Private Information Retrieval with Side Information

Swanand Kadhe, *Member, IEEE,* Brenden Garcia, Anoosheh Heidarzadeh, *Member, IEEE,* Salim El Rouayheb, *Member, IEEE,* and Alex Sprintson, *Senior Member, IEEE*

***Abstract*—We study the problem of Private Information Retrieval (PIR) in the presence of prior side information. The problem setup includes a database of K independent messages possibly replicated on several servers, and a user that needs to retrieve one of these messages. In addition, the user has some prior side information in the form of a subset of M messages, not containing the desired message and unknown to the servers. This problem is motivated by practical settings in which the user can obtain side information opportunistically from other users or has previously downloaded some messages using classical PIR schemes.**

The objective of the user is to retrieve the required message with downloading minimum amount of data from the servers while achieving information-theoretic privacy in one of the following two scenarios: (i) the user wants to protect jointly the identities of the demand and the side information; (ii) the user wants to protect only the identity of the demand, but not necessarily the side information. To highlight the role of side information, we focus first on the case of a single server (single database). In the first scenario, we prove that the minimum download cost is $K-M$ messages, and in the second scenario it is $\lceil K/(M+1)\rceil$ messages, which should be compared to K messages—the minimum download cost in the case of no side information. Then, we extend some of our results to the case of the database replicated on multiple servers. Our proof techniques relate PIR with side information to the index coding problem. We leverage this connection to prove converse results, as well as to design achievability schemes.

***Index Terms*—Private information retrieval, information-theoretic privacy, index coding**

I. Introduction

Consider the following Private Information Retrieval (PIR) setting first studied in [1], [2]: a user wishes to privately download a message belonging to a database with copies stored on a single or multiple remote servers, without revealing which message it is requesting. In a straightforward PIR scheme, the user would download all the messages in the database. This scheme may not be feasible due to its high communication cost. In the case of a single server (i.e., there is only one copy of the database), it can be shown that downloading the whole database is necessary to achieve perfect privacy in an information-theoretic sense [1]. If computational (cryptographic) privacy is desired, then PIR schemes with lower communication overhead do exist [3], [4], but they do not offer information-theoretic privacy guarantees and usually have high computational complexity. In contrast, in this paper, we design and analyze schemes that achieve information-theoretic privacy.

Interestingly, more efficient PIR schemes achieving perfect information-theoretic privacy can be constructed when the database is replicated on multiple servers with restriction on the servers' collusion. This replication-based model has been the one that is predominantly studied in the PIR literature (e.g., [5]–[8]) with breakthrough results in the past few years (e.g., [9]–[13]). Recently, there has been a renewed interest in PIR for the case in which the data is stored on the servers using erasure codes, which result in better storage overhead compared to the traditional replication techniques [12]–[19].

In this paper, we study the PIR problem when the user has prior side information about the database. In particular, we assume that the user already has a random subset of the database messages that is unknown to the server(s)[1]. This side information could have been obtained in several ways. For example, the user could have obtained these messages opportunistically from other users in its network, overheard them from a wireless broadcast channel, or downloaded them previously through classical PIR schemes. The next example illustrates how this side information could be leveraged to devise efficient PIR with side information (PIR-SI) schemes. In particular, the following example shows that perfect information-theoretic privacy can be achieved in the single server case without having to download the entire database.

Example 1 (single-server PIR with side information)**.** *Consider a remote server that has a database formed of an even number of binary messages denoted by $X_1,\dots,X_K$ of equal length. A user wants to download one of these messages from the server without revealing to the server which one. Moreover, the user has one message as side information chosen uniformly at random among all the other messages and unknown to the server. We propose two PIR-SI schemes that leverage the side*

This paper was presented in part at the 55th Annual Allerton Conference on Communication, Control, and Computing, Oct. 2017.

Swanand Kadhe is with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: swanand.kadhe@berkeley.edu). Part of this work was done while he was with the ECE department at Texas A&M University.

Brenden Garcia is with Epic Systems Corporation. Part of this work was done when he was was with the ECE department at Texas A&M University (e-mail: brendengarcia@tamu.edu).

Anoosheh Heidarzadeh and Alex Sprintson are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mails: {anoosheh, spalex}@tamu.edu).

Salim El Rouayheb is with the Department of Electrical and Computer Engineering, Rutgers University—The State University of New Jersey, New-Brunswick, NJ 08901-8554 USA (e-mail: salim.elrouayheb@rutgers.edu).

The work of S. Kadhe was supported in part by NSF grants CCF-1748692 and TWC-1748585.

The work of S. El Rouayheb was supported in part by NSF Grant CCF 1652867 and ARL Grant W911NF-17-1-0032.

The work of B. Garcia, A. Heidarzadeh, S. Kadhe, and A. Sprintson was supported by the National Science Foundation under Grants No. 1718658 and 1642983.

Communicated by Vinod M. Prabhakaran, Associate Editor for Shannon Theory.

[1] We assume that this side information subset does not contain the desired message. Otherwise, the problem is degenerate.

information and compare them to the straightforward scheme that downloads all the K messages.

1) Maximum Distance Separable (MDS) PIR-SI scheme. *The number of bits downloaded by this scheme is equivalent to $K-1$ messages. The user sends to the server the number of messages in their side information (one in this example). The server responds by coding all the messages using a $(2K-1, K)$ systematic MDS code and sending the $K-1$ parity symbols of the code. Therefore, the user can always decode all the messages using their side information and the coded messages received from the server. Perfect privacy is achieved here, since the protocol is independent of the index of the desired message (as well as the index of the side information).*
2) Partition and Code PIR-SI scheme. *The number of bits downloaded by this scheme is equivalent to $K/2$ messages. Suppose the message the user wants is X_W and the message in the user's side information is X_S for some $W, S \in \{1, \dots, K\}$, $W \neq S$. The user chooses a random partition of $\{1, \dots, K\}$ formed only of sets of size two and containing $\{W, S\}$, and sends indices of all pairs in the partition to the server. The server sends back the XOR of the messages indexed by each subset. Using this scheme, the user can always decode X_W because it always receives $X_W + X_S$. Intuitively, perfect privacy is achieved here because the index of the desired message can be in any subset of the partition, and in each subset it could be either one of the messages in the subset, since the server does not know the index of the side information.* ■

We will show later that the two schemes above are optimal, but achieve different privacy constraints. The MDS PIR-SI scheme protects both the indices of the desired message and that of the side information, whereas the Partition and Code PIR-SI scheme is designed to protect only the former.[2]

A. Our Contributions

We consider the PIR with side information (PIR-SI) problem as illustrated in Example 1. A user wishes to download a message from a set of K messages that belong to a database stored on a single remote server or replicated on several *non-colluding* servers. The user has a random subset of M messages as side information. The identity of the messages in this subset is unknown to the server. We focus on PIR-SI schemes that achieve information-theoretic privacy. The figure of merit that we consider for the PIR-SI schemes is the download rate, which dominates the total communication rate (download plus upload) for large message sizes. Under this setting, we distinguish between two types of privacy constraints:

(i) hiding both the identity of the requested message and that of the side information from the server; and
(ii) hiding only the identity of the desired message.

The latter, and less stringent, privacy constraint is justified when the side information is obtained opportunistically given that it is random and assumed to be independent of the user's request. In the case in which the side information messages were obtained previously through PIR, this constraint implies that the identity of these messages may be leaked to the server(s). However, this type of privacy can still be relevant when privacy is only desired for a certain duration of time, i.e., when the user is not concerned about protecting the identity of messages downloaded as long as it has happened far enough in the past.

First, we focus on the single server scenario as the canonical case to understand the role of side information in PIR. We characterize the capacity of PIR-SI problem in the case of a single server for the two privacy constraints mentioned above. We show that when protecting the request and the side information jointly, the maximum download rate[3] is $(K-M)^{-1}$, and this can be achieved by a generalization of the MDS PIR-SI scheme in Example 1. Moreover, we show that when protecting only the request, the maximum download rate is $\lceil K/(M+1) \rceil^{-1}$, and this can be achieved by a generalization of the Partition and Code PIR-SI scheme in Example 1. We present achievability and converse proofs that use, among other things, connections to index coding.

Second, we tackle the case of $N > 1$ servers storing replicas of the database. In this case, we devise a PIR-SI scheme that achieves a download rate equal to

$$\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{\lceil \frac{K}{M+1} \rceil - 1}}\right)^{-1}.$$

Our scheme for the multiple servers uses ideas from our single server scheme in conjunction with the scheme due to Sun and Jafar [9] for the setting with no side information.

B. Related Work

Single-server PIR: Our initial motivation for this work grew from the need to construct single-server, computationally efficient PIR schemes with efficient download rate. Before this work, single-server PIR has been predominantly studied in the computational privacy setting. In particular, the authors of [3] presented a single-server computational PIR scheme that does not require downloading the entire database. This result was extended in several works to further reduce the communication costs, see e.g., [4], [20], [21]. However, computational PIR schemes typically suffer from heavy computational cost, and require homomorphic encryption [22].

On the other hand, for achieving information-theoretic privacy in the single-server case, it is well-known that the user has to download the entire database to hide which message they are interested in (see [1]). The key contribution of this paper is to demonstrate that having access to some side information (that is unknown to the server) enables the user to achieve

[2] It is worth noting that, in the above toy example, the Partition and Code scheme also protects the side information individually, but it does not protect the desired message index and the side information index jointly. In general, the Partition and Code scheme is guaranteed to protect only the desired message index and may leak some information about user's side information (see Remark 3 in Sec. IV-C).

[3] The download rate is defined as the inverse of the normalized download cost.

information-theoretic privacy without needing to download the entire database.

Multi-server PIR: The initial work on PIR in [1], [2] and in the literature that followed focused on designing PIR schemes for replicated data that have efficient communication cost accounting for both the size of the user queries and the servers' responses. PIR schemes with communication cost that is subpolynomial in the number of messages were devised in [7] and [23]. Information-theoretic bounds on the download rate (servers' responses) and achievable schemes were devised in [9] and [10].

Several recent works on the so-called *cache-aided PIR* [24]–[26] are conceptually related to our work in considering the role of side information in PIR. The main difference is that in the cache-aided PIR problem the user can choose the side information, which the servers may or may not know. On the contrary, in our setting, the user is given a certain (random) subset of messages as a side information, which is unknown to the servers. In [24], the capacity is derived for the case when the side information can be any function of the database, and is known to all the servers. In [25], the authors present capacity results when the servers are partially aware of the uncoded side information. In [26], the authors present capacity results when the side information is uncoded and not known to the servers.

It is worth mentioning that, following the initial version of this paper in [27] and [28], references [29], [30] generalized the capacity result from single server to multiple non-colluding servers (storing replicas of the database). In particular, following up the privacy model where the user wants to protect both the requested message(s) and the side information messages, the work of [29] characterized the capacity for the multi-server single-message case, whereas [30] characterized the capacity for multi-server multi-message case for certain regimes. The multi-message version of our problem is considered in [31], [32], and the case of coded side information is considered in [33]. The work of [34] builds up on the Partition and Code scheme to design a computationally efficient single-server PIR scheme with computational privacy guarantees.

Privacy in broadcasting: Another related line of work is that of private broadcasting in [35], which considers the index coding setting with multiple users with side information and a single server. In this setup, the server *does know* the content of the side information at the users. The privacy constraint is to protect the request and side information of a user from the other users through a carefully designed encoding matrix. In contrast, the goal of our scheme is to protect the identity of the requested data from the server. We note that the case in which the side information is unknown at the server is also considered in the index coding literature under the name of blind index coding [36]. However, the goal there is to minimize the broadcast rate without privacy constraints.

II. Problem Formulation

For a positive integer K, denote $\{1,\dots,K\}$ by $[K]$. For a set $\{X_1,\dots,X_K\}$ and a subset $S \subseteq [K]$, let $X_S = \{X_j : j \in S\}$. For a subset $S \subseteq [K]$, let $\mathbf{1}_S$ denote the characteristic vector of the set S, which is a binary vector of length K such that, for all $j \in [K]$, its j-th entry is 1 if $j \in S$, otherwise it is 0. Let $\mathbb{F}_q$ denote the finite field of order q. We denote a random variable with a bold symbol, e.g., $\mathbf{X}$, and its realization without bold face, e.g., X.

We assume that the database consists of a set of K messages $\mathbf{X}_{[K]} = \{\mathbf{X}_1,\dots,\mathbf{X}_K\}$, with each message being independently and uniformly distributed over $\mathbb{F}_{2^t}$ (i.e., each message X_j is t bits long). We also assume that there are $N \geq 1$ non-colluding servers, which store identical copies of the K messages.

A user is interested in downloading a message $\mathbf{X}_W$ for some $W \in [K]$. We refer to W as the *demand index* and $\mathbf{X}_W$ as the *demand*. The user has the knowledge of a subset $\mathbf{X}_S$ of the messages for some $S \subseteq [K] \setminus \{W\}$, $|S|= M$, $M < K$. We refer to S as the *side information index set* and $\mathbf{X}_S$ as the *side information*.

Let $\mathbf{W}$ and $\mathbf{S}$ denote the random variables corresponding to the demand index and the side information index set, respectively. We restrict our attention to the class of distributions $p_{\mathbf{W}}(\cdot)$ of $\mathbf{W}$ such that $p_{\mathbf{W}}(W) > 0$ for every $W \in [K]$.

An important distribution of $\mathbf{W}$ and $\mathbf{S}$ that we focus on in this work is as follows. We assume that the side information index set $\mathbf{S}$ is distributed uniformly over over all subsets of $[K]$ of size M, i.e.,

$$p_{\mathbf{S}}(S) = \begin{cases} \frac{1}{\binom{K}{M}}, & \text{if } S \subset [K],\ |S|= M, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Further, we assume that the demand index set $\mathbf{W}$ has the following conditional distribution given the side information index set S:

$$p_{\mathbf{W}|\mathbf{S}}(W \mid S) = \begin{cases} \frac{1}{K-M}, & \text{if } W \in [K] \setminus S \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

We note that this implies the following joint distribution on $(\mathbf{W},\mathbf{S})$:

$$p_{\mathbf{W},\mathbf{S}}(W,S) = \begin{cases} \frac{1}{(K-M)\binom{K}{M}}, & \text{if } W \notin S, |S|= M, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Marginalizing (3) over $\mathbf{S}$, it is easy to see that the demand index $\mathbf{W}$ is distributed uniformly over $[K]$, i.e.,

$$p_{\mathbf{W}}(W) = \frac{1}{K}, \quad \forall\, W \in [K]. \tag{4}$$

We assume that the servers do not know the realization of the user's side information S, but they know the number of side information messages M and the *a priori* distributions $p_{\mathbf{S}}(S)$ and $p_{\mathbf{W}|\mathbf{S}}(W|S)$.

To download the message $\mathbf{X}_W$ given the side information $\mathbf{X}_S$, the user sends a query $Q_j^{[W,S]}$ from a finite alphabet $\mathcal{Q}$ to the j-th server. The j-th server responds to the query it receives with an answer $A_j^{[W,S]}$ over a finite alphabet $\mathcal{A}$. Let $\mathbf{Q}_j^{[W,S]}$ and $\mathbf{A}_j^{[W,S]}$ be the corresponding random variables. We refer to the set of queries and answers as the *PIR with side information (PIR-SI) scheme*. Our focus in this paper is on non-interactive (single round) schemes. Further, we assume that the servers do not collude with each other.

A PIR-SI scheme should satisfy the following requirements.

1. For every $j \in [N]$, the query $\mathbf{Q}_j^{[W,S]}$ to the server j is a (potentially stochastic) function of W and S[4]. We assume that the answer from the server is a deterministic function of the query and the messages, i.e., for all $j \in [N]$, $(\mathbf{W}, \mathbf{S}) \leftrightarrow (\mathbf{Q}_j^{[\mathbf{W},\mathbf{S}]}, \mathbf{X}_{[K]}) \leftrightarrow \mathbf{A}_j^{[\mathbf{W},\mathbf{S}]}$ forms a Markov chain, and
$$H\left(\mathbf{A}_j^{[\mathbf{W},\mathbf{S}]} \mid \mathbf{Q}_j^{[\mathbf{W},\mathbf{S}]}, \mathbf{X}_{[K]}, \mathbf{W}, \mathbf{S}\right) = 0. \qquad (5)$$
2. From the answers $A_1^{[W,S]}, \ldots, A_N^{[W,S]}$, the queries $Q_1^{[W,S]}, \ldots, Q_N^{[W,S]}$, and the side information X_S, the user should be able to decode the desired message X_W, i.e.,
$$H\Big(\mathbf{X}_\mathbf{W} \mid \mathbf{A}_1^{[\mathbf{W},\mathbf{S}]}, \cdots, \mathbf{A}_N^{[\mathbf{W},\mathbf{S}]}, \mathbf{Q}_1^{[\mathbf{W},\mathbf{S}]}, \cdots, \mathbf{Q}_N^{[\mathbf{W},\mathbf{S}]}, \mathbf{X}_\mathbf{S}, \mathbf{W}, \mathbf{S}\Big) = 0. \qquad (6)$$
3. The PIR-SI scheme should guarantee privacy for the user by ensuring one of the following two conditions, referred to as W-privacy and (W, S)-privacy, as defined below.

   Definition 1. *W-**privacy**: No server can infer any information about the demand index from the query, answer, and messages, i.e., for all $j \in [N]$, we have*
$$I\left(\mathbf{W} ; \mathbf{Q}_j^{[\mathbf{W},\mathbf{S}]}, \mathbf{A}_j^{[\mathbf{W},\mathbf{S}]}, \mathbf{X}_{[K]}\right) = 0. \qquad (7)$$

   Definition 2. *(W, S)-**privacy**: No server can infer any information about the demand index as well as the side information index set from the query, answer, and messages, i.e., for all $j \in [N]$, we have*
$$I\left(\mathbf{W}, \mathbf{S} ; \mathbf{Q}_j^{[\mathbf{W},\mathbf{S}]}, \mathbf{A}_j^{[\mathbf{W},\mathbf{S}]}, \mathbf{X}_{[K]}\right) = 0. \qquad (8)$$

   We refer to a PIR-SI scheme preserving W-privacy or (W, S)-privacy as W-PIR-SI or (W, S)-PIR-SI scheme, respectively.

The *rate* of a PIR-SI scheme is defined as the ratio of the message length (t bits) to the total length of the answers (in bits) as follows[5]:
$$R = \frac{t}{\sum_{j=1}^{N} H\left(\mathbf{A}_j^{[\mathbf{W},\mathbf{S}]}\right)}. \qquad (9)$$

The *capacity* of W-PIR-SI or (W, S)-PIR-SI problem, respectively denoted by C_W or $C_{W,S}$, is defined as the supremum of rates over all W-PIR-SI or (W, S)-PIR-SI schemes for a given N, K, and M, respectively.

Remark 1. *Note that, in general, queries can be a (potentially stochastic) function of the demand index W, the side information index set S,* as well as *the side information X_S. However, throughout the paper, we assume that queries depend only on W and S, and are independent of the message values X_S. Characterizing the capacities of W-PIR-SI and (W, S)-PIR-SI when queries also depend on X_S is left as a future work.*

III. Main Results

First, we summarize our main results for the single server case in Theorems 1 and 2, which characterize the capacity of W-PIR-SI and (W, S)-PIR-SI problems, respectively.

Theorem 1. *For the W-PIR-SI problem with $N = 1$ server, K messages, and side information size M, when the demand index $\mathbf{W}$ and the side information index set $\mathbf{S}$ are jointly distributed according to* (3)*, the capacity is*
$$C_W = \left\lceil \frac{K}{M+1} \right\rceil^{-1}. \qquad (10)$$

Our proof for Theorem 1 is based on two parts. We prove the converse in Section IV-B for any joint distribution of $(\mathbf{W}, \mathbf{S})$. Then, we construct an achievability scheme in Section IV-C for the distribution given in (3).

Theorem 2. *For the (W, S)-PIR-SI problem with $N = 1$ server storing K messages and for any arbitrary joint distribution of the demand index $\mathbf{W}$ and the side information index set $\mathbf{S}$ such that the size of $\mathbf{S}$ is equal to M, the capacity is*
$$C_{W,S} = (K - M)^{-1}. \qquad (11)$$

We begin by showing that the capacity $C_{W,S}$ of the (W, S)-PIR-SI problem with $N = 1$ server, K messages, and size information size M is upper bounded by $(K - M)^{-1}$ for any joint distribution of $(\mathbf{W}, \mathbf{S})$ (see Section V-A). Then, we construct a scheme based on maximum distance separable (MDS) codes, which achieves this bound (see Section V-B).

Second, we state our main result for multiple servers storing replicas of the database, which gives a lower bound on the capacity of W-PIR-SI problem based on an achievability scheme.

Theorem 3. *For the W-PIR-SI problem with N servers, each storing K messages, and side information size M, when the demand index $\mathbf{W}$ and the side information index set $\mathbf{S}$ are jointly distributed according to* (3)*, the capacity is lower bounded as*
$$C_W \geq \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{\lceil \frac{K}{M+1} \rceil - 1}}\right)^{-1}. \qquad (12)$$

Our proposed PIR scheme builds up on the scheme in [9], which is for the case of no side-information.

Remark 2. *The capacity of (W, S)-PIR-SI problem with N servers, each storing K messages, was characterized in [29] as $(1 + 1/N + \cdots + 1/N^{K-M-1})^{-1}$. Note that this is identical to the capacity of the PIR problem with N servers, each storing $K - M$ messages, and no side information (characterized in [9]). In other words, in the (W, S)-PIR problem, the side information effectively reduces the size of the database by M messages (i.e., from K to $K - M$ messages). On the other hand, our proposed achievability scheme shows that in the W-PIR-SI problem, the side information reduces the size of the database by a factor of $1/(M+1)$ (i.e., from K to $\lceil K/(M+1) \rceil$ messages). Whether this rate is optimal is an open question.*

[4] We assume that the queries are independent of the message values X_S. See Remark 1 for details.

[5] Note that the download rate dominates the total communication rate for large enough messages.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2019.2948845, IEEE Transactions on Information Theory

IV. Single-Server W-PIR-SI Problem

Our converse proofs for Theorems 1 and 2 in the single-server case use the following simple yet powerful observation[6].

Proposition 1. *For a W-PIR-SI scheme, given a demand index W, a side information index set S, and a query $Q^{[W,S]}$, the following two conditions hold:*

1) *For each $i \in [K] \setminus \{W\}$, there exist a subset $S_i \subset [K] \setminus \{i\}$ with $|S_i|= M$ and a decoding function $D_i^{Q^{[W,S]}}$ such that, for every message set realization $\{X_1, \dots, X_K\}$, the corresponding answer $A^{[W,S]}$ satisfies $D_i^{Q^{[W,S]}}\left(A^{[W,S]}, X_{S_i}\right) = X_i$.*
2) *There exists a decoding function $D_W^{Q^{[W,S]}}$ such that, for every message set realization $\{X_1, \dots, X_K\}$, the corresponding answer $A^{[W,S]}$ satisfies $D_W^{Q^{[W,S]}}\left(A^{[W,S]}, X_S\right) = X_W$.*

Proof. The result follows from the decodability requirement (6) and the W-privacy requirement (7)[7]. In particular, the second condition is implied by the decodability requirement (6). Further, observe that the two conditions together imply that, for each $i \in [K]$, there exist a subset $S_i \subset [K] \setminus \{i\}$ with $|S_i|= M$ and a decoding function $D_i^{Q^{[W,S]}}$ satisfying $D_i^{Q^{[W,S]}}\left(A^{[W,S]}, X_{S_i}\right) = X_i$. This is necessary to ensure the W-privacy requirement (7). Indeed, if this was not the case for some i, then the server would know that the i-th message is not requested by the user. Therefore, it holds that

$$\mathbb{P}\left(\mathbf{W} = i \mid \mathbf{Q}^{[\mathbf{W},\mathbf{S}]} = Q^{[W,S]}\right) = 0, \tag{13}$$

which, in turn, implies that $I\left(\mathbf{W}; \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}, \mathbf{A}^{[\mathbf{W},\mathbf{S}]}, \mathbf{X}_{[K]}\right) > 0$. This violates the W-privacy condition (7). It is worth noting that the result holds under the assumption that $\mathbf{W}$ has a distribution such that $p_{\mathbf{W}}(W) > 0$ for each $W \in [K]$. ■

The above proposition enables us to show a relation of the single-server PIR-SI problem with an instance of index coding with side information problem [37]–[40]. We begin with briefly reviewing the index coding problem. We refer the reader to [41] for an excellent recent survey.

A. Index Coding problem

Consider a server with K messages $\{X_1, \cdots, X_K\}$ with $X_j \in \{0,1\}^t$ for each $j \in [K]$. Consider L clients $R_1, \cdots, R_L$, $L \geq K$, where for each $i \in [L]$, R_i is interested in one message, denoted by $X_{f(i)} \in X$, and knows some subset X_{S_i} of the other messages, referred to as the side information. Here, $f : [L] \to [K]$ is a function that maps the index of a client to the index of the client's requested message. We refer to $I \triangleq \{f(i), S_i : i \in [L]\}$ as an *instance* of the index coding problem.

An index code of length ℓ for a given instance I is a set of codewords in $\{0,1\}^\ell$ together with an encoding function $E^I : \{0,1\}^{tK} \to \{0,1\}^\ell$, and a set of L decoding functions $D_1^I, \cdots, D_L^I$ such that

$$D_i^I\left(E^I\left(X_1, \cdots, X_K\right), X_{S_i}\right) = X_{f(i)} \tag{14}$$

for all $i \in [L]$. We refer to $E^I\left(X_1, \cdots, X_K\right)$ as a *solution* to the instance I of the index coding problem. Note that the solution as well as the decoding functions depend on the instance I.

When $L = K$ and every client requires a distinct message, the side information of all the clients can be represented by a simple directed graph $G = (V, E)$, where $V = \{1, 2, \cdots, K\}$ with vertex i corresponding to the message X_i, and there is an arc $(i, j) \in E$ if $j \in S_i$. We denote the out-neighbors of vertex i as $\mathcal{N}(i)$.

For a given instance of the index coding problem, the minimum encoding length ℓ, as a function of message-length t, is denoted as β_t, and the *broadcast rate* is defined as (see [38], [42])

$$\beta = \inf_t \frac{\beta_t}{t} \tag{15}$$

It is well-known that the broadcast rate can be lower bounded as follows [37], [41], [43].

Proposition 2. *[41, Theorem 5.1] For an index coding instance with side information graph G, the broadcast rate β is lower bounded by the size of a maximum acyclic induced subgraph (MAIS) of G, denoted as $MAIS(G)$[8].*

B. Converse for Theorem 1

The key step of the converse is to show that for any scheme that satisfies the W-privacy constraint (7), the answer from the server must be a solution to an instance of the index coding problem that satisfies certain requirements as specified in the following lemma.

Lemma 1. *For a W-PIR-SI scheme, given a demand index W, a side information index set S, and a query $Q^{[W,S]}$, for every message set realization $\{X_1, \cdots, X_K\}$, the answer $A^{[W,S]}$ from the server must be a solution to the following instance of the index coding problem:*

1) *The instance has the messages $X_1, \cdots, X_K$;*
2) *There are K clients such that each client wants to decode a distinct message from $X_1, \cdots, X_K$, and possesses a side information that includes M messages;*
3) *The client that wants X_W has the side information set X_S; for each other client the side information set has M arbitrary messages from $X_1, \cdots, X_K$.*

Proof. Fix a demand index W, a side information index set S, and a query $Q^{[W,S]}$. Now, for each $i \in [K]$, there must exist a subset S_i and a decoding function $D_i^{Q^{[W,S]}}$ satisfying the conditions mentioned in Proposition 1. These sets and decoding functions can be used to construct an instance I of the index

[6] For the single-server case ($N = 1$), we drop the subscript from the query and the answer, and denote them, respectively, as $Q^{[W,S]}$ and $A^{[W,S]}$ for any given demand W and side information set S.

[7] Note that (W, S)-privacy implies W-privacy. Thus, we consider the W-privacy requirement without loss of generality.

[8] Note that the size of a graph is the number of its vertices. A MAIS of a graph G is an acyclic vertex-induced subgraph of G that has the largest number of vertices amongst all acyclic vertex-induced subgraphs of G.

coding problem as follows. Let $X_{[K]} = \{X_1, \dots, X_K\}$ be an arbitrary message set realization, and let $A^{[W,S]}$ be the answer corresponding to $X_{[K]}$ given $Q^{[W,S]}$. The instance I has the message set $X_{[K]}$, and K clients $\{R_1, \cdots, R_K\}$ such that:

- Client R_W requires packet X_W and has the side information set X_S; and
- Each other client R_i, $i \neq W$, requires X_i and has side information set X_{S_i}.

By the construction, the instance I satisfies the three conditions stated in the statement of the lemma. It remains to show that $A^{[W,S]}$ is a solution for the instance I. Towards this end, first, from Proposition 1, for each $i \in [K] \setminus \{W\}$, there exist a subset $S_i \subset [K] \setminus \{i\}$ and a decoding function $D_i^{Q^{[W,S]}}$ such that $D_i^{Q^{[W,S]}} \left(A^{[W,S]}, X_{S_i}\right) = X_i$. Further, there exists a decoding function $D_W^{Q^{[W,S]}}$ such that $D_W^{Q^{[W,S]}} \left(A^{[W,S]}, X_S\right) = X_W$. Therefore, each of the K clients can recover their request from their side information and $A^{[W,S]}$, and thus, $A^{[W,S]}$ is a solution to the instance I (cf. (14)). This completes the proof. ■

Note that Lemma 1 shows that the answer $\mathbf{A}^{[W,S]}$ from the server must be a solution to an instance of the index coding problem in which out-degree of every vertex in the corresponding side information graph G is equal to M. Note that, since the query $Q^{[W,S]}$ is assumed to be independent of the message values, in the corresponding instance of the index coding problem no client obtains any information about the messages outside of their side information by knowing $Q^{[W,S]}$. Next, we lower bound the broadcast rate for an index coding problem with side information graph G such that out-degree of every vertex in G is M as follows.

Lemma 2. *Consider any instance of the index coding problem such that the out-degree of every vertex in the corresponding side information graph G is equal to M. Then, the broadcast rate of the instance is lower bounded by $\lceil K/(M+1) \rceil$.*

Proof. For an index coding instance with side information graph G, the broadcast rate β is lower bounded by the size of a maximum acyclic induced subgraph (MAIS) of G (see Proposition 2). We show that for any graph G that satisfies the conditions of the lemma (i.e., the out-degree of each of the K vertices of G is M), it holds that

$$MAIS(G) \geq \left\lceil \frac{K}{M+1} \right\rceil.$$

Specifically, we build an acyclic subgraph of G induced by a vertex set Z of size at least $\lceil K/(M+1) \rceil$, through the following procedure. Recall that the vertex set of a side information graph G is denoted by $V = \{1, \dots, K\}$, and the set of out-neighbors of each vertex $i \in V$ is represented by $\mathcal{N}(i)$.

```
1. Set $Z = \emptyset$ and a candidate set of vertices $V' = V$;
2. Add an arbitrary vertex $i \in V'$ into Z, i.e., $Z \leftarrow Z \cup \{i\}$;
3. Set $V' \leftarrow V' \setminus (\mathcal{N}(i) \cup \{i\})$;
4. There are two cases:
     Case 1: If $V' \neq \emptyset$, then repeat Steps 2-4.
     Case 2: If $V' = \emptyset$, then terminate the procedure and
       return $Z$.
```

It is easy to see that the vertices in set Z returned by the procedure induce an acyclic subgraph of G. More specifically, if the vertices are ordered in the order they are added to Z, then there can only be an edge (i, j) if j was added to Z before i. This implies that the subgraph induced by Z cannot contain a cycle.

Further, note that the set Z contains at least $\lceil K/(M+1) \rceil$ vertices. At each removal step, there are at most $M+1$ vertices removed from V. Thus, the procedure iterates at least $\lceil K/(M+1) \rceil$ times, and in each iteration we add one vertex to Z. This implies that the size of Z is at least $\lceil K/(M+1) \rceil$, as was to be shown. ■

Corollary 1 (Converse of Theorem 1)**.** *For the W-PIR-SI problem with $N = 1$ server, K messages, and side information size M, the capacity is at most $\lceil K/(M+1) \rceil^{-1}$.*

Proof. Lemmas 1 and 2 imply that the length of the answer $A^{[W,S]}$ is at least $\lceil K/(M+1) \rceil t$ bits for any given W, S, and $X_{[K]}$. Then, by (9), the rate of any W-PIR-SI scheme for $N = 1$ server, K messages, and side information size M is upper bounded by $\lceil K/(M+1) \rceil^{-1}$. ■

C. Achievability for Theorem 1

In this section, we propose a W-PIR-SI scheme for $N = 1$ server, K messages, and side information size M, which achieves the rate $\lceil K/(M+1) \rceil^{-1}$. Recall that we assume that the distribution of the side information index set $\mathbf{S}$ and the conditional distribution of the demand index $\mathbf{W}$ given $\mathbf{S}$ are given respectively in (1) and (2). We describe the proposed scheme, referred to as the *Partition and Code* PIR-SI scheme, in the following.

Partition and Code PIR-SI Scheme: Given K, M, W, and S, denote $g \triangleq \lceil K/(M+1) \rceil$. The scheme consists of the following three steps.

Step 1. The user creates a partition of the K messages into g sets. For the ease of understanding, we describe the special case of $(M+1) \mid K$ first.

(a) Special case of $(M+1) \mid K$: Denote $P_1 \triangleq W \cup S$. The user randomly partitions the set of messages $[K] \setminus P_1$ into $g-1$ sets, each of size $M+1$, denoted as $P_2, \dots, P_g$.

(b) General case: Let $P_1, \dots, P_g$ be a collection of g empty sets. Note that, although empty at the beginning, once constructed, the sets $P_1, \dots, P_{g-1}$ will be of size $M+1$, and the set P_g will be of size $K-(g-1)(M+1)$. The user begins by assigning probabilities to the sets according to their sizes: the sets $P_1, \dots, P_{g-1}$ are each assigned a probability $(M+1)/K$, and the set P_g is assigned a probability $(K-(g-1)(M+1))/K$. Then, the user chooses a set randomly according to the assigned probabilities of the sets.

If the chosen set is a set $P \in \{P_1, \dots, P_{g-1}\}$, then the user fills the set P with the demand index W and the side information index set S. Next, it fills the remaining sets choosing one index at a time from the set of indices of the remaining messages uniformly at random until all the message indices are filled.

If the chosen set is the set P_g, then it fills P_g with the demand index W, and fills the remaining $K-(g-1)(M+1)-1$ places in the set P_g with randomly chosen elements from the side information index set S. (Note that once P_g is filled, it is possible that not all of the indices in the side information index set S are placed in the set.) Next, it fills the remaining sets by choosing one index at a time from the set of indices of the unplaced messages uniformly at random until all message indices are placed.

Step 2. The user sends to the server a uniform random permutation of the partition $\{P_1, \cdots, P_g\}$, i.e., they send $\{P_1, \cdots, P_g\}$ in a random order.

Step 3. The server computes the answer $A^{[W,S]}$ as a set of g inner products given by $A^{[W,S]} = \{A_{P_1}, \dots, A_{P_g}\}$, where $A_P = [X_1, \dots, X_K] \cdot \mathbf{1}_P$ for all $P \in \{P_1, \dots, P_g\}$. (Recall that $\mathbf{1}_P$ denotes the characteristic vector of the set P.)

Upon receiving the answer from the server, the user decodes X_W by subtracting off the contributions of the side information X_S from A_P for some $P \in \{P_1, \dots, P_g\}$ such that $W \in P$.

Example 2. *Assume that $K = 8$ and $M = 2$. Assume that the user demands the message X_2 and has two messages X_4 and X_6 as side information, i.e., $W = 2$ and $S = \{4, 6\}$. Following the Partition and Code PIR-SI scheme, the user labels three sets as P_1, P_2, and P_3, and assigns probability $3/8$ to each of the two sets P_1 and P_2, and assigns probability $2/8$ to the set P_3. Next, the user chooses one of these sets at random according to the assigned probabilities. Assume the user has chosen the set P_3. The user then places 2 into the set P_3, and chooses another element from $\{4, 6\}$ uniformly at random to place in P_3 as well. Say the user chooses 6 from the set $\{4, 6\}$, then the set P_3 becomes $P_3 = \{2, 6\}$. Then the user fills the other sets P_1 and P_2 randomly to exhaust the elements from $\{1, 2, 3, 5, 7, 8\}$. Say the user chooses $P_1 = \{1, 7, 8\}$ and $P_2 = \{3, 4, 5\}$. Then the user sends to the server a random permutation of $\{P_1, P_2, P_3\}$ as the query $Q^{[2,\{4,6\}]}$. The server sends three coded messages back to the user: $Y_1 = X_1 + X_7 + X_8$, $Y_2 = X_3 + X_4 + X_5$, and $Y_3 = X_2 + X_6$. The user can decode for X_2 by computing $X_2 = Y_3 - X_6$.*

From the server's perspective the user's demand is in either $\{1, 7, 8\}$ or $\{3, 4, 5\}$ with probability $3/8$ each, or in $\{2, 6\}$ with probability $2/8$. Given a set P_i, since any message in the set is equally likely to be the demand, it follows that

$$\mathbb{P}(\mathbf{W} = W' | \mathbf{Q}^{[\mathbf{W},\mathbf{S}]} = Q^{[1,\{2,3\}]}) = \frac{1}{8} = p_{\mathbf{W}}(W')$$

for any $W' \in [K]$.

In the following, we show that the Partition and Code PIR-SI scheme satisfies the W-privacy requirement for the setting in which the user's side information index set $\mathbf{S}$ and demand index $\mathbf{W}$ (given S) are distributed according to (1) and (2), respectively.

Lemma 3 (Achievability of Theorem 1)**.** *Consider the scenario of a W-PIR-SI problem in which:*

- *The server has messages $\mathbf{X}_{[K]} = \{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_K\}$;*
- *The side information index set $\mathbf{S}$ and the demand index $\mathbf{W}$ given the side information index set $\mathbf{S}$ follow the distributions given in* (1) *and* (2), *respectively.*

In this scenario, the Partition and Code PIR-SI scheme satisfies the W-privacy, and has rate $R = \lceil K/(M+1) \rceil^{-1}$.

Proof. First, we show that, for the Partition and Code protocol, we have

$$H\left(\mathbf{W} \mid \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}, \mathbf{A}^{[\mathbf{W},\mathbf{S}]}, \mathbf{X}_{[K]}\right) = H\left(\mathbf{W} \mid \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}\right). \quad (16)$$

To see this, we expand $H\left(\mathbf{W}, \mathbf{A}^{[\mathbf{W},\mathbf{S}]}, \mathbf{X}_{[K]} \mid \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}\right)$ in two ways as

$$\begin{aligned} &H\left(\mathbf{W}, \mathbf{A}^{[\mathbf{W},\mathbf{S}]}, \mathbf{X}_{[K]} \mid \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}\right) \\ &\quad = H\left(\mathbf{X}_{[K]} \mid \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}\right) + H\left(\mathbf{A}^{[\mathbf{W},\mathbf{S}]} \mid \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}, \mathbf{X}_{[K]}\right) \\ &\qquad + H\left(\mathbf{W} \mid \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}, \mathbf{X}_{[K]}, \mathbf{A}^{[\mathbf{W},\mathbf{S}]}\right) \quad (17) \\ &\quad = H\left(\mathbf{W} \mid \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}\right) + H\left(\mathbf{X}_{[K]} \mid \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}, \mathbf{W}\right) \\ &\qquad + H\left(\mathbf{A}^{[\mathbf{W},\mathbf{S}]} \mid \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}, \mathbf{W}, \mathbf{X}_{[K]}\right). \quad (18) \end{aligned}$$

Now, for the Partition and Code PIR-SI scheme, the answer is a deterministic function of the query and the message values. Thus, $H\left(\mathbf{A}^{[\mathbf{W},\mathbf{S}]} \mid \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}, \mathbf{X}_{[K]}\right) = 0$, and since conditioning cannot increase the entropy, we also get $H\left(\mathbf{A}^{[\mathbf{W},\mathbf{S}]} \mid \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}, \mathbf{W}, \mathbf{X}_{[K]}\right) = 0$. Further, since the demand is independent of the message values, and since the queries in the Partition and Code protocol are also independent of the message values, we have $H\left(\mathbf{X}_{[K]} \mid \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}\right) = H\left(\mathbf{X}_{[K]} \mid \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}, \mathbf{W}\right) = H\left(\mathbf{X}_{[K]}\right)$. Then, from (17) and (18), we get the desired result (16).

Next, it follows from (7) and (16) that, to show that the Partition and Code PIR-SI scheme satisfies the W-privacy, it suffices to show that

$$\mathbb{P}(\mathbf{W} = W' | \mathbf{Q}^{[\mathbf{W},\mathbf{S}]} = Q^{[W,S]}) = p_{\mathbf{W}}(W') = \frac{1}{K},$$

for any $W' \in [K]$.

Towards this end, first observe that

$$\mathbb{P}(\mathbf{W} \in P_i | \mathbf{Q}^{[\mathbf{W},\mathbf{S}]} = Q^{[W,S]}) = \frac{|P_i|}{K},$$

which follows from how the user chooses a set that would contain the demand index. Next, note that for any $i \in [g]$, we have

$$\begin{aligned} &\mathbb{P}(\mathbf{W} = W' | \mathbf{W} \in P_i, \mathbf{Q}^{[\mathbf{W},\mathbf{S}]} = Q^{[W,S]}) \\ &\quad = \begin{cases} \frac{1}{|P_i|} & \text{if } W' \in P_i, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (19)$$

This is because, given $\mathbf{W} \in P_i$, every index in P_i is equally likely to be the demand by the construction of the sets P_i and from (1) and (2).

Thus, for any $W' \in [K]$, we have

$$\begin{aligned}&\mathbb{P}(\mathbf{W} = W' | \mathbf{Q}^{[\mathbf{W},\mathbf{S}]} = Q^{[W,S]})\\ &= \sum_{i=1}^{g} \mathbb{P}(\mathbf{W} = W' | \mathbf{W} \in P_i, \mathbf{Q}^{[\mathbf{W},\mathbf{S}]} = Q^{[W,S]})\\ &\qquad \times \mathbb{P}(\mathbf{W} \in P_i | \mathbf{Q}^{[\mathbf{W},\mathbf{S}]} = Q^{[W,S]})\\ &= \frac{1}{K}.\end{aligned}$$

This completes the proof of W-privacy.

To compute the rate of the scheme, note that for any feasible (W, S), we have

$$\begin{aligned} H(\mathbf{A}^{[W,S]}) &= H(\mathbf{A}_{P_1}, \mathbf{A}_{P_2}, \dots, \mathbf{A}_{P_g})\\ &= \sum_{i=1}^{g} H(\mathbf{A}_{P_i})\\ &= g \times t,\end{aligned}$$

where the last two equalities follow since the messages $\mathbf{X}_j$'s (and hence the answers $\mathbf{A}_P$'s) are independently and uniformly distributed over $\mathbb{F}_{2^t}$. Further, observe that $H\left(\mathbf{A}^{[\mathbf{W},\mathbf{S}]}\right) = g \times t$, since $H\left(\mathbf{A}^{[W,S]}\right) = g \times t$ for all feasible (W, S). Thus, the Partition and Code PIR-SI scheme has rate

$$R = \frac{t}{g \times t} = \frac{1}{g} = \left\lceil \frac{K}{M+1} \right\rceil^{-1},$$

as was to be shown. ■

Remark 3. *It is easy to see that the Partition and Code scheme does not protect the desired message index and the side information index jointly. In fact, the Partition and Code scheme even leaks some information about user's side information when $M \geq 2$. To see this, let $P = \{P_1, P_2, \cdots, P_g\}$ be the partition in the query. Let $X_{i_1} \in P_i$ and $X_{j_1} \in P_j$. Then, the server learns that the user cannot have X_{i_1} and X_{j_1} together in their side information. On the other hand, it is easy to see that, when $M = 1$, the Partition and Code scheme also protects the side information individually.*

Remark 4. *It is worth noting that the Partition and Code scheme is reminiscent of the "clique covering" scheme that is well-known in the index coding literature, see e.g., [41, Chapter 6].*

V. Single-Server (W, S)-PIR-SI Problem

In this section we consider the single-server PIR-SI problem when (W, S)-privacy is required. We show the proof of the converse and the achievability for Theorem 2 through a reduction to an index coding instance and an MDS coding scheme, respectively.

A. Converse for Theorem 2

Similar to the W-privacy case, the key step of the converse is to show that for any scheme that satisfies the (W, S)-privacy constraint (7), the answer from the server must be a solution to an instance of the index coding problem that satisfies certain requirements as specified in the following lemma.

Lemma 4. *For a (W, S)-PIR-SI scheme, given a demand index W, a side information index set S, and a query $Q^{[W,S]}$, for every message set realization $X_{[K]} = \{X_1, \cdots, X_K\}$, the answer $A^{[W,S]}$ from the server must be a solution to the following instance of the index coding problem:*

1) *The instance has the messages $X_1, \cdots, X_K$;*
2) *There are $L = (K - M)\binom{K}{M}$ clients such that each client wants to decode one message and possesses a side information set that includes M other messages.*
3) *For each $i \in [K]$, for each $S_i \subset [K] \setminus \{i\}$ such that $|S_i| = M$, there exists a client that demands X_i and possesses X_{S_i} as their side information.*

Proof. Fix a demand index W, a side information index set S, and a query $Q^{[W,S]}$. We note that the (W, S)-privacy requirement implies that, for each message $i \in [K]$ and every set $S_i \subseteq [K] \setminus \{i\}$ of size M, there exists a decoding function $D_{i,S_i}^{Q^{[W,S]}}$, such that, for every message set realization $X_{[K]}$, the corresponding answer $A^{[W,S]}$ satisfies $D_{i,S_i}^{Q^{[W,S]}}\left(A^{[W,S]}, X_{S_i}\right) = X_i$. Otherwise, for a particular $\{i, S_i\}$, the server will know that the user cannot possess the side information S_i and demand the i-th message. Therefore, it holds that

$$\mathbb{P}\left(\mathbf{W} = i, \mathbf{S} = S_i \mid \mathbf{Q}^{[\mathbf{W},\mathbf{S}]} = Q^{[W,S]}\right) = 0, \qquad (20)$$

which, in turn, implies that $I\left(\mathbf{W}, \mathbf{S}; \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}, \mathbf{A}^{[\mathbf{W},\mathbf{S}]}, \mathbf{X}_{[K]}\right) > 0$. This violates the (W, S)-privacy requirement (8).

Now, let $X_{[K]} = \{X_1, \dots, X_K\}$ be an arbitrary message set realization, and let $A^{[W,S]}$ be the answer corresponding to X given $Q^{[W,S]}$. Consider an instance I of the index coding problem as stated in the lemma. For each $i \in [K]$, $S_i \subset [K] \setminus \{i\}$, $|S_i| = M$, a client can use the decoding function $D_{i,S_i}^{Q^{[W,S]}}$ to decode X_i from their side information X_{S_i} and $A^{[W,S]}$. Thus, $A^{[W,S]}$ is a solution to the instance I. ■

Next, we give a lower bound on the broadcast rate for an instance satisfying the conditions in Lemma 4.

Lemma 5. *For any instance of the index coding problem satisfying the conditions specified in Lemma 4, the broadcast rate is at least $K - M$.*

Proof. Let J denote an instance of the index coding problem satisfying the conditions in Lemma 4. Let J' be an instance of the index coding problem with the K messages $X_{[K]} = \{X_1, \cdots, X_K\}$ and $K - M$ clients. Each client has the side information X_S and wants to decode one distinct message from $X_{[K]} \setminus X_S$. Clearly, a solution to instance J is also a solution to instance J'. Observe that the side information graph for J' is acyclic, and thus, the broadcast rate for J' is at least $K - M$ from the MAIS lower bound in Proposition 2. This completes the proof. ■

Corollary 2 (Converse of Theorem 2)**.** *For the (W, S)-PIR-SI problem with $N = 1$ server, K messages, and side information size M, the capacity is at most $(K - M)^{-1}$.*

Proof. Lemmas 4 and 5 imply that the length of the answer $A^{[W,S]}$ is at least $(K - M)t$ for any given W, S, and X. Thus, by using (9), it follows that the rate of any (W,S)-PIR-SI scheme is upper bounded by $(K-M)^{-1}$. ∎

B. Achievability for Theorem 2

In this section, we give a (W,S)-PIR-SI scheme based on a maximum distance separable (MDS) code that achieves the rate of $(K-M)^{-1}$.

MDS PIR-SI Scheme: Given a demand index W and a side information index set S of size M, the user queries the server to send the $K-M$ parity symbols of a systematic $(2K-M,K)$ MDS code over the finite field $\mathbb{F}_{2^t}$. We assume that $t \geq \log_2(2K-M)$, or equivalently, $2^t \geq 2K-M$. Thus, it is possible construct a $(2K-M,K)$ MDS code over $\mathbb{F}_{2^t}$. The answer $A^{[W,S]}$ from the server consists of the $K-M$ parity symbols.

Lemma 6 (Achievability of Theorem 2)**.** *The MDS PIR-SI scheme satisfies the decodability condition in* (6) *and the* (W,S)*-privacy condition in* (8)*, and it has the rate of* $R=(K-M)^{-1}$.

Proof. For a $(2K-M,K)$ systematic MDS code, given the $K-M$ parity symbols and any M out of the K messages, the user can decode all of the remaining $K-M$ messages as the code is MDS. Thus, the user can recover their demanded message.

To ensure the (W,S)-privacy, note that the query and the answer are independent of the particular realizations of $\mathbf{W}$ and $\mathbf{S}$, but only depend on the size M of the side information index set. As the server already knows the size of the side information index set, it does not get any other information about $\mathbf{W}$ and $\mathbf{S}$ from the query and the answer. Thus, the MDS PIR-SI scheme satisfies the (W,S)-privacy requirement.

To compute the rate, note that for any W and S, the answer $A^{[W,S]}$ of the MDS PIR-SI scheme consists of $K-M$ parity symbols of a $(2K-M,K)$ systematic MDS code over $\mathbb{F}_{2^t}$. For an MDS code, any parity symbol is a linear combination of all the messages. Thus, as each message is distributed independently and uniformly over $\mathbb{F}_{2^t}$ and the parity symbols are linearly independent, every parity symbol is also independently and uniformly distributed over $\mathbb{F}_{2^t}$. Hence, we have $H(\mathbf{A}^{[W,S]}) = (K-M)t$. Therefore, the rate of the MDS PIR-SI scheme is $R=(K-M)^{-1}$. ∎

VI. W-PIR-SI Problem with Multiple Servers

In this section, we present a W-PIR-SI scheme, when data is replicated on multiple non-colluding servers. The rate achieved by the proposed scheme gives a lower bound on the capacity of multi-server W-PIR-SI problem. Our scheme builds up on the scheme proposed by Sun and Jafar in [9], referred to as the *Sun-Jafar protocol*, which deals with the case of no side information (i.e., $M=0$). Next, we use an example to describe the Sun-Jafar protocol. (The details can be found in [9].)

Example 3. *(Sun-Jafar Protocol [9])* $N=2$ *servers,* $K=2$ *messages, and* $M=0$*, i.e., no side information. The protocol assumes that each of the messages is* $t=N^K=4$ *bits long. For a message* X_i*, let* $[X_{i,1},\cdots,X_{i,t}]$ *be a uniform random permutation of its* t *bits. The user chooses a random permutation of the bits of* X_1*, and an independent random permutation of the bits of* X_2*. Suppose that the user is interested in downloading* X_1*. Then, they request the bits from the first server (S1) and the second server (S2) as given in first (i.e., left) part of Table I. The second (i.e., right) part of Table I corresponds to the case when the user is interested in* X_2*.*

TABLE I
Queries for the Sun-Jafar protocol for $N=2$ servers and $K=2$ messages (and no side-information), when the user demands the message X_1 and X_2, respectively. Each message X_i consists of $t=4$ bits $X_{i,1}, X_{i,2}, X_{i,3}, X_{i,4}$.

S1	S2
$X_{1,1}$	$X_{1,2}$
$X_{2,1}$	$X_{2,2}$
$X_{1,3}+X_{2,2}$	$X_{1,4}+X_{2,1}$

S1	S2
$X_{1,1}$	$X_{1,2}$
$X_{2,1}$	$X_{2,2}$
$X_{2,3}+X_{1,2}$	$X_{2,4}+X_{1,1}$

Note that the user can decode the four bits of the desired message from the answers it gets. To ensure privacy, note that each server is asked for a randomly chosen bit of each message and a sum of different pair of randomly chosen bits from each message irrespective of the demand. Therefore, a server cannot distinguish about which message is requested by the user.

Next, we give an example to outline our proposed scheme for multi-server PIR-SI before describing it formally.

Example 4. *(Multi-Server W-PIR-SI Scheme)* $N=2$ *servers,* $K=4$ *messages, and* $M=1$ *message as side information. Our scheme assumes that each message is* $t=N^{\frac{K}{M+1}}=4$ *bits long. The demand is privately chosen by the user, uniformly at random. The side information set has size* $M=1$*. It is chosen uniformly at random from the other messages, and is unknown to the servers.*

Consider an instance when the user demands $W=1$*, and the side information index set* $S=\{2\}$*. First step is that the user forms a partition of* $[K]$ *into* $g=K/(M+1)=2$ *sets* $\{P_1,P_2\}$*, where* $P_1=\{1,2\}$*, and* $P_2=\{3,4\}$[9]*. Next, the user sends a random permutation of* $\{P_1,P_2\}$ *to both the servers. The user and the servers form two coded messages* $\hat{X}_1$ *and* $\hat{X}_2$ *by taking the sum of the messages indexed by* P_1 *and* P_2 *as follows:* $\hat{X}_1=X_1+X_2$ *and* $\hat{X}_2=X_3+X_4$*. The last step is that the user and the servers apply the Sun-Jafar protocol using the two coded messages* $\hat{X}_1$ *and* $\hat{X}_2$*, such that the user can download* $\hat{X}_1$*. The form of the queries is given in Table I.*

From the answers, the user obtains $\hat{X}_1$*, from which it can decode the desired message* X_1 *using the side-information* X_2*. Note that the privacy property of the Sun-Jafar protocol guarantees that no server can distinguish which coded message is requested by the user. Thus, since the desired coded message can be either one, and in a coded message, any of the messages can be the demand, the privacy of the demand index is ensured.*

Note that in the above example the proposed scheme requires to download 6 bits, achieving the rate of $2/3$.

[9] The general procedure for forming the partition is elaborated in the formal description of the scheme.

It is shown in [9, Theorem 1] that the capacity of PIR with N servers and K messages and no side information is $(1+1/N+\cdots+1/N^{K-1})^{-1}$. Therefore, if the user attempts to download the demand without using their side information, then the capacity is $(1+1/N+1/N^2+1/N^3)^{-1} = 8/15$, which is smaller than $2/3$.

Next, we describe our W-PIR-SI scheme for N non-colluding servers storing identical copies of the K messages, when the user has a side information set of size M. For the sake of simplicity, here we describe the scheme for the case $(M+1) \mid K$, and refer the reader to Remark 5 for the case $(M+1) \nmid K$. Suppose the messages are $t = N^{K/(M+1)}$ bits long. Recall that, for a subset $S \subset [K]$, $\mathbf{1}_S$ denotes the characteristic vector of the set S. Let $g \triangleq K/(M+1)$.

Multi-Server W-PIR-SI Scheme: The scheme consists of the following three steps.

Step 1. Given the demand index W and the side information index set S, let $P_1 = W \cup S$. The user randomly partitions the set of messages $[K] \setminus P_1$ into $g-1$ sets of size $M+1$ each, denoted as $\{P_2, \cdots, P_g\}$.

Step 2. The user sends to all the servers a uniform random permutation of the partition $\{P_1, \cdots, P_g\}$, i.e., they send $\{P_1, \cdots, P_g\}$ in a random order. Then, the user and the servers form g *coded messages* $\{\hat{\mathbf{X}}_1, \dots, \hat{\mathbf{X}}_g\}$, where $\hat{\mathbf{X}}_i = [\mathbf{X}_1, \dots, \mathbf{X}_K] \cdot \mathbf{1}_{P_i}$ for $i \in [g]$.

Step 3. The user and the N servers utilize the Sun-Jafar protocol with g coded messages in such a way that the user can download the message $\hat{X}_1$.

Lemma 7. *Consider the scenario of a W-PIR-SI problem in which:*

- *The N non-colluding servers store identical copies of K messages $\mathbf{X}_{[K]} = \{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_K\}$;*
- *The side information index set $\mathbf{S}$ and the demand index $\mathbf{W}$ given the side information index set $\mathbf{S}$ follow the distributions given in* (1) *and* (2)*, respectively.*

In this scenario, the multi-server W-PIR-SI scheme satisfies the W-privacy, and has rate

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

Proof. Since the messages $\mathbf{X}_{[K]}$ are uniform and independent, the coded messages $\{\hat{\mathbf{X}}_1, \dots, \hat{\mathbf{X}}_g\}$ are uniform and independent as well. Thus, the rate of the scheme is that of the Sun-Jafar protocol for N servers and $K/(M+1)$ messages, which is given by $\left(1+1/N+\cdots+1/N^{K/(M+1)-1}\right)^{-1}$, see [9, Theorem 1].

To prove the privacy of the scheme, first note that it easy to show the following, similar to the proof of Lemma 3.

$$H\left(\mathbf{W} \mid \mathbf{Q}_j^{[\mathbf{W},\mathbf{S}]}, \mathbf{A}_j^{[\mathbf{W},\mathbf{S}]}, \mathbf{X}_{[K]}\right) = H\left(\mathbf{W} \mid \mathbf{Q}_j^{[\mathbf{W},\mathbf{S}]}\right), \quad (21)$$

for every $j \in [N]$. Next, it follows from (7) and (21) that, to show the W-privacy, it suffices to show that

$$\mathbb{P}(\mathbf{W} = W' | \mathbf{Q}_j^{[\mathbf{W},\mathbf{S}]} = Q_j^{[W,S]}) = p_{\mathbf{W}}(W') = \frac{1}{K},$$

for any $W' \in [K]$ and $j \in [N]$.

Towards this end, we note that the Sun-Jafar protocol protects the privacy of the demanded coded message, i.e., no server can have any information about which coded message the user is trying to download. Therefore, from the perspective of each server, every coded message is equally likely to include the demanded message. Further, any one of the $M+1$ messages in a coded message is equally likely to be the demanded message. In other words, for every server $j \in [N]$ and each $i \in [g]$, we have

$$\mathbb{P}\left(\mathbf{W} \in P_i \mid \mathbf{Q}_j^{[\mathbf{W},\mathbf{S}]} = Q_j^{[W,S]}\right) = \frac{M+1}{K},$$

and

$$\mathbb{P}\left(\mathbf{W} = W' | \mathbf{W} \in P_i, \mathbf{Q}_j^{[\mathbf{W},\mathbf{S}]} = Q_j^{[W,S]}\right) = \begin{cases} \frac{1}{M+1} & \text{if } W' \in P_i, \\ 0 & \text{otherwise.} \end{cases} \quad (22)$$

Hence, for any $W' \in [K]$ and $j \in [N]$, we have

$$\mathbb{P}\left(\mathbf{W} = W' | \mathbf{Q}_j^{[\mathbf{W},\mathbf{S}]} = Q_j^{[W,S]}\right) = \frac{1}{K}.$$

This completes the privacy proof. ∎

Remark 5. *Consider the case $(M+1) \nmid K$. Let $g \triangleq \lceil K/(M+1) \rceil$. For this case, we form the partitions $P_1, P_2, \dots, P_g$ in Step 1 of the multi-server W-PIR-SI scheme in the same way as described in Step 1 Case (b) of the Partition and Code PIR-SI scheme. It is straightforward to adapt the proof of privacy as in Lemma 7 for this case. Further, it is easy to see that the rate of the underlying scheme for this case is* $\left(1+1/N+\cdots+1/N^{\lceil K/(M+1)\rceil-1}\right)^{-1}$.

It is easy to see that Theorem 3 follows from Lemma 7 and Remark 5.

VII. Conclusion

In this paper we considered the problem of Private Information Retrieval (PIR) with side information (PIR-SI), in which the user has *a priori* a subset of the messages at the server obtained from other sources. The goal of the user is to download a message, which is not in their side information, from the server(s) while satisfying a certain privacy constraint. We considered two types of privacy constraints: W-privacy—in which the user wants to protect the identity of the demand (i.e., which message the user wishes to download) from the server(s); and (W,S)-privacy—in which the user wants to protect the identities of the demand and the side information jointly, from the server(s). First, we focused on the case of single server (i.e., single database). We established the capacity of single-server PIR-SI problem when (W,S)-privacy is required for arbitrary distribution of the demand index W and the side information index set S. In the case of W-privacy, we established the capacity of single-server PIR-SI problem for the uniform distribution. Second, we extended our single-server PIR-SI scheme for W-privacy to the case of multiple servers (i.e., multiple copies of the database). Our multi-server PIR-SI scheme uses ideas from our single-server PIR-SI scheme in conjunction with the no-side-information scheme of Sun and Jafar in [9]. The capacity of multi-server PIR-SI problem under the W-privacy constraint remains open.

References

[1] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," in *IEEE Symposium on Foundations of Computer Science*, 1995, pp. 41–50.

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

[3] E. Kushilevitz and R. Ostrovsky, "Replication is not needed: Single database, computationally-private information retrieval," in *IEEE Symposium on Foundations of Computer Science*, 1997, p. 364.

[4] C. Cachin, S. Micali, and M. Stadler, "Computationally private information retrieval with polylogarithmic communication," in *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 1999, pp. 402–414.

[5] S. Yekhanin, "Private information retrieval," *Communications of the ACM*, vol. 53, no. 4, pp. 68–73, 2010.

[6] A. Beimel and Y. Ishai, "Information-theoretic private information retrieval: A unified construction," in *Automata, Languages and Programming*. Springer, 2001, pp. 912–926.

[7] A. Beimel, Y. Ishai, E. Kushilevitz, and J.-F. Raymond, "Breaking the $\mathrm{O}(n^{1/(2k-1)})$ barrier for information-theoretic private information retrieval," in *43rd Annual IEEE Symposium on Foundations of Computer Science*, 2002, pp. 261–270.

[8] W. Gasarch, "A survey on private information retrieval," in *Bulletin of the EATCS*. Citeseer, 2004.

[9] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4075–4088, July 2017.

[10] H. Sun and S. A. Jafar, "The capacity of robust private information retrieval with colluding databases," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2361–2370, April 2018.

[11] K. Banawan and S. Ulukus, "Multi-message private information retrieval: Capacity results and near-optimal schemes," *IEEE Transactions on Information Theory*, vol. 64, no. 10, pp. 6842–6862, Oct 2018.

[12] ——, "The capacity of private information retrieval from coded databases," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1945–1956, March 2018.

[13] R. Tajeddine, O. W. Gnilke, and S. El Rouayheb, "Private information retrieval from mds coded data in distributed storage systems," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 7081–7093, Nov 2018.

[14] N. Shah, K. Rashmi, and K. Ramchandran, "One extra bit of download ensures perfectly private information retrieval," in *2014 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2014, pp. 856–860.

[15] T. H. Chan, S. Ho, and H. Yamamoto, "Private information retrieval for coded storage," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 2842–2846.

[16] R. Tajeddine and S. El Rouayheb, "Private information retrieval from MDS coded data in distributed storage systems," *2016 IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2016.

[17] A. Fazeli, A. Vardy, and E. Yaakobi, "Codes for distributed PIR with low storage overhead," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 2852–2856.

[18] S. R. Blackburn and T. Etzion, "PIR array codes with optimal PIR rates," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 2658–2662.

[19] R. Freij-Hollanti, O. W. Gnilke, C. Hollanti, and D. A. Karpuk, "Private information retrieval from coded databases with colluding servers," *SIAM Journal on Applied Algebra and Geometry*, vol. 1, no. 1, pp. 647–664, 2017.

[20] B. Chor, N. Gilboa, and M. Naor, *Private information retrieval by keywords*. Citeseer, 1997.

[21] C. Gentry and Z. Ramzan, "Single-database private information retrieval with constant communication rate," in *International Colloquium on Automata, Languages, and Programming*. Springer, 2005, pp. 803–815.

[22] R. Sion and B. Carbunar, "On the computational practicality of private information retrieval," in *In Proceedings of the Network and Distributed Systems Security Symposium*, 2007.

[23] Z. Dvir and S. Gopi, "2-Server PIR with subpolynomial communication," *Journal of the ACM (JACM)*, vol. 63, no. 4, p. 39, 2016.

[24] R. Tandon, "The capacity of cache aided private information retrieval," in *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2017, pp. 1078–1082.

[25] Y. Wei, K. Banawan, and S. Ulukus, "Cache-aided private information retrieval with partially known uncoded prefetching: Fundamental limits," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 6, pp. 1126–1139, June 2018.

[26] ——, "Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching," *IEEE Transactions on Information Theory*, vol. 65, no. 5, pp. 3215–3232, May 2019.

[27] S. Kadhe, B. Garcia, A. Heidarzadeh, S. E. Rouayheb, and A. Sprintson, "Private information retrieval with side information: The single server case," in *2017 55th Annual Allerton Conference on Communication, Control, and Computing*, Oct 2017, pp. 1099–1106.

[28] ——, "Private information retrieval with side information," *CoRR*, vol. abs/1709.00112, 2017. [Online]. Available: http://arxiv.org/abs/1709.00112

[29] Z. Chen, Z. Wang, and S. A. Jafar, "The capacity of private information retrieval with private side information," *CoRR*, vol. abs/1709.03022, 2017. [Online]. Available: http://arxiv.org/abs/1709.03022

[30] S. P. Shariatpanahi, M. J. Siavoshani, and M. A. Maddah-Ali, "Multi-message private information retrieval with private side information," May 2018. [Online]. Available: arXiv:1805.11892

[31] A. Heidarzadeh, B. Garcia, S. Kadhe, S. E. Rouayheb, and A. Sprintson, "On the capacity of single-server multi-message private information retrieval with side information," in *2018 56th Annual Allerton Conf. on Commun., Control, and Computing*, Oct 2018.

[32] S. Li and M. Gastpar, "Single-server multi-message private information retrieval with side information," in *2018 56th Annual Allerton Conf. on Commun., Control, and Computing*, Oct 2018.

[33] A. Heidarzadeh, F. Kazemi, and A. Sprintson, "Capacity of single-server single-message private information retrieval with coded side information," in *2018 IEEE Information Theory Workshop (ITW)*, Nov 2018, pp. 1–5.

[34] S. Patel, G. Persiano, and K. Yeo, "Private stateful information retrieval," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '18. New York, NY, USA: ACM, 2018, pp. 1002–1019. [Online]. Available: http://doi.acm.org/10.1145/3243734.3243821

[35] M. Karmoose, L. Song, M. Cardone, and C. Fragouli, "Private broadcasting: An index coding approach," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 2543–2547.

[36] D. T. Kao, M. A. Maddah-Ali, and A. S. Avestimehr, "Blind index coding," *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2076–2097, 2017.

[37] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1479–1494, March 2011.

[38] N. Alon, E. Lubetzky, U. Stav, A. Weinstein, and A. Hassidim, "Broadcasting with side information," in *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, Oct 2008, pp. 823–832.

[39] M. Effros, S. El Rouayheb, and M. Langberg, "An equivalence between network coding and index coding," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2478–2487, 2015.

[40] S. El Rouayheb, A. Sprintson, and C. Georghiades, "On the index coding problem and its relation to network coding and matroid theory," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3187–3195, 2010.

[41] F. Arbabjolfaei and Y.-H. Kim, "Fundamentals of index coding," *Foundations and Trends in Communications and Information Theory*, vol. 14, no. 3-4, pp. 163–346, 2018. [Online]. Available: http://dx.doi.org/10.1561/0100000094

[42] A. Blasiak, R. Kleinberg, and E. Lubetzky, "Broadcasting with side information: Bounding and approximating the broadcast rate," *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5811–5823, Sept 2013.

[43] F. Arbabjolfaei, "Index coding: Fundamental limits, coding schemes, and structural properties," in *PhD Thesis, UC San Diego*, 2017.

Swanand Kadhe (S'14–M'19) is currently a postdoctoral research scholar at University of California Berkeley. He received the Ph.D. degree in electrical engineering from Texas A&M University, College Station, USA, in 2017. Prior to that, he received the B.E degree in electronics and telecommunications engineering from the University of Pune, India, in 2007, and the M.Tech. degree in electrical engineering from Indian Institute of Technology Kanpur, India, in 2009. He was a researcher at the Innovation Labs of Tata Consultancy Services (TCS) Bangalore, India, (2009–2012). He is a recipient of Tata Consultancy Services best paper award in 2010. His research interests lie in the areas of distributed storage, computing, and learning systems with a focus on privacy and security.

Brenden Garcia is currently a software developer at Epic Systems Corporation. He received his B.S. in Computer Engineering and Applied Mathematical Sciences from Texas A&M University in 2016, and his M.S. in Computer Engineering from Texas A&M University in 2018. His research interests include private information retrieval, network coding, and network security.

Anoosheh Heidarzadeh (S'08–M'13) received the Ph.D. degree in electrical and computer engineering from Carleton University, Ottawa, ON, Canada, in 2012; and he was a postdoctoral research fellow at California Institute of Technology, Pasadena, CA, USA, from 2013 to 2014. He is currently a Visiting Assistant Professor at the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. His research interests include network coding and network information theory, private information retrieval, private function computation, coded distributed computing, group testing, compressed sensing, and game theory.

Salim El Rouayheb (S'07–M'09) is currently an assistant professor in the ECE department at Rutgers University, New Jersey. He is the recipient of the NSF career award and the Google Faculty Research Award. He received the Diploma degree in electrical engineering from the Lebanese University, Faculty of Engineering, Roumieh, Lebanon, in 2002, and the M.S. degree from the American University of Beirut, Lebanon, in 2004. He received the Ph.D. degree in electrical engineering from Texas A&M University, College Station, in 2009. He was a postdoctoral research fellow at UC Berkeley (2010–2011) and a research scholar at Princeton University (2012–2013). His research interests are in information theory and coding theory with a focus on applications to data security and privacy.

Alex Sprintson is a faculty member in the Department of Electrical and Computer Engineering, Texas A&M University, College Station, USA. He is currently serving as a rotating program director at the US National Science Foundation (NSF). Dr. Sprintson received his B.S. degree (*summa cum laude*), M.S., and Ph.D. degrees in electrical engineering from the Technion in 1995, 2001, and 2003, respectively. From 2003 to 2005, he was a Postdoctoral Research Fellow at the California Institute of Technology, Pasadena. His research interests lie in the general area of communication networks with a focus on wireless network coding, distributed storage, and software-defined networks. Dr. Sprintson received the Wolf Award for Distinguished Ph.D.students, the Viterbi Postdoctoral Fellowship, the TAMU College of Engineering Outstanding Contribution Award, and the NSF CAREER award. From 2013 to 2019 he served as an Associate Editor of the IEEE Transactions on Wireless Communications. He has been a member of the Technical Program Committee for the IEEE Infocom 2006–2020.