\mathbb{S}_3 = \{0, 1, 2\}$ to include every option in $\mathbb{S}_8 = \{0, 1, \ldots, 7\}$. Thus, when $K = 8$, X and Y can either be two one-symbol vectors from S_8 or two two-symbol vectors from S_3.*

Remark 2 *As in CDS and CDMS, we can use a bipartite graph to specify two-database SPIR constraints. As introduced in [52], [53], CDS can be viewed as a data storage system over a bipartite graph where the nodes in each side of the graph are used to denote the input values in each party, and the connectivity of the links is used to indicate the satisfaction of the condition after selecting two nodes (input values) from two parties. In the extension to CDMS, we assign a distinct color c_k to each independent secret $S_k, \forall k \in [K]$. Hence, in CDMS, the color of links is used to indicate which secret should be revealed while keeping all the other secrets completely private. Following CDMS, in two-database SPIR, the nodes are used to denote the queries received by the databases, and the links with different colors are used to indicate which message should be retrieved while keeping all the other messages completely private, which implies reliability and database privacy.*

Remark 3 *In the bipartite graph, the links that are incident to any node should include all possible colors with equal number, due to user privacy.*

Example 1 *In this example, we will show the use of bipartite graphs for SPIR for $N = 2$, $K = 3$ and two example upload costs of $U = 2\log_2 3$ and $U = 4$. We use colors red, yellow and green to denote messages W_1, W_2 and W_3, respectively.*

For upload cost of $U = 2\log_2 3$, we use one-symbol vectors X and Y where X and Y are both uniformly selected from $\mathbb{S}_3$ s.t. $X + Y = \theta - 1$ for message W_θ. In this case, globally known condition functions are set accordingly as: $f_i(X, Y) = X + Y + 2 - i$, for $i \in [3]$. Then, we use A_0, A_1, A_2 to denote the three choices for the queries in database 1, and B_0, B_1, B_2

to denote the three choices for the queries in database 2. The corresponding bipartite graph is shown in Fig. 2.

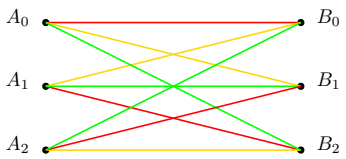

Fig. 2. Bipartite graph for $K = 3$ messages and $U = 2\log_2 3$ upload cost.

For upload cost of $U = 4$, we use two-symbol vectors $X = \{X_2, X_1\}$ and $Y = \{Y_2, Y_1\}$ where X_1, X_2, Y_1, Y_2 are all uniformly selected from $\mathbb{S}_2$ s.t. $2(X_2 + Y_2) + (X_1 + Y_1) = \theta - 1$ for message W_θ. The setting of globally known condition functions is similar: $f_i(X, Y) = 2(X_2+Y_2)+(X_1+Y_1)+2-i$, for $i \in [3]$. Then, $A_{00}, A_{01}, A_{10}, A_{11}$ and $B_{00}, B_{01}, B_{10}, B_{11}$ are used to denote the choices for the queries in two databases. The corresponding bipartite graph is shown in Fig. 3.

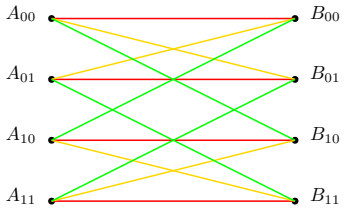

Fig. 3. Bipartite graph for $K = 3$ messages and $U = 4$ upload cost.

Remark 4 *Given an achievable scheme for two-database SPIR with $K = P$ messages with known upload cost U and download cost D, we can construct a new achievable scheme for $K = 2P$ messages with upload cost $U + 2$ and download cost $2D$. We use the following simple example to illustrate the idea of the general construction.*

Example 2 *Consider two-database SPIR with $K = 4$ messages, where colors red, yellow, green, blue are assigned to messages W_1, W_2, W_3, W_4, respectively. Now, first consider a two-database SPIR with $K = 2$ messages with a special bipartite graph provided in Fig. 4. Following this bipartite graph, we generate an SPIR achievable scheme for $K = 2$ and $L = 1$, with $U = 2$ and $D = 2$ as follows:*

$$A_0 = R_1, \qquad B_0 = W_1 + R_1 \tag{21}$$
$$A_1 = W_1 + W_2 + R_1, \qquad B_1 = W_2 + R_1 \tag{22}$$

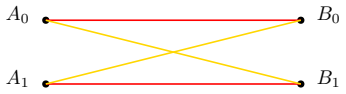

Fig. 4. Bipartite graph for $K = 2$ messages and $U = 2$ upload cost.

Now, we use the bipartite graph in Fig. 4 as a building block to construct an SPIR scheme for $K = 4$ messages as stated in Remark 4. First, we replicate this bipartite graph, thus, we need to use one extra bit to describe the query choices in each database, see the left part of Fig. 5. Then, we replicate the whole left part, change the color of links to green and blue, and then also exchange the order of query choices in the second column, see the right part of Fig. 5. Combining the left part and the right part in Fig. 5, we can verify that this new bipartite graph is a valid one by checking Remark 2 and Remark 3. Moreover, following this bipartite graph for $K = 4$, the corresponding upload cost increases by 2 and the corresponding download cost doubles; see the following achievable scheme with $L = 1$:

$$A_{00} = \{R_1, R_3\}, \quad B_{00} = \{W_1 + R_1, W_3 + R_4\} \tag{23}$$
$$A_{01} = \{W_1 + W_2 + R_1, W_3 + W_4 + R_3\}, \tag{24}$$
$$B_{01} = \{W_2 + R_1, W_4 + R_4\} \tag{25}$$
$$A_{10} = \{R_2, R_4\}, \quad B_{10} = \{W_1 + R_2, W_3 + R_3\} \tag{26}$$
$$A_{11} = \{W_1 + W_2 + R_2, W_3 + W_4 + R_4\}, \tag{27}$$
$$B_{11} = \{W_2 + R_2, W_4 + R_3\} \tag{28}$$

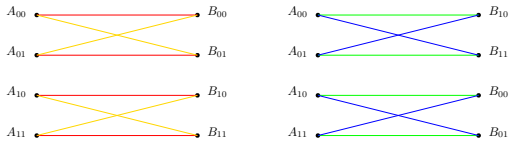

Fig. 5. Bipartite graph for $K = 4$ messages and $U = 4$ upload cost.

IV. Exact Upload-Download Region $N = 2$, $K = 3$

In this section, we give the exact achievable (U, D) cost region of two-database SPIR for $K = 3$ messages with $L = 1$ using the results of the previous section. In particular, for the upload cost of $U = 2\log_2 3$, we achieve a download cost of $D = 3$. This outperforms the best-known result of $D = 4$ in [3]. We show $(2\log_2 3, 3)$ corner point to be optimum with a converse. Further, by increasing the query selection for each database by one, we achieve a download cost of $D = 2$. This means that $U = 4$ is sufficient to achieve $D = 2$, and having $U = 6$ is not necessary as in [2]. We show $(4, 2)$ corner point to be optimum as well with a converse.

Theorem 2 *In the two-database SPIR with $K = 3$ messages with message length L, when the upload cost is $U = 2\log_2 3$, the optimal download cost is $D = 3L$ and the minimal amount of required common randomness is $2L$.*

Corollary 1 *In the two-database SPIR with $K = 3$ messages, if the message length is confined to be $L = 1$, the optimal total communication cost is $2\log_2 3 + 3$ with minimal amount of required common randomness being 2.*

Proof: We present the converse proof first. First, we select two random nodes from the two columns. Without loss of generality, let them be A_1 and B_1, respectively. From A_1, B_1, we can recover one random message W_p without learning anything about the remaining messages $W_{\bar{p}}$. Next, we select another two nodes A_i, $i \neq 1$ and B_j, $j \neq 1$ such that W_p

can be recovered from A_i, B_j once again with no knowledge about $W_{\bar{p}}$. Thus, from A_i, B_1, we can only recover another random message $W_q, q \neq p$. Then, we have,

$$
\begin{aligned}
&H(A_1|\mathcal{F}) + H(B_1|\mathcal{F}) \\
&\geq H(A_1|A_i, B_1, \mathcal{F}) + H(B_1|A_i, B_j, \mathcal{F}) && (29)\\
&= H(A_1, A_i, B_1, \mathcal{F}) + H(A_i, B_1, B_j, \mathcal{F}) \\
&\quad - H(A_i, B_1, \mathcal{F}) - H(A_i, B_j, \mathcal{F}) && (30)\\
&= H(W_p, A_1, A_i, B_1, \mathcal{F}) + H(W_p, A_i, B_1, B_j, \mathcal{F}) \\
&\quad - H(A_i, B_1, \mathcal{F}) - H(A_i, B_j, \mathcal{F}) && (31)\\
&\geq H(W_p, A_i, B_1, \mathcal{F}) + H(W_p, A_1, A_i, B_1, B_j, \mathcal{F}) \\
&\quad - H(A_i, B_1, \mathcal{F}) - H(A_i, B_j, \mathcal{F}) && (32)\\
&= H(W_p) + H(W_p, A_1, A_i, B_1, B_j, \mathcal{F}) - H(A_i, B_j, \mathcal{F}) && (33)\\
&= H(W_p) + H(W_{\bar{p}}, A_1, A_i, B_1, B_j, \mathcal{F}) - H(A_i, B_j, \mathcal{F}) && (34)\\
&\geq H(W_p) + H(W_{\bar{p}}, A_i, B_j, \mathcal{F}) - H(A_i, B_j, \mathcal{F}) && (35)\\
&= H(W_p) + H(W_{\bar{p}}) && (36)\\
&= 3L && (37)
\end{aligned}
$$

where (31) follows from the decodability of message W_p from A_1, B_1 and from A_i, B_j, (32) follows form the fact that conditioning cannot increase entropy, i.e., $H(A_1|W_p, A_i, B_1, \mathcal{F}) \geq H(A_1|W_p, A_i, B_1, B_j, \mathcal{F})$, (33) and (36) both come from the database privacy (8), (34) follows from the fact that we can decode $W_{1:3}$ from $A_1, B_1, A_i, B_j, \mathcal{F}$, which can be readily proved by contradiction in the bipartite graph. As a result, we reach the desired converse result regarding the download cost,

$$D \geq H(A_1) + H(B_1) \geq H(A_1|\mathcal{F}) + H(B_1|\mathcal{F}) \geq 3L \quad (38)$$

Next, we prove $H(\mathcal{R}) \geq 2L$:

$$
\begin{aligned}
0 &= I(W_{\bar{p}}; A_1, B_1, \mathcal{F}) && (39)\\
&= I(W_{\bar{p}}; A_1, B_1|W_p, \mathcal{F}) && (40)\\
&= H(A_1, B_1|W_p, \mathcal{F}) - H(A_1, B_1|W_{1:K}, \mathcal{F}) \\
&\quad + H(A_1, B_1|W_{1:K}, \mathcal{F}, \mathcal{R}) && (41)\\
&= H(A_1, B_1|W_p, \mathcal{F}) - I(A_1, B_1; \mathcal{R}|W_{1:K}, \mathcal{F}) && (42)\\
&\geq H(A_1, B_1|W_p, \mathcal{F}) - H(\mathcal{R}) && (43)
\end{aligned}
$$

where (40) follows from the combination of (2) and (3), (41) follows from the deterministic answers (5), and (43) follows from (3) again. Therefore, we turn to find a lower bound for the expression $H(A_1, B_1|W_p, \mathcal{F})$,

$$
\begin{aligned}
&H(A_1, B_1|W_p, \mathcal{F}) \\
&\geq H(A_1|W_p, A_i, B_1, \mathcal{F}) + H(B_1|W_p, A_i, B_j, \mathcal{F}) && (44)\\
&= H(A_1, A_i, B_1, \mathcal{F}) + H(A_i, B_1, B_j, \mathcal{F}) && (45)\\
&\quad - H(W_p, A_i, B_1, \mathcal{F}) - H(A_i, B_j, \mathcal{F}) && (46)\\
&= H(A_1, A_i, B_1, \mathcal{F}) + H(A_i, B_1, B_j, \mathcal{F}) && (47)\\
&\quad - H(A_i, B_1, \mathcal{F}) - H(A_i, B_j, \mathcal{F}) - H(W_p) && (48)\\
&\geq H(W_p) + H(W_{\bar{p}}) - H(W_p) && (49)\\
&= 2L && (50)
\end{aligned}
$$

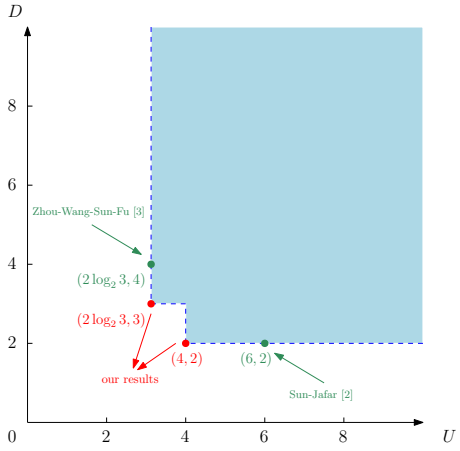

Fig. 6. Achievable (U, D) region for two-database SPIR with $K = 3$, $L = 1$.

where (49) exactly follows from the steps between (30)-(36). As a consequence, we reach the desired converse result regarding the minimal amount of required common randomness,

$$H(\mathcal{R}) \geq 2L \quad (51)$$

Next, we move to the achievability. We use the structure in Fig. 2 and the corresponding answers for $L = 1$ are as follows (we use this achievable scheme multiple times for larger L),

$$A_0{=}(R_1, R_2), B_0{=}W_1{+}R_1 \quad (52)$$
$$A_1{=}(W_1{+}W_2{+}R_1, W_2{+}W_3{+}R_2), B_1{=}W_2{+}R_2 \quad (53)$$
$$A_2{=}(W_1{+}W_3{+}R_1, W_1{+}W_2{+}R_2), B_2{=}W_3{+}R_1{+}R_2 \quad (54)$$

The achievability in (52)-(54) together with converses in (38) and (51) complete the proof. ■

Theorem 3 *In the two-database SPIR with $K = 3$ messages with message length L, when the upload cost is $U = 2\log_2 4 = 4$, the optimal download cost is $D = 2L$ and the minimal amount of required common randomness is L.*

Corollary 2 *In the two-database SPIR with $K = 3$ messages, if the message length is confined to be $L = 1$, the optimal total communication cost is $4 + 2 = 6$ with minimal amount of required common randomness being* 1.

Proof: The converse proof comes from [2, Thm. 1]. The achievability comes from the following answers corresponding to the structure in Fig. 3,

$$A_{00} = R_1,\ B_{00} = W_1 + R_1 \quad (55)$$
$$A_{01} = W_1 + W_2 + R_1,\ B_{01} = W_2 + R_1 \quad (56)$$
$$A_{10} = W_1 + W_3 + R_1,\ B_{10} = W_3 + R_1 \quad (57)$$
$$A_{11} = W_2 + W_3 + R_1,\ B_{11} = W_1 + W_2 + W_3 + R_1 \quad (58)$$

which complete the proof. ■

Combining Theorem 2 and Theorem 3, we obtain the achievable (U, D) region for two-database SPIR for $K = 3$ and $L = 1$ in Fig. 6. Any point within the light blue area is achievable, while all the remaining points are not achievable. Thus, the optimal communication cost is $4 + 2 = 6$.

REFERENCES

[1] Y. Gertner, Y. Ishai, E. Kushilevitz, and T. Malkin. Protecting data privacy in private information retrieval schemes. In *ACM STOC*, May 1998.

[2] H. Sun and S. A. Jafar. The capacity of symmetric private information retrieval. *IEEE Trans. on Info. Theory*, 65(1):322–329, January 2019.

[3] Y. Zhou, Q. Wang, H. Sun, and S. Fu. The minimum upload cost of symmetric private information retrieval. In *IEEE ISIT*, June 2020.

[4] Z. Wang, K. Banawan, and S. Ulukus. Private set intersection: A multi-message symmetric private information retrieval perspective. *IEEE Trans. on Info. Theory*, 68(3):2001–2019, March 2022.

[5] Z. Wang, K. Banawan, and S. Ulukus. Multi-party private set intersection: An information-theoretic approach. *IEEE Jour. on Selected Areas in Info. Theory*, 2(1):366–379, March 2021.

[6] Z. Wang and S. Ulukus. Symmetric private information retrieval with user-side common randomness. In *IEEE ISIT*, July 2021.

[7] Q. Wang and M. Skoglund. On PIR and symmetric PIR from colluding databases with adversaries and eavesdroppers. *IEEE Trans. on Info. Theory*, 65(5):3183–3197, May 2019.

[8] Q. Wang, H. Sun, and M. Skoglund. Symmetric private information retrieval with mismatched coded messages and randomness. In *IEEE ISIT*, July 2019.

[9] Q. Wang and M. Skoglund. Symmetric private information retrieval from MDS coded distributed storage with non-colluding and colluding servers. *IEEE Trans. on Info. Theory*, 65(8):5160–5175, August 2019.

[10] J. Cheng, N. Liu, and W. Kang. The capacity of symmetric private information retrieval under arbitrary collusion and eavesdropping patterns. Available at arXiv:2010.08249.

[11] H. Sun and S. A. Jafar. The capacity of private information retrieval. *IEEE Trans. on Info. Theory*, 63(7):4075–4088, July 2017.

[12] C. Tian, H. Sun, and J. Chen. Capacity-achieving private information retrieval codes with optimal message size and upload cost. *IEEE Trans. on Info. Theory*, 65(11):7613–7627, November 2019.

[13] N. B. Shah, K. V. Rashmi, and K. Ramchandran. One extra bit of download ensures perfectly private information retrieval. In *IEEE ISIT*, June 2014.

[14] 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.

[15] 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.

[16] 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. *IEEE Trans. on Info. Theory*, 65(6):3898–3906, June 2019.

[17] S. Kumar, H.-Y. Lin, E. Rosnes, and A. G. i Amat. Achieving maximum distance separable private information retrieval capacity with linear codes. *IEEE Trans. on Info. Theory*, 65(7):4243–4273, July 2019.

[18] R. G. L. D'Oliveira and S. El Rouayheb. One-shot PIR: Refinement and lifting. *IEEE Trans. on Info. Theory*, 66(4):2443–2455, April 2020.

[19] R. Zhou, C. Tian, H. Sun, and T. Liu. Capacity-achieving private information retrieval codes from MDS-coded databases with minimum message size. *IEEE Trans. on Info. Theory*, 66(8):4904–4916, August 2020.

[20] K. Banawan and S. Ulukus. Private information retrieval from non-replicated databases. In *IEEE ISIT*, July 2019.

[21] N. Raviv, I. Tamo, and E. Yaakobi. Private information retrieval in graph-based replication systems. *IEEE Trans. on Info. Theory*, 66(6):3590–3602, June 2020.

[22] H. Sun and S. A. Jafar. The capacity of private computation. *IEEE Trans. on Info. Theory*, 65(6):3880–3897, June 2019.

[23] K. Banawan and S. Ulukus. Asymmetry hurts: Private information retrieval under asymmetric traffic constraints. *IEEE Trans. on Info. Theory*, 65(11):7628–7645, November 2019.

[24] K. Banawan and S. Ulukus. Noisy private information retrieval: On separability of channel coding and information retrieval. *IEEE Trans. on Info. Theory*, 65(12):8232–8249, December 2019.

[25] K. Banawan and S. Ulukus. Private information retrieval through wiretap channel II: Privacy meets security. *IEEE Trans. on Info. Theory*, 66(7):4129–4149, July 2020.

[26] Z. Chen, Z. Wang, and S. A. Jafar. The asymptotic capacity of private search. *IEEE Trans. on Info. Theory*, 66(8):4709–4721, August 2020.

[27] X. Yao, N. Liu, and W. Kang. The capacity of private information retrieval under arbitrary collusion patterns for replicated databases. *IEEE Trans. on Info. Theory*, 67(10):6841–6855, October 2021.

[28] S. Vithana, K. Banawan, and S. Ulukus. Semantic private information retrieval. *IEEE Trans. on Info. Theory*, 68(4):2635–2652, April 2022.

[29] S. Kadhe, B. Garcia, A. Heidarzadeh, S. El Rouayheb, and A. Sprintson. Private information retrieval with side information. *IEEE Trans. on Info. Theory*, 66(4):2032–2043, April 2020.

[30] S. Li and M. Gastpar. Single-server multi-message private information retrieval with side information. In *Allerton Conference*, October 2018.

[31] S. Kumar, A. G. i Amat, E. Rosnes, and L. Senigagliesi. Private information retrieval from a cellular network with caching at the edge. *IEEE Trans. on Commun.*, 67(7):4900–4912, July 2019.

[32] R. Tandon. The capacity of cache aided private information retrieval. In *Allerton Conference*, October 2017.

[33] Y.-P. Wei, K. Banawan, and S. Ulukus. Cache-aided private information retrieval with partially known uncoded prefetching: Fundamental limits. *IEEE JSAC*, 36(6):1126–1139, June 2018.

[34] Y.-P. Wei, K. Banawan, and S. Ulukus. Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching. *IEEE Trans. on Info. Theory*, 65(5):3215–3232, May 2019.

[35] Y.-P. Wei, K. Banawan, and S. Ulukus. The capacity of private information retrieval with partially known private side information. *IEEE Trans. on Info. Theory*, 65(12):8222–8231, December 2019.

[36] Y.-P. Wei and S. Ulukus. The capacity of private information retrieval with private side information under storage constraints. *IEEE Trans. on Info. Theory*, 66(4):2023–2031, April 2020.

[37] Z. Chen, Z. Wang, and S. Jafar. The capacity of T-private information retrieval with private side information. *IEEE Trans. on Info. Theory*, 66(8):4761–4773, August 2020.

[38] M. J. Siavoshani, S. P. Shariatpanahi, and M. A. Maddah-Ali. Private information retrieval for a multi-message scenario with private side information. *IEEE Trans. on Commun.*, 69(5):3235–3244, May 2021.

[39] T. Guo, R. Zhou, and C. Tian. On the information leakage in private information retrieval systems. *IEEE Trans. on Info. Forensics and Security*, 15:2999–3012, 2020.

[40] I. Samy, M. Attia, R. Tandon, and L. Lazos. Asymmetric leaky private information retrieval. *IEEE Trans. on Info. Theory*, 67(8):5352–5369, August 2021.

[41] 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.

[42] Z. Jia, H. Sun, and S. A. Jafar. Cross subspace alignment and the asymptotic capacity of X-secure T-private information retrieval. *IEEE Trans. on Info. Theory*, 65(9):5783–5798, September 2019.

[43] Z. Jia and S. A. Jafar. X-secure T-private information retrieval from MDS coded storage with Byzantine and unresponsive servers. *IEEE Tran. on Info. Theory*, 66(12):7427–7438, December 2020.

[44] Y.-P. Wei, B. Arasli, K. Banawan, and S. Ulukus. The capacity of private information retrieval from decentralized uncoded caching databases. *Information*, 10, December 2019.

[45] K. Banawan, B. Arasli, Y.-P. Wei, and S. Ulukus. The capacity of private information retrieval from heterogeneous uncoded caching databases. *IEEE Trans. on Info. Theory*, 66(6):3407–3416, June 2020.

[46] M. A. Attia, D. Kumar, and R. Tandon. The capacity of private information retrieval from uncoded storage constrained databases. *IEEE Trans. on Info. Theory*, 66(11):6617–6634, November 2020.

[47] K. Banawan, B. Arasli, and S. Ulukus. Improved storage for efficient private information retrieval. In *IEEE ITW*, August 2019.

[48] C. Tian. On the storage cost of private information retrieval. *IEEE Trans. on Info. Theory*, 66(12):7539–7549, December 2020.

[49] H.-M. Sun and S.-P. Shieh. Secret sharing in graph-based prohibited structures. In *IEEE Infocom*, April 1997.

[50] A. Sahai and B. Waters. Fuzzy identity-based encryption. In *Advances in Cryptology - EUROCRYPT*, May 2005.

[51] B. Applebaum, B. Arkis, P. Raykov, and P. N. Vasudevan. Conditional disclosure of secrets: Amplification, closure, amortization, lower-bounds, and separations. In *Advances in Cryptology – CRYPTO*, 2017.

[52] Z. Li and H. Sun. Conditional disclosure of secrets: A noise and signal alignment approach. *IEEE Trans. on Commun.*, 2022. Early Access.

[53] Z. Li and H. Sun. On the linear capacity of conditional disclosure of secrets. Available at arXiv preprint arXiv:2106.04483.

[54] B. Albert-László. *Network Science*. Cambridge university press, 2016.