The Capacity of Symmetric Private Information Retrieval

Hua Sun, *Member, IEEE*, and Syed Ali Jafar, *Fellow, IEEE*

***Abstract*— Private information retrieval (PIR) is the problem of retrieving, as efficiently as possible, one out of K messages from N non-communicating replicated databases (each holds all K messages) while keeping the identity of the desired message index a secret from each individual database. Symmetric PIR (SPIR) is a generalization of PIR to include the requirement that beyond the desired message, the user learns nothing about the other $K-1$ messages. The information theoretic capacity of SPIR (equivalently, the reciprocal of minimum download cost) is the maximum number of bits of desired information that can be privately retrieved per bit of downloaded information. We show that the capacity of SPIR is $1-1/N$ regardless of the number of messages K, if the databases have access to common randomness (not available to the user) that is independent of the messages, in the amount that is at least $1/(N-1)$ bits per desired message bit. Otherwise, if the amount of common randomness is less than $1/(N-1)$ bits per message bit, then the capacity of SPIR is zero. Extensions to the capacity region of SPIR and the capacity of finite length SPIR are provided.**

***Index Terms*— Capacity, private information retrieval, symmetric private information retrieval.**

I. Introduction

THE private information retrieval (PIR) problem [1], [2] seeks the most efficient way for a user to retrieve a desired message from a set of distributed databases, each of which stores all the messages, without revealing any information about which message is being retrieved to any individual database. This seemingly impossible mission has a trivial (expensive) solution, i.e., the user can request all the messages to hide his interest. The goal of the PIR problem is to find the most efficient solution. The capacity of PIR is defined as the maximum number of bits of desired message that can be privately downloaded per bit of downloaded information. In our recent work [3], the capacity of PIR with K messages and N databases was shown to be $C_{PIR} = (1 + 1/N + \cdots + 1/N^{K-1})^{-1}$.

Manuscript received July 11, 2017; revised April 16, 2018; accepted June 14, 2018. Date of publication June 19, 2018; date of current version December 19, 2018. This work was supported in part by ONR under Grants N00014-16-1-2629 and N00014-18-1-2057, in part by NSF under Grants CCF-1317351, CCF-1617504, and CNS-1731384, and in part by ARL under Grant W911NF-16-1-0215. This paper was presented in part at the 2016 IEEE Workshop on Network Coding and Applications.

H. Sun is with the Department of Electrical Engineering, University of North Texas, Denton, TX 76203 USA (e-mail: hua.sun@unt.edu).

S. A. Jafar is with the Center for Pervasive Communications and Computing, Department of Electrical Engineering and Computer Science, University of California at Irvine, Irvine, CA 92697 USA (e-mail: syed@uci.edu).

Communicated by A. Khisti, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2018.2848977

The original formulation of PIR only considers the privacy of the user. The privacy of the undesired messages is ignored. However, it is often desirable to restrict the user to retrieve nothing beyond his chosen message. This new constraint is called database privacy, and with this constraint, the problem is called symmetric[1] PIR (SPIR) [4]. Symmetric PIR is especially challenging because the databases must individually learn nothing about the identity of the desired message, but must still collectively allow the user to retrieve his desired message in such a way that the user learns nothing about any other message besides his desired message. For example, the trivial solution of downloading everything, is no longer acceptable. The main contribution of this work is the characterization of the capacity of SPIR, i.e., the maximum number of bits of desired message that can be privately retrieved by a user per bit of downloaded information, without leaking any information about undesired messages to the user. For K messages and N databases, we show that the capacity is $1 - 1/N$. Extensions of the main result, from capacity to capacity region and from infinite message length to arbitrary message length, are also provided.

Besides its direct applications, PIR is especially significant as a fundamental problem that lies at the intersection of several open problems in cryptography [5], [6], coding theory [7]–[9] and complexity theory [10]. SPIR inherits many of these connections from PIR. For example, SPIR is essentially a (distributed) form of oblivious transfer [11], [12], where the typical objective is that the transmitter(s) should not know which message is received by the receiver and the receiver should obtain nothing more than the desired message. Oblivious transfer is an important building block (primitive) in cryptography, whose feasibility leads to many other cryptographic protocols [13], [14]. Fundamental limits on the communication efficiency of various forms of oblivious transfer therefore represent an important class of open problems [15], [16]. The capacity characterization of SPIR is a promising step in this direction.

In addition to this work, we note that there is much recent interest in characterizing the capacity of PIR and SPIR under various models. The extensions include the capacity of PIR with colluding and non-responsive servers [17]–[19], the capacity of PIR with finite message length [20], the capacity of PIR with multiple rounds [21], the capacity of PIR with multiple desired messages [22], the capacity of PIR with

[1] Symmetry means that the privacy of both the user and the database is considered.

coded storage [23]–[25], the capacity of colluding servers and MDS coded storage [26]–[28], the capacity of PIR with side information and/or caching [29]–[33], the capacity of PIR with byzantine and colluding servers and/or eavesdroppers [34], [35], the capacity of PIR with dependent messages [36]–[39], the capacity of PIR with colluding servers over small fields [40], the capacity of PIR with asymmetric traffic constraints [41], the capacity of SPIR with MDS coded storage [42], and the capacity of SPIR with colluding and byzantine servers and eavesdroppers [43].

Notation: For $n_1, n_2 \in \mathbb{Z}, n_1 \leq n_2$, define the notation $[n_1 : n_2]$ as the set $\{n_1, n_1+1, \cdots, n_2\}$. For an index set $\mathcal{I} = \{i_1, i_2, \cdots, i_n\}$, with $i_1 < i_2 < \cdots < i_n$, the notation $A_{\mathcal{I}}$ represents the vector $(A_{i_1}, A_{i_2}, \cdots, A_{i_n})$. For an element i_θ in the set $\mathcal{I} = \{i_1, i_2, \cdots, i_n\}$, i.e., $i_\theta \in \mathcal{I}$, the notation $\overline{i_\theta}$ represents the complement of $\{i_\theta\}$, i.e., $\overline{i_\theta} \triangleq \{i_1, \cdots, i_{\theta-1}, i_{\theta+1}, \cdots, i_n\}$.

II. Problem Statement

Consider K independent messages $W_1, \cdots, W_K, W_k \in \mathbb{F}_p^{l_k L \times 1}, k \in [1 : K], l_k \in \mathbb{Z}_+, L \in \mathbb{Z}_+$, where W_k is represented as an $l_k L \times 1$ vector comprised of $l_k L$ i.i.d. uniform symbols from a finite field $\mathbb{F}_p$ for a prime p. In p-ary units,

$$H(W_1, \cdots, W_K) = H(W_1) + \cdots + H(W_K), \quad (1)$$

$$H(W_k) = l_k L, \quad \forall k \in [1 : K]. \quad (2)$$

There are N databases. Each database stores all the messages $W_1, \cdots, W_K$.

Let us use $\mathcal{F}$ to denote a random variable privately generated by the user, whose realization is not available to the databases. $\mathcal{F}$ represents the randomness in the strategies followed by the user. The user privately generates θ uniformly from $[1 : K]$ and wishes to retrieve W_θ privately. The databases do not want to give out any information beyond the one message of the user's choosing (W_θ). In order to achieve database-privacy, we assume that the databases share a common random variable S that is not known to the user. It has been shown that without such common randomness, SPIR is not feasible [4].[2] For a pictorial illustration of an example of the SPIR problem with K messages and 2 databases, see Figure 1. $\mathcal{F}$ is generated independently and before the realizations of the messages, the common randomness or the desired message index are known, so that

$$H(\theta, \mathcal{F}, W_1, \cdots, W_K, S) = H(\theta) + H(\mathcal{F}) + H(W_1) + \cdots + H(W_K) + H(S). \quad (3)$$

[2]The common randomness model was introduced in the first SPIR paper [4], where it was argued that this is the minimal extension of the original PIR model in the following sense. If no common randomness is available (even if independent private randomness is allowed), SPIR protocols cannot simultaneously satisfy user-privacy, database-privacy and correctness constraints. The intuition might be seen as follows. In order to prohibit the answers from revealing non-desired information, the answers must be mixed with some randomness, and the randomness must be correlated so that the user may cancel the randomness and extract the desired message.

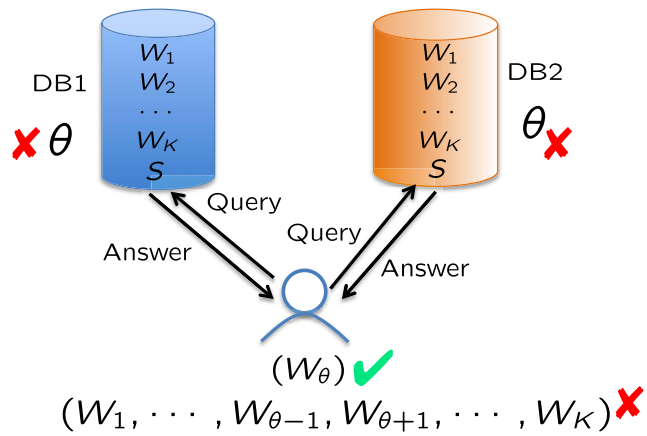

Fig. 1. The SPIR problem with K messages and 2 databases.

Suppose $\theta = k$. In order to retrieve message $W_k, k \in [1 : K]$ privately, the user privately generates N queries $Q_1^{[k]}, \cdots, Q_N^{[k]}$.

$$H(Q_1^{[k]}, \cdots, Q_N^{[k]} | \mathcal{F}) = 0, \forall k \in [1 : K]. \quad (4)$$

The user sends query $Q_n^{[k]}$ to the n-th database, $n \in [1 : N]$. Upon receiving $Q_n^{[k]}$, the n-th database generates an answering string $A_n^{[k]}$, which is a function of $Q_n^{[k]}$, all messages $W_1, \cdots, W_K$, and the common randomness S,

$$H(A_n^{[k]} | Q_n^{[k]}, W_1, \cdots, W_K, S) = 0. \quad (5)$$

Each database returns to the user its answer $A_n^{[k]}$.

From all the information that is now available to the user $(Q_{1:N}^{[k]}, A_{1:N}^{[k]}, \mathcal{F})$, the user decodes the desired message W_k according to a decoding rule that is specified by the SPIR scheme. Let P_e denote the probability of error achieved with the specified decoding rule.

To protect the user's privacy, the K strategies must be indistinguishable (identically distributed) from the perspective of any individual database, i.e., the following user-privacy constraint must be satisfied,[3]

$$\text{[User-Privacy]} \quad (Q_n^{[k]}, A_n^{[k]}, W_{1:K}, S) \sim (Q_n^{[k']}, A_n^{[k']}, W_{1:K}, S), \quad \forall k, k' \in [1 : K], \quad \forall n \in [1 : N]. \quad (6)$$

Symmetric PIR also requires protecting the privacy of the database, i.e., it must be ensured that the user learns nothing more than the desired message W_k. So the vector $W_{\overline{k}} = (W_1, \cdots, W_{k-1}, W_{k+1}, \cdots, W_K)$, must be independent of all the information available to the user. Thus, the following database-privacy constraint must be satisfied:

$$\text{[DB-Privacy]} \; I(W_{\overline{k}}; Q_{1:N}^{[k]}, A_{1:N}^{[k]}, \mathcal{F}) = 0, \forall k \in [1 : K]. \quad (7)$$

The SPIR rate of W_k characterizes the amount of desired information retrieved per downloaded symbol, and is defined as follows.

$$R_k \triangleq \frac{l_k L}{D} \quad (8)$$

[3]The User-Privacy constraint is equivalently expressed as $I(\theta; Q_n^{[\theta]}, A_n^{[\theta]}, W_{1:K}, S) = 0$.

where D is the maximum value of the total number of symbols downloaded by the user from all the databases.

A rate tuple $(R_1, \cdots, R_K)$ is said to be ϵ-error achievable if $\forall k \in [1:K]$, there exists a sequence of PIR schemes, indexed by L, where the rate of W_k is greater than or equal to R_k and $P_e \to 0$ as $L \to \infty$. Note that for such a sequence of SPIR schemes, from Fano's inequality, we must have

$$[\text{Correctness}]\ o(L) = H(W_k | Q_{1:N}^{[k]}, A_{1:N}^{[k]}, \mathcal{F}) \tag{9}$$

$$\overset{(4)}{=} H(W_k | A_{1:N}^{[k]}, \mathcal{F}) \tag{10}$$

where any function of L, say $f(L)$, is said to be $o(L)$ if $\lim_{L\to\infty} = f(L)/L = 0$. The closure of the set of all ϵ-error achievable rate tuples is called the capacity region $\mathcal{C}$.

III. Results

A. Capacity of SPIR

In the typical setting of SPIR, the sizes of the messages are the same, i.e., $l_k = 1, \forall k \in [1:K]$ and the rate (refer to (8)) of each message is the same. Then the capacity region is characterized by one single parameter, i.e., the supremum of the achievable rate, named the capacity. We denote the capacity as C.

When there is only $K = 1$ message, note that the database-privacy constraint is satisfied trivially, so that SPIR reduces to the PIR setting and the capacity is 1. For $K \geq 2$, it is known that some common randomness S is necessary for the feasibility of SPIR [4]. Let us define ρ as the amount of common randomness relative to the message size

$$\rho = \frac{H(S)}{H(W)} = \frac{H(S)}{L}. \tag{11}$$

The capacity should depend on ρ, and because availability of common randomness at the databases is a non-trivial requirement, this dependence is of some interest.

When there is only $N = 1$ database, it is easy to see that the database-privacy constraint, the user-privacy constraint and correctness constraint conflict with each other such that SPIR is not feasible and the capacity is zero. The reason is as follows. First, because of the user-privacy constraint (6), the answer from the only database $A_1^{[k]}$ is identically distributed for all $k \in [1:K]$. Second, from the correctness constraint (10), from $A_1^{[k]}, \mathcal{F}$, one can decode W_k. Combining these two facts, we have that from $A_1^{[k]}, \mathcal{F}$, one can decode all messages $W_1, \cdots, W_K$. This contradicts the database-privacy constraint (7). Therefore, when $N = 1$ and $K \geq 2$, SPIR is not feasible.

The following theorem states the capacity of SPIR, when we have $N \geq 2$ databases and $K \geq 2$ messages.

Theorem 1: For SPIR with $K \geq 2$ messages and $N \geq 2$ databases, the capacity is

$$C_{SPIR} = \begin{cases} 1 - 1/N & \text{if } \rho \geq \frac{1}{N-1}, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

The following observations place Theorem 1 in perspective.

1) We notice a surprising threshold phenomenon in the dependence of SPIR capacity, C_{SPIR}, on the amount of common randomness ρ. When $\rho < \frac{1}{N-1}$, SPIR is not feasible and $C_{SPIR} = 0$. However, when $\rho \geq \frac{1}{N-1}$, SPIR is not only possible, but the rate can immediately be increased to the maximum possible, i.e., the capacity. Therefore, the minimum common randomness required to achieve any positive rate is already sufficient to achieve the capacity of SPIR. A pictorial illustration of the SPIR capacity and its dependency on the amount of common randomness appears in Figure 2.

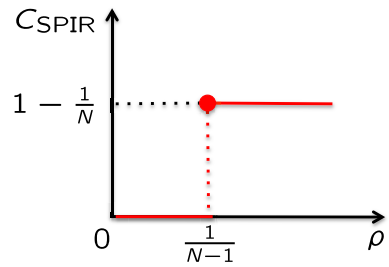

Fig. 2. SPIR capacity.

2) The capacity of SPIR is independent of the number of messages, K.
3) When the capacity is non-zero, the capacity is strictly increasing in the number of databases, N, and when N approaches infinity, the capacity approaches 1.
4) It is interesting to compare the capacity of SPIR and the capacity of PIR [3],

$$C_{PIR} = \left(1 + 1/N + 1/N^2 + \cdots + 1/N^{K-1}\right)^{-1}. \tag{13}$$

We see that the capacity of SPIR is strictly smaller than the capacity of PIR (the additional requirement of preserving database-privacy strictly hurts) and the capacity of PIR approaches the capacity of SPIR when the number of messages, K, approaches infinity (in the large number of messages regime, the penalty vanishes), i.e., $C_{PIR} > C_{SPIR}$ for any finite K and $C_{PIR} \to C_{SPIR}$ when $K \to \infty$. Specifically, the gap is computed as

$$\begin{aligned} &C_{PIR} - C_{SPIR} \\ &= \frac{1}{1 + 1/N + 1/N^2 + \cdots + 1/N^{K-1}} - (1 - 1/N) \\ &= \frac{1 - 1/N}{1 - 1/N^K} - (1 - 1/N) = \frac{1 - 1/N}{N^K - 1}. \end{aligned}$$

Therefore, the gap decreases exponentially with the number of messages, K. A pictorial illustration of this gap as a function of the number of messages, K, for various number of databases, N, is shown in Figure 3.

5) The achievable scheme presented in Section IV-A.1 has exactly zero error. Further, in the achievability proof for Theorem 1, the message size is $N - 1$ bits per message. Therefore, to achieve capacity, message size is not required to approach infinity. By employing the scheme multiple times, we know that when message size is equal to an integer multiple of $N-1$ bits, the capacity

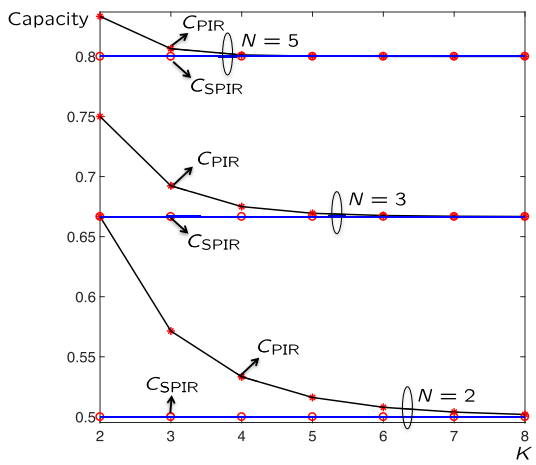

Fig. 3. The comparison of PIR capacity and SPIR capacity.

is achieved as well. When the message size is not equal to an integer multiple of $N-1$ bits, it turns out that there is a penalty in the form of a ceiling operation. This extension of SPIR to finite length messages is considered in Theorem 3, to be presented in Section III-C.

6) We note that the converse (upper bound, presented in Section IV-A.2) holds for arbitrary message size L when we require exactly zero error, by replacing the $o(L)$ terms with zero.
7) The extension to unequal message sizes is considered in Section III-B.

In the following sections, we relax each one of the two assumptions by itself, i.e., equal message sizes and message length L going to infinity.

B. Capacity Region of SPIR

In this section, we relax the assumption of equal message sizes, i.e., $l_k, \forall k \in [1:K]$ are arbitrary. Therefore, going beyond the (symmetric) capacity, we wish to characterize the capacity region of SPIR.

When we only have $K=1$ message, similar to the previous section, the capacity region is characterized by the capacity of one message, which is 1. When we only have $N=1$ database and $K \geq 2$ messages, similar to the previous section, SPIR is not feasible and the capacity region is the zero vector. Therefore, we consider $K \geq 2$ messages and $N \geq 2$ databases, where the capacity region of SPIR is characterized in the following theorem. Here the amount of common randomness is normalized with respective to the largest message size.

$$\rho = \frac{H(S)}{\max_{i:i\in[1:K]} H(W_i)} = \frac{H(S)}{\max_{i:i\in[1:K]} l_i L}. \tag{14}$$

Theorem 2: For SPIR with $K \geq 2$ messages and $N \geq 2$ databases, the capacity region $\mathcal{C}$ is

$$\mathcal{C} = \Big\{(R_1, \cdots, R_K) : R_k \leq \frac{l_k}{\max_i l_i}(1-\frac{1}{N}), \quad \forall k \in [1:K]\Big\}, \text{ if } \rho \geq \frac{1}{N-1} \tag{15}$$

and the zero vector otherwise.

Remark: The optimal (minimum) normalized download cost $D/L = l_k/R_k = \max_i l_i \frac{N}{N-1}$ is the same for each message.

C. Capacity of Finite Length SPIR

In this section, we again assume that all messages have the same length, but relax the assumption that L approaches infinity. Instead, we assume that L is an arbitrary finite value. As L is finite, we consider zero error achievable rates and define its supremum as zero error capacity, denoted as C_o. This setting can be obtained from the general problem statement by setting $l_k = 1, \forall k \in [1:K]$, and L finite.

Similar to Section III-A, we restrict to $K \geq 2$ and $N \geq 2$ cases as the problem is trivial when $K=1$ or $N=1$. The capacity of finite length SPIR is characterized in the following theorem. The relative size of the common randomness, ρ, is defined as in (11).

Theorem 3: For SPIR with $K \geq 2$ messages, $N \geq 2$ databases, where each message is of size $L \in \mathbb{Z}_+$ symbols, the zero error capacity is

$$C_{o,LSPIR} = \begin{cases} L/\lceil \frac{L}{C_{SPIR}} \rceil = L/\lceil \frac{L}{1-1/N} \rceil & \text{if } \rho \geq \frac{\lceil L/(N-1) \rceil}{L}, \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

The above two extensions - capacity region and finite length messages may be combined, i.e., the zero error capacity region of finite length SPIR is characterized in the following theorem. ρ is defined as in (14).

Theorem 4: For SPIR with $K \geq 2$ messages and $N \geq 2$ databases, where message $W_k, k \in [1:K]$ is of size $l_k L$ symbols, $l_k \in \mathbb{Z}_+, L \in \mathbb{Z}_+$, the zero error capacity region $\mathcal{C}_{o,LSPIR}$ is

$$\mathcal{C}_{o,LSPIR} = \Big\{(R_1, \cdots, R_K) : R_k \leq l_k L/\lceil \frac{L \max_i l_i}{1-1/N} \rceil, \quad \forall k \in [1:K]\Big\}, \text{ if } \rho \geq \frac{\lceil L/(N-1) \rceil}{L} \tag{17}$$

and the zero vector otherwise.

It is obvious that Theorem 2 and Theorem 3 are special cases of Theorem 4. Therefore, we only present the proof of Theorem 4, in Section IV-B.

IV. Proofs

A. Proof of Theorem 1

1) Achievability: In this section, we present the scheme that achieves rate $1-1/N$, when $\rho = 1/(N-1)$. To this end, we assume each message consists of $N-1$ bits and each answering string is 1 bit. Specifically, we assume $W_k = (x_{k,1}, \cdots, x_{k,N-1}), \forall k \in [1:K]$ where each $x_{k,i}, i \in [1:N-1]$ is one bit. We further assume the entropy of the common random variable S is 1 bit, i.e., S is uniformly distributed over $\{0,1\}$. Note that S is independent of the messages.

Next we specify the queries. To retrieve W_k privately, the user first generates a random vector of length $(N-1)K$, $[h_{1,1}, \cdots, h_{1,N-1}, \cdots, h_{k,1}, \cdots, h_{K,N-1}]$, where each

element is uniformly distributed over $\{0, 1\}$. Then the queries are set as follows.

$$\begin{aligned} Q_1^{[k]} &= [h_{1,1}, \cdots, h_{k,1}, \cdots, h_{k,N-1}, \cdots, h_{K,N-1}] \\ Q_2^{[k]} &= [h_{1,1}, \cdots, h_{k,1}+1, \cdots, h_{k,N-1}, \cdots, h_{K,N-1}] \\ &\cdots \\ Q_N^{[k]} &= [h_{1,1}, \cdots, h_{k,1}, \cdots, h_{k,N-1}+1, \cdots, h_{K,N-1}] \end{aligned} \tag{18}$$

where the addition '+' is over the binary field. The answering strings are generated by using the query vector as the combining coefficients and producing the corresponding linear combination of message bits. We further add the common random variable to each answer.

$$\begin{aligned} A_1^{[k]} &= \sum_{j=1}^{K}\sum_{i=1}^{N-1} h_{j,i}x_{j,i} + S \\ A_2^{[k]} &= \sum_{j=1}^{K}\sum_{i=1}^{N-1} h_{j,i}x_{j,i} + x_{k,1} + S \\ &\cdots \\ A_N^{[k]} &= \sum_{j=1}^{K}\sum_{i=1}^{N-1} h_{j,i}x_{j,i} + x_{k,N-1} + S \end{aligned} \tag{19}$$

The user obtains $x_{k,i}, i \in [1 : N-1]$ by subtracting $A_1^{[k]}$ from $A_{i+1}^{[k]}$. Therefore, the correctness condition is satisfied.

Privacy of the user is guaranteed because each query is independent of the desired message index k. This is because regardless of the desired message index k, each of the query vectors $Q_n^{[k]}, \forall n$ is individually comprised of elements that are i.i.d. uniform over $\{0, 1\}$. Thus, each database learns nothing about which message is requested.

We now show that database-privacy is preserved as well.

$$\begin{aligned} & I(W_{\overline{k}}; A_1^{[k]}, A_2^{[k]}, \cdots, A_N^{[k]}, Q_{1:N}^{[k]}, \mathcal{F}) && (20)\\ &= I(W_{\overline{k}}; A_1^{[k]}, A_1^{[k]} + x_{k,1}, \cdots, \\ &\qquad A_1^{[k]} + x_{k,N-1}, Q_{1:N}^{[k]}, \mathcal{F}) && (21)\\ &= I(W_{\overline{k}}; A_1^{[k]}, x_{k,1}, \cdots, x_{k,N-1}, Q_{1:N}^{[k]}, \mathcal{F}) && (22)\\ &= I(W_{\overline{k}}; A_1^{[k]}, W_k, Q_{1:N}^{[k]}, \mathcal{F}) && (23)\\ &\overset{(3)(4)}{=} I(W_{\overline{k}}; A_1^{[k]} | W_k, Q_{1:N}^{[k]}, \mathcal{F}) && (24)\\ &= 0 && (25) \end{aligned}$$

where in each step, the transformation on the variables is invertible such that mutual information remains the same. The last step follows from the independence of the messages and the common randomness (refer to (3)).

Note that because each answering string is 1 bit and the message is $L = N-1$ bits, the rate achieved is $(N-1)/N = 1 - 1/N$ which matches the capacity. Also note that only the minimum threshold amount of common randomness is utilized, i.e., $\rho = 1/(N-1)$. ■

2) Converse: Although Theorem 1 restricts to the setting where $l_k = 1, \forall k \in [1 : K]$, we do not assume this at the beginning in the proof of converse. This will make the converse general such that some of the intermediate steps can be used in the converse proofs of Theorem 2 and Theorem 3 as well.

For the converse we allow any feasible SPIR scheme, and prove that its rate cannot be larger than C_{SPIR}. Let us start with two lemmas that will be used later in the proof.

Lemma 1:

$$H(A_n^{[k]}|W_k, Q_n^{[k]}) = H(A_n^{[k']}|W_k, Q_n^{[k']}) \tag{26}$$

$$H(A_n^{[k]}|Q_n^{[k]}) = H(A_n^{[k']}|Q_n^{[k']}) \quad \forall n \in [1 : N], \quad k, k' \in [1 : K] \tag{27}$$

Proof: Since the proofs of (26) and (27) follow from the same arguments, here we will present only the proof of (26). From the User-Privacy constraint (6) we know that $\forall k \in [1 : K], \forall n \in [1 : N]$, $I(\theta; A_n^{[\theta]}, W_k, Q_n^{[\theta]}) = 0$. Therefore, we must have $\forall k' \in [1 : K]$,

$$H(A_n^{[k]}, W_k, Q_n^{[k]}) = H(A_n^{[k']}, W_k, Q_n^{[k']}) \tag{28}$$

$$H(W_k, Q_n^{[k]}) = H(W_k, Q_n^{[k']}) \tag{29}$$

Combining (28) and (29), we obtain $H(A_n^{[k]}|W_k, Q_n^{[k]}) = H(A_n^{[k']}|W_k, Q_n^{[k']})$. ■

Lemma 2:

$$H(A_n^{[k]}|W_k, \mathcal{F}, Q_n^{[k]}) = H(A_n^{[k]}|W_k, Q_n^{[k]}), \quad \forall n \in [1 : N] \tag{30}$$

Proof: Since

$$\begin{aligned} & H(A_n^{[k]}|W_k, Q_n^{[k]}) - H(A_n^{[k]}|W_k, \mathcal{F}, Q_n^{[k]}) \\ &\qquad = I(A_n^{[k]}; \mathcal{F}|W_k, Q_n^{[k]}) \geq 0, \end{aligned} \tag{31}$$

we only need to prove $I(A_n^{[k]}; \mathcal{F}|W_k, Q_n^{[k]}) \leq 0$.

$$\begin{aligned} & I(A_n^{[k]}; \mathcal{F}|W_k, Q_n^{[k]}) && (32)\\ &\leq I(A_n^{[k]}, W_1, \cdots, W_K, S; \mathcal{F}|W_k, Q_n^{[k]}) && (33)\\ &= I(W_1, \cdots, W_K, S; \mathcal{F}|W_k, Q_n^{[k]}) \\ &\quad + \underbrace{I(A_n^{[k]}; \mathcal{F}|W_1, \cdots, W_K, S, W_k, Q_n^{[k]})}_{=0} && (34)\\ &\leq I(W_1, \cdots, W_K, S; \mathcal{F}, Q_n^{[k]}) && (35)\\ &= 0 && (36) \end{aligned}$$

where the second term in (34) is zero because of (5) and (36) follows from (3), (4). ■

a) The proof for $R \leq C_{SPIR}$: For every feasible SPIR scheme, we must satisfy the database-privacy constraint (7),

$$0 = I(W_{\overline{k'}}; A_1^{[k']}, \cdots, A_N^{[k']}, Q_1^{[k']}, \cdots, Q_N^{[k']}, \mathcal{F}) \tag{37}$$

such that $\forall n \in [1 : N], \forall k \in [1 : K], k \neq k'$,

$$\begin{aligned} 0 &= I(W_k; A_n^{[k']}, Q_n^{[k']}) && (38)\\ &= H(A_n^{[k']}|Q_n^{[k']}) - H(A_n^{[k']}|W_k, Q_n^{[k']}) && (39)\\ &\overset{(26)}{=} H(A_n^{[k']}|Q_n^{[k']}) - H(A_n^{[k]}|W_k, Q_n^{[k]}) && (40) \end{aligned}$$

Now, consider the answering strings $A_1^{[k]}, \cdots, A_N^{[k]}$, from which we can decode W_k.

$$\begin{aligned} l_k L &= H(W_k) \overset{(3)}{=} H(W_k|\mathcal{F}) && (41)\\ &\overset{(10)}{=} I(W_k; A_1^{[k]}, \cdots, A_N^{[k]}|\mathcal{F}) + o(L) && (42)\\ &= H(A_1^{[k]}, \cdots, A_N^{[k]}|\mathcal{F}) - H(A_1^{[k]}, \cdots, A_N^{[k]}|W_k, \mathcal{F}) \\ &\quad + o(L) && (43)\\ &\overset{(4)}{\leq} H(A_1^{[k]}, \cdots, A_N^{[k]}|\mathcal{F}) - H(A_n^{[k]}|W_k, \mathcal{F}, Q_n^{[k]}) + o(L) && (44)\\ &\overset{(30)}{=} H(A_1^{[k]}, \cdots, A_N^{[k]}|\mathcal{F}) - H(A_n^{[k]}|W_k, Q_n^{[k]}) + o(L) && (45)\\ &\overset{(40)}{=} H(A_1^{[k]}, \cdots, A_N^{[k]}|\mathcal{F}) - H(A_n^{[k']}|Q_n^{[k']}) + o(L) && (46)\\ &\overset{(27)}{=} H(A_1^{[k]}, \cdots, A_N^{[k]}|\mathcal{F}) - H(A_n^{[k]}|Q_n^{[k]}) + o(L) && (47)\\ &\overset{(4)}{\leq} H(A_1^{[k]}, \cdots, A_N^{[k]}|\mathcal{F}) - H(A_n^{[k]}|\mathcal{F}) + o(L) && (48) \end{aligned}$$

Adding (48) for all $n \in [1:N]$, we have

$$\begin{aligned} N l_k L &\leq N H(A_1^{[k]}, \cdots, A_N^{[k]}|\mathcal{F}) - \sum_{n\in[1:N]} H(A_n^{[k]}|\mathcal{F}) \\ &\quad + o(L) && (49)\\ &\leq (N-1)\, H(A_1^{[k]}, \cdots, A_N^{[k]}|\mathcal{F}) + o(L) && (50)\\ &\leq (N-1) \sum_{n=1}^{N} H(A_n^{[k]}) + o(L) && (51)\\ &\leq (N-1)\, D + o(L) && (52)\\ R_k &= \frac{l_k L}{D} \leq 1 - \frac{1}{N} \quad \text{(Letting } L \to \infty) && (53) \end{aligned}$$

Thus, the rate of any feasible SPIR scheme cannot be more than C_{SPIR}.

Remark: The intuition behind the outer bound is as follows. Because of user-privacy, the answer from any individual database should be independent of the desired message index. Because of database-privacy, the answer from any individual database should contain no information about the non-desired messages. Combining these two facts, we know that the answer from any individual database should contain no information about any individual message (including the desired one). As a result, the useful information about the desired message can only come from the other $N-1$ databases. Therefore, each answer should contain at least $1/(N-1)$ of the entropy of the desired message and the outer bound follows.

b) The proof for $\rho \geq 1/(N-1)$: Suppose a feasible SPIR scheme exists that achieves a non-zero SPIR rate. Then we will show that it must have $\rho \geq 1/(N-1)$. Consider the answering strings $A_1^{[k]}, \cdots, A_N^{[k]}$, from which we can decode W_k. From the database-privacy constraint, we have

$$\begin{aligned} 0 &= I(W_{\bar{k}}; A_1^{[k]}, \cdots, A_N^{[k]}, \mathcal{F}) && (54)\\ &\overset{(3)}{=} I(W_{\bar{k}}; A_1^{[k]}, \cdots, A_N^{[k]}|\mathcal{F}) && (55)\\ &\overset{(10)}{=} I(W_{\bar{k}}; A_1^{[k]}, \cdots, A_N^{[k]}, W_k|\mathcal{F}) + o(L) && (56)\\ &\overset{(3)}{=} I(W_{\bar{k}}; A_1^{[k]}, \cdots, A_N^{[k]}|W_k, \mathcal{F}) + o(L) && (57)\\ &\geq I(W_{\bar{k}}; A_n^{[k]}|W_k, \mathcal{F}) + o(L) && (58)\\ &= H(A_n^{[k]}|W_k, \mathcal{F}) - H(A_n^{[k]}|W_1, \cdots, W_K, \mathcal{F}) \\ &\quad + o(L) && (59)\\ &\overset{(4)(5)}{=} H(A_n^{[k]}|W_k, \mathcal{F}) - H(A_n^{[k]}|W_1, \cdots, W_K, \mathcal{F}) \\ &\quad + H(A_n^{[k]}|W_1, \cdots, W_K, \mathcal{F}, S) + o(L) && (60)\\ &= H(A_n^{[k]}|W_k, \mathcal{F}) - I(S; A_n^{[k]}|W_1, \cdots, W_K, \mathcal{F}) \\ &\quad + o(L) && (61)\\ &\overset{(4)}{\geq} H(A_n^{[k]}|W_k, \mathcal{F}, Q_n^{[k]}) - H(S) + o(L) && (62)\\ &\overset{(30)}{=} H(A_n^{[k]}|W_k, Q_n^{[k]}) - H(S) + o(L) && (63)\\ &\overset{(40)}{=} H(A_n^{[k']}|Q_n^{[k']}) - H(S) + o(L) && (64)\\ &\overset{(27)}{=} H(A_n^{[k]}|Q_n^{[k]}) - H(S) + o(L) && (65) \end{aligned}$$

Adding (65) for $n \in [1:N]$, we have

$$\begin{aligned} 0 &\geq \sum_{n\in[1:N]} H(A_n^{[k]}|Q_n^{[k]}) - N H(S) + o(L) && (66)\\ &\geq H(A_1^{[k]}, \cdots, A_N^{[k]}|\mathcal{F}) - N H(S) + o(L) && (67)\\ &\overset{(50)}{\geq} \frac{N}{N-1} l_k L - N H(S) + o(L) && (68)\\ \Rightarrow H(S) &\geq \frac{1}{N-1} l_k L + o(L) && (69)\\ \Rightarrow \rho &= \frac{H(S)}{l_k L} \geq \frac{1}{N-1} \quad \text{(Letting } L \to \infty) && (70) \end{aligned}$$

Thus, the amount of common randomness relative to the message size of any feasible SPIR scheme cannot be less than $1/(N-1)$.

Remark: The intuition behind the bound on the amount of common randomness is as follows. From the proof for the rate of SPIR, we know that the average size of the answer from any individual database must be at least $1/(N-1)$ of the size of the desired message. Because of database-privacy, the answer from each individual database is independent of the messages. Therefore, to protect the answer, the amount of common randomness must be larger than or equal to the size of the answer, and the desired bound follows.

B. Proof of Theorem 4

1) Achievability: Without loss of generality, we assume that $l_1 \leq l_2 \leq \cdots \leq l_K$. Further, we append $(l_K - l_k)$ dummy message symbols to $W_k, \forall k \in [1:K]$ so that each message is made up of $l_K L$ symbols. We need to achieve the download cost of

$$D = \lceil \frac{L \max_i l_i}{1 - 1/N} \rceil = \lceil \frac{l_K L}{1 - 1/N} \rceil \tag{71}$$

with common randomness

$$\rho = \frac{\lceil L/(N-1) \rceil}{L}. \tag{72}$$

Here is such a scheme. Suppose $l_K L = G_1(N-1) + L_1$, where $G_1 = \lfloor l_K L/(N-1) \rfloor$ and $L_1 \in [0 : N-2]$. Note that the capacity achieving scheme for SPIR when $l_K L$ is not restricted is based on dividing the messages to blocks of length $N-1$ (refer to Theorem 1). The optimal scheme for finite $l_K L$ setting is constructed by first using the capacity achieving SPIR scheme G_1 times to retrieve $G_1(N-1)$ bits, and then

for the remaining L_1 bits, we use the capacity achieving SPIR schemes with only $L_1+1 \leq N$ databases (say, the first L_1+1 databases), if $L_1 \geq 1$. Otherwise if $L_1 = 0$, then we are done. Note that for the SPIR scheme that uses only L_1+1 databases, the rate is $1-1/(L_1+1)$, the message size is $L_1+1-1 = L_1$ bits, and the common randomness ratio is $\rho = 1/(L_1+1-1) = 1/L_1$. Therefore, overall, the download cost and the amount of common randomness are as follows.

$$D = \begin{cases} G_1 N, & \text{if } L_1 = 0, \\ G_1 N + L_1 + 1, & \text{otherwise.} \end{cases} \tag{73}$$

$$\rho = \begin{cases} \frac{G_1}{G_1(N-1)} = \frac{1}{N-1}, & \text{if } L_1 = 0, \\ \frac{G_1+1}{G_1(N-1)+L_1} = \frac{\lfloor l_K L/(N-1) \rfloor + 1}{l_K L}, & \text{otherwise.} \end{cases} \tag{74}$$

$$= \frac{\lceil L/(N-1) \rceil}{L}. \tag{75}$$

where (75) follows from the fact that when $l_K \geq 2$ (the case where $l_K = 1$ is immediate),

$$\frac{\lfloor l_K L/(N-1) \rfloor + 1}{l_K L} < \frac{l_K L/(N-1) + 1 + 1}{l_K L} \tag{76}$$

$$\leq \frac{l_K L/(N-1) + l_K}{l_K L} \tag{77}$$

$$= \frac{L/(N-1) + 1}{L} \tag{78}$$

and

$$\frac{\lfloor l_K L/(N-1) \rfloor + 1}{l_K L} > \frac{l_K L/(N-1)}{l_K L} = \frac{L/(N-1)}{L}. \tag{79}$$

Next, we prove that the download cost achieved in (73) matches that in (71), i.e., $\lceil \frac{l_K L}{1-1/N} \rceil$. When $L_1 = 0$ ($l_K L$ is an integer multiple of $N-1$), the claim follows trivially as

$$\lceil \frac{l_K L}{1-1/N} \rceil = N \frac{l_K L}{N-1} = G_1 N. \tag{80}$$

Hereafter, we consider $L_1 > 0$. It suffices to show that the download cost, $D = G_1 N + L_1 + 1$, satisfies $D \in [\frac{l_K L}{1-1/N}, \frac{l_K L}{1-1/N} + 1)$. In the converse proof to be presented in the next section, we will show that for arbitrary L, l_K and all SPIR schemes, $D \geq \frac{l_K L}{1-1/N}$ holds. So we are left to show that $D < \frac{l_K L}{1-1/N} + 1$.

$$D = G_1 N + L_1 + 1 \tag{81}$$

$$< \frac{G_1(N-1) + L_1}{1-1/N} + 1 \quad (N \geq 2) \tag{82}$$

$$= \frac{l_K L}{1-1/N} + 1 \tag{83}$$

Therefore the achievability proof is complete.

2) Converse: The converse proof follows closely from that of Theorem 1. Note that (52) holds for all l_k, all k and when we require exactly zero error, then we have

$$N \max_i l_i \leq (N-1) D/L \tag{84}$$

$$\Rightarrow D \geq \frac{N L \max_i l_i}{N-1} = \frac{L \max_i l_i}{1-1/N} \tag{85}$$

Note that since the downloads are assumed to be in terms of symbols from the same field as the message symbols, the download cost must be an integer value. From (85), we have

$$D \geq \lceil \frac{L \max_i l_i}{1-1/N} \rceil \tag{86}$$

$$\Rightarrow R_k = \frac{l_k L}{D} \leq l_k L / \lceil \frac{L \max_i l_i}{1-1/N} \rceil \tag{87}$$

Therefore the rate bound is proved. Next we proceed to the common randomness bound. From (70), which holds for all $k \in [1:K]$, we have

$$\rho L \geq L/(N-1) \tag{88}$$

Similar to the download cost, which is restricted to be integers, in the finite length regime the amount of common randomness, ρL is restricted to take integer values as well. Therefore, from (88), we have $\rho L \geq \lceil L/(N-1) \rceil$.

V. Conclusion

For K messages and N databases, the capacity of SPIR was shown to be $C = 1 - 1/N$. In order to achieve any positive rate for SPIR, the minimum amount of common randomness needed among the databases was shown to be $1/(N-1)$ bits per message bit. Remarkably, this is also sufficient to achieve the capacity of SPIR. The insights extend to settings with unequal message sizes and finite length messages.

References

[1] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," in *Proc. IEEE 36th Annu. Found. Comput. Sci.*, Oct. 1995, pp. 41–50.

[2] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private information retrieval," *J. ACM*, vol. 45, no. 6, pp. 965–981, 1998.

[3] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4075–4088, Jul. 2017.

[4] Y. Gertner, Y. Ishai, E. Kushilevitz, and T. Malkin, "Protecting data privacy in private information retrieval schemes," in *Proc. 30th Annu. ACM Symp. Theory Comput.*, 1998, pp. 151–160.

[5] W. Gasarch, "A survey on private information retrieval," *Bull. EATCS*, vol. 82, p. 72–107, Feb. 2004.

[6] S. Yekhanin, "Private information retrieval," *Commun. ACM*, vol. 53, no. 4, pp. 68–73, 2010.

[7] J. Katz and L. Trevisan, "On the efficiency of local decoding procedures for error-correcting codes," in *Proc. 32nd Annu. ACM Symp. Theory Comput.*, 2000, pp. 80–86.

[8] S. Yekhanin, "Locally decodable codes and private information retrieval schemes," Ph.D. dissertation, Dept. EECS, MIT, Cambridge, MA, USA, 2007.

[9] Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai, "Batch codes and their applications," in *Proc. 36th Annu. ACM Symp. Theory Comput.*, 2004, pp. 262–271.

[10] Y. Ishai and E. Kushilevitz, "On the hardness of information-theoretic multiparty computation," in *Advances in Cryptology—EUROCRYPT*. Berlin, Germany: Springer-Verlag, 2004, pp. 439–455.

[11] M. O. Rabin, "How to exchange secrets with oblivious transfer," Aiken Comput. Lab., Harvard Univ., Cambridge, MA, USA, Tech. Rep. TR-81, 1981.

[12] S. Even, O. Goldreich, and A. Lempel, "A randomized protocol for signing contracts," *Commun. ACM*, vol. 28, no. 6, pp. 637–647, 1985.

[13] J. Kilian, "Founding cryptography on oblivious transfer—Efficiently," in *Proc. 20th Annu. ACM Symp. Theory Comput.*, 1988, pp. 20–31.

[14] Y. Ishai, M. Prabhakaran, and A. Sahai, "Founding cryptography on oblivious transfer—Efficiently," in *Proc. Annu. Int. Cryptol. Conf.* Berlin, Germany: Springer, 2008, pp. 572–591.

[15] R. Ahlswede and I. Csiszár, "On oblivious transfer capacity," in *Information Theory, Combinatorics, and Search Theory*. Berlin, Germany: Springer, 2013, pp. 145–166.
[16] A. C. A. Nascimento and A. Winter, "On the oblivious-transfer capacity of noisy resources," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2572–2581, Jun. 2008.
[17] H. Sun and S. A. Jafar, "The capacity of robust private information retrieval with colluding databases," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2361–2370, Apr. 2018.
[18] R. Tajeddine, O. W. Gnilke, D. Karpuk, R. Freij-Hollanti, C. Hollanti, and S. E. Rouayheb, "Private information retrieval schemes for coded data with arbitrary collusion patterns," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 1908–1912.
[19] Z. Jia, H. Sun, and S. A. Jafar, "The capacity of private information retrieval with disjoint colluding sets," in *Proc. IEEE GLOBECOM*, Dec. 2017, pp. 1–6.
[20] H. Sun and S. A. Jafar, "Optimal download cost of private information retrieval for arbitrary message length," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 12, pp. 2920–2932, Dec. 2017.
[21] H. Sun and S. A. Jafar, "Multiround private information retrieval: Capacity and storage overhead," *IEEE Trans. Inf. Theory*, to be published, doi: 10.1109/TIT.2018.2789426.
[22] K. Banawan and S. Ulukus, "Multi-message private information retrieval: Capacity results and near-optimal schemes," *IEEE Trans. Inf. Theory*, to be published, doi: 10.1109/TIT.2018.2828310.
[23] R. Tajeddine, O. W. Gnilke, and S. E. Rouayheb, "Private information retrieval from mds coded data in distributed storage systems," *IEEE Trans. Inf. Theory*, to be published, doi: 10.1109/TIT.2018.2815607.
[24] K. Banawan and S. Ulukus, "The capacity of private information retrieval from coded databases," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1945–1956, Mar. 2018.
[25] H.-Y. Lin, S. Kumar, E. Rosnes, and A. G. I. Amat. (2018). "An MDS-PIR capacity-achieving protocol for distributed storage using non-MDS linear codes." [Online]. Available: https://arxiv.org/abs/1801.04923
[26] R. Freij-Hollanti, O. W. Gnilke, C. Hollanti, and D. A. Karpuk, "Private information retrieval from coded databases with colluding servers," *SIAM J. Appl. Algebra Geometry*, vol. 1, no. 1, pp. 647–664, 2017.
[27] 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 Trans. Inf. Theory*, vol. 64, no. 2, pp. 1000–1022, Feb. 2018.
[28] Y. Zhang and G. Ge. (2017). "A general private information retrieval scheme for MDS coded databases with colluding servers." [Online]. Available: https://arxiv.org/abs/1704.06785
[29] R. Tandon. (2017). "The capacity of cache aided private information retrieval." [Online]. Available: https://arxiv.org/abs/1706.07035
[30] S. Kadhe, B. Garcia, A. Heidarzadeh, S. E. Rouayheb, and A. Sprintson. (2017). "Private information retrieval with side information." [Online]. Available: https://arxiv.org/abs/1709.00112
[31] Y.-P. Wei, K. Banawan, and S. Ulukus. (2017). "Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching." [Online]. Available: https://arxiv.org/abs/1709.01056
[32] Z. Chen, Z. Wang, and S. Jafar. (2017). "The capacity of private information retrieval with private side information." [Online]. Available: https://arxiv.org/abs/1709.03022
[33] Y.-P. Wei, K. Banawan, and S. Ulukus. (2017). "The capacity of private information retrieval with partially known private side information." [Online]. Available: https://arxiv.org/abs/1710.00809
[34] K. Banawan and S. Ulukus. (2017). "The capacity of private information retrieval from byzantine and colluding databases." [Online]. Available: https://arxiv.org/abs/1706.01442
[35] Q. Wang and M. Skoglund. (2017). "Secure private information retrieval from colluding databases with eavesdroppers." [Online]. Available: https://arxiv.org/abs/1710.01190
[36] H. Sun and S. A. Jafar. (2017). "The capacity of private computation." [Online]. Available: https://arxiv.org/abs/1710.11098
[37] M. Mirmohseni and M. A. Maddah-Ali. (2017). "Private function retrieval." [Online]. Available: https://arxiv.org/abs/1711.04677
[38] Z. Chen, Z. Wang, and S. Jafar. (2018). "The asymptotic capacity of private search." [Online]. Available: https://arxiv.org/abs/1801.05768
[39] D. Karpuk. (2018). "Private computation of systematically encoded data with colluding servers." [Online]. Available: https://arxiv.org/abs/1801.02194
[40] J. Xu and Z. Zhang. (2018). "Building capacity-achieving pir schemes with optimal sub-packetization over small fields." [Online]. Available: https://arxiv.org/abs/1801.02324
[41] K. Banawan and S. Ulukus. (2017). "Asymmetry hurts: Private information retrieval under asymmetric traffic constraints." [Online]. Available: https://arxiv.org/abs/1801.03079
[42] Q. Wang and M. Skoglund, "Symmetric private information retrieval for MDS coded distributed storage," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.
[43] Q. Wang and M. Skoglund. (2017). "Secure symmetric private information retrieval from colluding databases with adversaries." [Online]. Available: https://arxiv.org/abs/1707.02152

Hua Sun (S'12–M'17) received his B.E. in Communications Engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2011, M.S. in Electrical and Computer Engineering from University of California Irvine, USA, in 2013, and Ph.D. in Electrical Engineering from University of California Irvine, USA, in 2017. He is an Assistant Professor in the Department of Electrical Engineering at the University of North Texas, USA. His research interests include information theory and its applications to communications, privacy, networking, and storage.

Dr. Sun received the IEEE Jack Keil Wolf ISIT Student Paper Award in 2016, an IEEE GLOBECOM Best Paper Award in 2016, and the University of California Irvine CPCC Fellowship for the year 2011–2012.

Syed Ali Jafar (S'99–M'04–SM'09–F'14) received his B. Tech. from IIT Delhi, India, in 1997, M.S. from Caltech, USA, in 1999, and Ph.D. from Stanford, USA, in 2003, all in Electrical Engineering. His industry experience includes positions at Lucent Bell Labs and Qualcomm. He is a Professor in the Department of Electrical Engineering and Computer Science at the University of California Irvine, Irvine, CA USA. His research interests include multiuser information theory, wireless communications and network coding.

Dr. Jafar is a recipient of the New York Academy of Sciences Blavatnik National Laureate for *Physical Sciences and Engineering*, the NSF CAREER Award, the ONR Young Investigator Award, the UCI Academic Senate Distinguished Mid-Career Faculty Award for Research, the School of Engineering Mid-Career Excellence in Research Award and the School of Engineering Maseeh Outstanding Research Award. His co-authored papers have received the IEEE Information Theory Society Best Paper Award, IEEE Communications Society Best Tutorial Paper Award, IEEE Communications Society Heinrich Hertz Award, IEEE Signal Processing Society Young Author Best Paper Award, IEEE Information Theory Society Jack Wolf ISIT Best Student Paper Award, and three IEEE GLOBECOM Best Paper Awards. Dr. Jafar received the UC Irvine EECS Professor of the Year award six times, in 2006, 2009, 2011, 2012, 2014 and 2017 from the Engineering Students Council, a School of Engineering Teaching Excellence Award in 2012, and a Senior Career Innovation in Teaching Award in 2018. He was a University of Canterbury Erskine Fellow in 2010 and an IEEE Communications Society Distinguished Lecturer for 2013–2014. Dr. Jafar was recognized as a Thomson Reuters/Clarivate Analytics Highly Cited Researcher and included by Sciencewatch among The World's Most Influential Scientific Minds in 2014, 2015, 2016 and 2017. He served as Associate Editor for IEEE TRANSACTIONS ON COMMUNICATIONS 2004–2009, for IEEE COMMUNICATIONS LETTERS 2008–2009 and for IEEE TRANSACTIONS ON INFORMATION THEORY 2009–2012.