Cross Subspace Alignment and the Asymptotic Capacity of X-Secure T-Private Information Retrieval

Zhuqing Jia, *Student Member, IEEE*, Hua Sun, *Member, IEEE*, and Syed Ali Jafar, *Fellow, IEEE*

***Abstract*— X-secure and T-private information retrieval (XSTPIR) is a form of private information retrieval where data security is guaranteed against collusion among up to X servers and the user's privacy is guaranteed against collusion among up to T servers. The capacity of XSTPIR is characterized for an arbitrary number of servers N and arbitrary security and privacy thresholds X and T, in the limit as the number of messages $K \to \infty$. Capacity is also characterized for any number of messages if either $N = 3, X = T = 1$ or if $N \leq X+T$. Insights are drawn from these results, about aligning versus decoding noise, dependence of PIR rate on field size, and robustness to symmetric security constraints. In particular, the idea of cross subspace alignment, i.e., introducing a subspace dependence between Reed–Solomon code parameters, emerges as the optimal way to align undesired terms while keeping desired terms resolvable.**

***Index Terms*— Capacity, security, privacy, distributed storage.**

I. Introduction

MOTIVATED by the importance of security and privacy in the era of big data and distributed storage, in this work we explore the information theoretic capacity of private information retrieval (PIR) in a secure distributed storage system. Specifically, our focus is on the X-secure and T-private information retrieval problem (XSTPIR). A PIR scheme is said to be T-private if it allows a user to retrieve a desired message from a database of K messages stored at N distributed servers, without revealing any information about the identity of the desired message to any group of up to T colluding servers. Similarly, a distributed storage scheme is said to be X-secure[1] if it guarantees that any group of up to X colluding servers learn nothing about the stored data. The T and X parameters may be chosen arbitrarily depending on the relative importance of security and privacy for any given application.

The rate of a PIR scheme is the ratio of the number of bits retrieved by the user to the total number of bits downloaded from all servers. The supremum of achievable rates is called the capacity of PIR. The capacity of the basic PIR setting was found in [1] to be

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

The result was generalized subsequently in [2] to the T-PIR setting, as

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

Further generalizations of T-privacy, e.g., when privacy is required only against certain specified collusion patterns [3], [4] have also been explored. In particular, capacity is known for disjoint colluding sets [4].

The rapidly growing body of literature in this area has produced capacity results for PIR under a rich variety of constraints [5]–[20]. However, the capacity for the natural setting of secure storage remains unknown, and relatively unexplored. While a number of efforts are motivated by security concerns, such efforts have focused largely on other models, e.g., wiretap models where data security is desired against eavesdroppers listening to the communication between the user and the servers [21], [22], Byzantine models where the servers may respond incorrectly by introducing erasures or errors in their response to the user's queries [23]–[27], and so called symmetric security models [28]–[30] that allow the user to learn nothing about the data besides his desired message. An exception in this regard is the recent work in [31] where PIR with distributed storage is explored and the asymptotic (large K) capacity for the $X = T = 1$ setting is bounded as

$$\left(1 - \frac{1}{\sqrt{N}}\right)^2 \leq \lim_{K\to\infty} C_{\text{XSTPIR}}(N,K,X=1,T=1) \leq \left(1 - \frac{1}{N}\right). \quad (3)$$

As the main result of our work, we close this gap and characterize the asymptotic capacity of XSTPIR for all N, X, T

Manuscript received August 22, 2018; revised April 30, 2019; accepted May 1, 2019. Date of publication May 10, 2019; date of current version August 16, 2019. The work of Z. Jia and S. A. Jafar was supported in part by the NSF under Grant CCF-1617504, Grant CNS-1731384, in part by the Office of Naval Research (ONR) Grant N00014-18-1-2057, and in part by the Army Research Office (ARO) Grant W911NF-16-1-0215.

Z. Jia and S. A. Jafar are 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: zhuqingj@uci.edu; syed@uci.edu).

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

Communicated by P. Mitran, Associate Editor for Communications.

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.2019.2916079

[1] In other words, everything that is stored at any X servers must be independent of the K messages. Besides X-security, no other constraints are imposed on the storage. The storage and the PIR scheme are jointly optimized to maximize the capacity of XSTPIR.

as follows.

$$\lim_{K\to\infty} C_{\text{XSTPIR}}(N,K,X,T) = \begin{cases} 1-\left(\frac{X+T}{N}\right), & N > X+T \\ 0, & N \le X+T. \end{cases} \tag{4}$$

The asymptotic capacity characterization leads us to supplementary results which include a general upper bound on the capacity of XSTPIR, the exact capacity characterization for any number of messages K if $N \le X+T$, and the exact capacity characterization for any K if $X = T = 1, N = 3$. The results also lead us to interesting observations about aligning versus decoding noise, dependence of PIR rate on field size, robustness to symmetric security constraints, and a particularly useful idea called cross subspace alignment. When privately retrieving multiple symbols from a desired message in a secure distributed storage system, the structure (say, $1, \beta, \beta^2, \cdots$, for one symbol and $1, \gamma, \gamma^2, \cdots$ for another, as in Reed-Solomon (RS) codes) of storage and queries for each symbol determines the number of dimensions occupied by interference and the resolvability of desired symbols. Choosing identical RS codes ($\beta = \gamma$) for each symbol of the same message would cause desired signals to align among themselves, while making the RS codes insufficiently dependent would cause interference to occupy too many dimensions. Cross subspace alignment is achieved by drawing the code parameters as linear combinations from the same subspace (say, $\beta = 1+\alpha, \gamma = 2+\alpha$), which turns out to be the optimal way to align interference while keeping desired symbols resolvable. For a summary of results and a better explanation of the main observations we refer the reader directly to Section III.

Let us start by defining our notation.

Notation: Let $[m:n]$ denote the set $\{m, m+1, \ldots, n\}$ for any two integers m, n such that $m \le n$. For sake of simplicity, let $X_{[m:n]}$ denote the set of random variables $\{X_m, X_{m+1}, \ldots, X_n\}$. For an index set $\mathcal{I} = \{i_1, i_2, \ldots, i_n\}$, let $X_{\mathcal{I}}$ denote the set $\{X_{i_1}, X_{i_2}, \ldots, X_{i_n}\}$. For variables $a_n, n \in [1:N]$ and an arbitrary function $f(\cdot)$, we denote the $N \times 1$ vector whose n^{th} term is $f(a_n)$, as $\overrightarrow{f(a)}$. Similarly, $\overrightarrow{g(b)}$ denotes the vector $(g(b_1), \cdots, g(b_n))^T$ for variables $b_n, n \in [1:N]$ and a function $g(\cdot)$. For such $N \times 1$ vectors $\overrightarrow{f(a)}$ and $\overrightarrow{g(b)}$, let $\overrightarrow{f(a)} \circ \overrightarrow{g(b)}$ denote their Hadamard product, i.e., the $N \times 1$ vector whose n^{th} term is $f(a_n) \times g(b_n)$. The notation $X \sim Y$ is used to indicate that X and Y are identically distributed. When a natural number, say $\ell \in \mathbb{N}$, is used to represent an element of a finite field $\mathbb{F}_q$, it denotes the sum of ℓ ones in $\mathbb{F}_q$, i.e., ℓ is identified[2] with $\sum_{l=1}^{\ell} 1$, where the addition is over $\mathbb{F}_q$.

II. XSTPIR: Problem Statement

Consider data that is stored at N distributed servers. The data consists of K independent messages, $W_1, W_2, \cdots, W_K$, and each message is represented[3] by L random symbols from the finite field $\mathbb{F}_q$.

$$H(W_1) = H(W_2) = \cdots = H(W_K) = L, \tag{5}$$

$$H(W_1, W_2, \ldots, W_K) = KL, \tag{6}$$

in q-ary units. There are N servers. The information stored at the n^{th} server is denoted by $S_n, n \in [1:N]$. An X-secure scheme, $0 \le X < N$, guarantees that any X (or fewer) colluding servers learn nothing about the data.

$$[X\text{-Security}] \qquad I(S_{\mathcal{X}}; W_1, \ldots, W_K) = 0, \quad \forall \mathcal{X} \subset [1:N], |\mathcal{X}| = X. \tag{7}$$

Besides X-security, we place no other constraint[4] on the amount of storage or the storage code used at each server, all of which is jointly optimized to maximize the capacity of XSTPIR. To ensure information retrieval is possible, note that the set of messages $W_1, \cdots, W_K$ must be a function of $S_{[1:N]}$.

$$H(W_1, \cdots, W_K \mid S_{[1:N]}) = 0. \tag{8}$$

The user generates a desired message index θ privately and uniformly from $[1:K]$. In order to retrieve W_θ privately, the user generates N queries, $Q_1^{[\theta]}, Q_2^{[\theta]}, \ldots, Q_N^{[\theta]}$. The query $Q_n^{[\theta]}$ is sent to the n^{th} server. The user has no prior knowledge of the information stored at the servers, i.e.,

$$I(S_{[1:N]}; Q_{[1:N]}^{[\theta]}, \theta) = 0. \tag{9}$$

T-privacy, $1 \le T \le N$, guarantees that any T (or fewer) colluding servers learn nothing about θ.

$$[T\text{-Privacy}] \quad I(Q_{\mathcal{T}}^{[\theta]}, S_{\mathcal{T}}; \theta) = 0, \quad \forall \mathcal{T} \subset [1:N], |\mathcal{T}| = T. \tag{10}$$

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

$$H(A_n^{[\theta]} | Q_n^{[\theta]}, S_n) = 0. \tag{11}$$

From all the answers the user must be able to recover the desired message W_θ,

$$[\text{Correctness}] \qquad H(W_\theta | A_{[1:N]}^{[\theta]}, Q_{[1:N]}^{[\theta]}, \theta) = 0. \tag{12}$$

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

$$R = \frac{L}{D}, \tag{13}$$

[2] Recall that $q = p^n$ for some prime p and integer $n \ge 1$, and the elements of $\mathbb{F}_q$ are the polynomials over $\mathbb{F}_p$ whose degree is strictly less than n. The natural number ℓ represents a polynomial of degree 0, i.e., simply an element of $\mathbb{F}_p$. Recall that elements of $\mathbb{F}_p$ may be represented as integers modulo p. Therefore, there are only p distinct values of ℓ. The prime p is called the characteristic of the field.

[3] As usual for an information theoretic formulation, the actual size of each message is allowed to approach infinity. The parameters L and q partition the data into blocks and may be chosen freely by the coding scheme to match the code dimensions. Since the coding scheme for a block can be repeated for each successive block of data with no impact on rate, it suffices to consider one block of data subject to optimization over L and q.

[4] The amount of storage at each server is not constrained *a priori*, however, it is remarkable that none of the XSTPIR schemes in this work end up storing more than KL symbols at each server. Thus the amount of storage used is not worse than a data replication scheme in the absence of security constraints.

where D is the expected value (with respect to the random queries) of the number of q-ary symbols downloaded by the user from all servers. The capacity of XSTPIR, denoted $C_{\text{XSTPIR}}(N, K, X, T)$, is the supremum of achievable rates.

Finally, note that setting $X = 0$ and $T = 1$ reduces the XSTPIR problem to the basic PIR setting where data storage is not secure and the user's privacy is only guaranteed if no collusion takes place among servers. Setting $X = 0$ for arbitrary T, reduces XSTPIR to the T-PIR problem. Setting $T = 0$ for arbitrary X reduces XSTPIR to an X-secure storage scheme with no privacy constraint.

III. Capacity of XSTPIR: Results and Observations

The results of this work are presented in this section, followed by some observations.

A. Results

Our first result, presented in the following theorem, is an upper bound on the capacity of XSTPIR.

Theorem 1:

$$C_{\text{XSTPIR}}(N, K, X, T) \leq \left(\frac{N-X}{N}\right) C_{\text{TPIR}}(N-X, K, T). \quad (14)$$

The proof of Theorem 1 appears in Section IV. The intuition behind Theorem 1 may be understood through a thought experiment as follows. Without loss of generality, suppose the expected number of bits downloaded from each server is the same. Now, relax the constraints so that $S_{[1:X]}$, i.e., the stored information at the first X servers is made available globally (to all servers and to the user) for free, the messages $W_1, W_2, \cdots, W_K$ are made available to all servers, and the data-security constraint is eliminated. None of this can hurt capacity because any XSTPIR scheme from before can still be used with the relaxed constraints. So any upper bound on capacity of this relaxed setting is still an upper bound on the capacity of the original XSTPIR setting. The relaxed setting is analogous to the T-PIR problem with K messages and $N - X$ servers, for which we already know the optimal download per server from the existing capacity results for T-PIR. Thus, the statement of Theorem 1 follows. However, formalizing this intuition into a proof is not trivial because of the correlated side-information generated at the user and servers in the process of relaxing the constraints. Indeed, the formal proof presented in Section IV takes a less direct approach.

It turns out the bound in Theorem 1 is quite powerful. In fact, we suspect that this bound might be tight in general. An immediate observation is that if we set $X = 0$, i.e., remove the data storage security constraint, then the bound is tight because it gives us the capacity of T-PIR. Similarly, if we set $T = 0$, i.e., the privacy constraint is removed, then the bound is also tight, and the capacity in the absence of privacy constraints is easily seen to be $C_{\text{XSTPIR}}(N, K, X, T = 0) = 1 - \frac{X}{N}$, which is achievable by a simple secret-sharing scheme. We further prove the tightness of this bound for the cases identified in our next set of results. The first setting identifies a somewhat degenerate extreme where it is optimal to download everything.

Theorem 2: If $N \leq X + T$, then[5] for arbitrary K,

$$C_{\text{XSTPIR}}(N, K, X, T) = \left(\frac{N-X}{N}\right) C_{\text{TPIR}}(N-X, K, T) \quad (15)$$

$$= \frac{N-X}{NK}. \quad (16)$$

The proof of Theorem 2 is presented in Section V. Since the upper bound is already provided by Theorem 1, only a proof of achievability is needed. Furthermore, since retrieving the desired message in this setting amounts to downloading everything stored at all servers regardless of which message is desired, the only thing required for the achievable scheme is a secure storage scheme, which is readily achieved by including X uniformly random noise symbols for every $N - X$ symbols of each message.

Next, the main result of this paper is the asymptotic capacity characterization presented in the following theorem.

Theorem 3: As the number of messages $K \to \infty$, for arbitrary N, X, T,

$$\lim_{K\to\infty} C_{\text{XSTPIR}}(N, K, X, T)$$

$$= \lim_{K\to\infty} \left(\frac{N-X}{N}\right) C_{\text{TPIR}}(N-X, K, T) \quad (17)$$

$$= \begin{cases} 1 - \left(\frac{X+T}{N}\right), & N > X + T \\ 0, & N \leq X + T. \end{cases} \quad (18)$$

The proof of Theorem 3 appears in Section VI. Theorem 3 is significant for two reasons. First, asymptotic capacity results are particularly relevant for PIR problems because the capacity approaches its asymptotic value extremely quickly — the gap is negligible even for moderate values of K, and K is typically a large value. Second, the asymptotic capacity result showcases a new idea, *cross subspace alignment*, that is interesting by itself.

Insights from the asymptotically optimal scheme allow us to settle the exact capacity of XSTPIR with $X = T = 1$, $N = 3$ and arbitrary K.

Theorem 4: If the number of servers, $N = 3$, and $X = T = 1$, then for arbitrary number of messages, K,

$$C_{\text{XSTPIR}}(N = 3, K, X = 1, T = 1)$$

$$= \left(\frac{N-X}{N}\right) C_{\text{TPIR}}(N-X, K, T) \quad (19)$$

$$= \frac{2}{3}\left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{K-1}}\right)^{-1}. \quad (20)$$

Theorem 4 is proved in Section VII. The capacity achieving scheme introduces a new insight. For almost all PIR settings studied so far, asymptotic capacity achieving schemes have been found that send a uniformly random query vector to each server and download a product of the query vector and information stored at the server. Suppose the query vector is uniform over $\mathbb{F}_q^M$. Then with probability $1/q^M$ the query

[5] Note that $N > X$ by definition.

vector is all zero, and the scheme requests nothing from the server. Typically M depends on the number of messages K. As K approaches infinity the probability of requesting nothing approaches zero, so this does not help in the asymptotic sense. However, if the same scheme is used for finite K, then M is also finite, $1/q^M > 0$, and the average download is reduced by the factor $(1-1/q^M)$, which improves the achieved rate of the scheme. It is remarkable that the rate achieved in this way depends on the field size. This idea is essential to the capacity achieving scheme for Theorem 4.

Next we present some observations that place our results in perspective.

B. Observations

1) Alignment of Noise and Interference: Consider the simplest non-trivial setting for XSTPIR, where $X = 1, T = 1$, and the number of servers, $N \geq 3$. A natural idea for providing $X = 1$ secure storage is to include 1 independent uniformly random noise symbol along with the L symbols of each message, creating a new message with $M = L + 1$ symbols. This new message is stored across N servers according to an (N, M) MDS code, essentially storing a linear combination of the M message symbols at each server, where the coefficients for the noise symbol at each server must be non-zero. Capacity is known for PIR with coded storage (MDS-PIR [5]), and one might wonder if such an MDS-PIR scheme might suffice to achieve capacity with secure storage. It is not difficult to see that the best rate achievable with such an MDS-PIR scheme is

$$R_{\text{MDS-PIR}} = \frac{M-1}{M}\left(1+\left(\frac{M}{N}\right)+\cdots+\left(\frac{M}{N}\right)^{K-1}\right)^{-1}. \tag{21}$$

The $\frac{M-1}{M}$ penalty appears because one of the M symbols of the *decoded* message is the noise symbol. As $K \to \infty$, the rate approaches $R_{\text{MDS-PIR},\infty} = \frac{M-1}{M}\left(1-\left(\frac{M}{N}\right)\right)$. This expression takes its maximum value when $M = \sqrt{N}$, so it can be bounded as,

$$\begin{aligned} R_{\text{MDS-PIR},\infty} &\leq \frac{\sqrt{N}-1}{\sqrt{N}}\left(1-\left(\frac{\sqrt{N}}{N}\right)\right) \\ &= \left(1-\frac{1}{\sqrt{N}}\right)^2. \end{aligned} \tag{22}$$

Note that this expression matches the achievable rate bound of [31]. However, it is strictly smaller than, $1 - 2/N$, the asymptotic capacity of XSTPIR for this setting. Evidently, the natural MDS-PIR solution, and the secret sharing based scheme of [31], are asymptotically suboptimal. In fact, the MDS-PIR solution falls short of the asymptotic $(K \to \infty)$ capacity of XSTPIR, even if the MDS-PIR scheme is only required to deal with $K = 2$ messages. Denoting the corresponding rate of the MDS-PIR scheme as $R_{\text{MDS-PIR},2}$, we have,

$$\begin{aligned} R_{\text{MDS-PIR},2} &\leq \frac{\sqrt{N+1}}{\sqrt{N+1}+1}\left(1+\left(\frac{\sqrt{N+1}+1}{N}\right)\right)^{-1} \\ &\leq 1-\frac{2}{N}. \end{aligned} \tag{23}$$

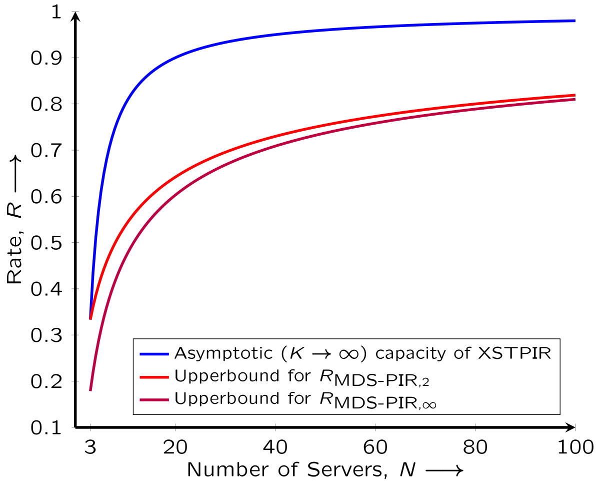

Fig. 1. Suboptimality of the rate achieved by the $X = 1$ secure MDS-PIR alternative that allows the user to decode noise relative to the rate achieved with the asymptotically optimal XSTPIR scheme where the noise is aligned with other interference.

Figure 1 shows that the gap between the $X = 1$ secure MDS-PIR alternative and the XSTPIR scheme is significant. Intuitively, the reason for this gap is the following. The secure MDS-PIR alternative allows the user to *decode* the artificial noise symbol which is added to the message to guarantee security. However, in the XSTPIR scheme, the user is able to decode only the desired message, and not the noise protecting it. In fact this noise is *aligned* with other interfering symbols, e.g., the noise terms protecting other message symbols, thus creating a more efficient solution. Incidentally, the alignment of noise provides another unexpected benefit, in some cases it automatically makes the scheme symmetrically secure, as explained next.

2) Symmetric Security: Capacity of Sym-XSPIR: Let us fix $T = 1$, thereby relaxing the T-privacy constraint to its minimum value for PIR. Now, suppose in addition to X-secure storage, we also include the so called 'symmetric' security constraint, that the user should learn nothing about the data besides his desired message, i.e.,

$$[\text{Sym-Security}] \qquad I(W_{[1:K]}; A_{[1:N]}^{[\theta]} \mid Q_{[1:N]}^{[\theta]}, W_\theta, \theta) = 0. \tag{24}$$

Capacity of the basic $(X = 0, T = 1, K > 1)$ Sym-PIR setting was shown in [28] to be

$$C_{\text{Sym-PIR}}(K, N) = 1 - \frac{1}{N}. \tag{25}$$

Note that there is a loss of capacity due to the additional symmetric security constraint. Furthermore, the capacity without the symmetric security constraint depends on the number of messages K while the capacity with the symmetric security constraint does not.

XSTPIR with the symmetric security constraint and with $T = 1$, in short the Sym-XSPIR setting (note that we drop the T because $T = 1$ is the degenerate case for T-privacy), reveals a surprising aspect of our XSTPIR schemes, that imposing the symmetric security constraint does not affect[6] our capacity results for $T = 1$. This is made explicit in the following corollaries for Sym-XSPIR, that match the corresponding theorems for XSTPIR.

Corollary 1:

$$C_{\text{Sym-XSPIR}}(N, K, X) \leq \left(\frac{N-X}{N}\right) C_{\text{PIR}}(N-X, K). \tag{26}$$

Corollary 2: If $N = X + 1$, then[7] for arbitrary K,

$$C_{\text{Sym-XSPIR}}(N, K, X) = \left(\frac{N-X}{N}\right) C_{\text{PIR}}(N-X, K) \tag{27}$$

$$= \frac{1}{NK}. \tag{28}$$

Corollary 3: As the number of messages $K \to \infty$, for arbitrary N, X,

$$\lim_{K\to\infty} C_{\text{Sym-XSPIR}}(N, K, X) = \lim_{K\to\infty} \left(\frac{N-X}{N}\right) C_{\text{PIR}}(N-X, K) \tag{29}$$

$$= \begin{cases} 1 - \left(\frac{X+1}{N}\right), & N > X+1 \\ 0, & N \leq X+1. \end{cases} \tag{30}$$

Corollary 4: If the number of servers, $N = 3$, and $X = 1$, then for arbitrary number of messages, K,

$$C_{\text{Sym-XSPIR}}(N = 3, K, X = 1)$$

$$= \left(\frac{N-X}{N}\right) C_{\text{PIR}}(N-X, K) \tag{31}$$

$$= \frac{2}{3}\left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{K-1}}\right)^{-1}. \tag{32}$$

The proofs of all 4 corollaries appear in Appendix B. Surprisingly, note that there is no loss of capacity in each case due to the additional symmetric security constraint. Also note that according to Corollary 4, unlike Sym-PIR, the capacity of Sym-XSPIR depends on the number of messages K for all $K > 1$.

3) Cross Subspace Alignment: Conceptually, the most intriguing aspect of the asymptotically optimal XSTPIR scheme is the extent to which it is able to align interference. Interference alignment is central to PIR [1], [32], and nearly all existing PIR constructions use some form of interference alignment. The strength of XSTPIR lies in the novel idea of cross subspace alignment, that we explain intuitively in this section through an example. Consider the setting of $X = 2$ secure and $T = 1$ PIR with $N = 5$ servers. Let w_1 be a symbol from a desired message W. For simplicity (and because identical alignments are applied to all messages), it suffices to focus on only this message for the purpose of this explanation. In order to guarantee $X = 2$ security, w_1 is mixed with 2 random noise symbols z_{11}, z_{12}, according to the following RS Code, so that the n^{th} row is stored at the n^{th} server, $n \in [1:5]$.

$$\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix} w_1 + \begin{bmatrix}\beta_1\\\beta_2\\\beta_3\\\beta_4\\\beta_5\end{bmatrix} z_{11} + \begin{bmatrix}\beta_1^2\\\beta_2^2\\\beta_3^2\\\beta_4^2\\\beta_5^2\end{bmatrix} z_{12}$$

$$\triangleq \overrightarrow{1} w_1 + \overrightarrow{\beta} z_{11} + \overrightarrow{\beta^2} z_{12}.$$

To ensure privacy, the query symbol q_θ ($q_\theta = 1$, i.e., this message is desired) is similarly mixed with a noise symbol z_1'.

$$\overrightarrow{1} q_\theta + \overrightarrow{\beta} z_1' = \overrightarrow{1} + \overrightarrow{\beta} z_1'$$

and the n^{th} row of this query vector is sent to the n^{th} server. Each server returns the product of the noisy query symbol and the noisy stored symbol, so that the user receives the 5 answers.

$$\left(\overrightarrow{1} w_1 + \overrightarrow{\beta} z_{11} + \overrightarrow{\beta^2} z_{12}\right) \circ \left(\overrightarrow{1} + \overrightarrow{\beta} z_1'\right)$$

$$= \overrightarrow{1} w_1 + \overrightarrow{\beta}\left(w_1 z_1' + z_{11}\right) + \overrightarrow{\beta^2}\left(z_{11} z_1' + z_{12}\right) + \overrightarrow{\beta^3} z_{12} z_1'.$$

The desired symbol w_1 appears along the vector $\overrightarrow{1}$ while the remaining 5 undesired symbols align along 3 dimensions. Specifically, the undesired symbols $w_1 z_1'$ and z_{11} align along the vector $\overrightarrow{\beta}$; undesired symbols $z_{11} z_1'$ and z_{12} align along the vector $\overrightarrow{\beta^2}$ and undesired symbol $z_{12} z_1'$ appears along the vector $\overrightarrow{\beta^3}$. This type of alignment, enabled by using the same $\overrightarrow{\beta}$ in the storage and query, is indeed very useful and has been used previously by Freij-Hollanti et al. for MDS-TPIR [6]. However, note that we have a 5 dimensional space (all vectors are 5×1) and we are so far only using 4 dimensions (one desired, three interference), so there is room for improvement.

In order to improve the efficiency of the retrieval scheme, suppose we try to retrieve another symbol, w_2, from the same desired message $W = (w_1, w_2)$. The challenge is that because of the $X = 2$ security requirement w_2 is mixed with new (independent) noise symbols z_{21}, z_{22} according to an RS code parameterized by γ,

$$\overrightarrow{1} w_2 + \overrightarrow{\gamma} z_{21} + \overrightarrow{\gamma^2} z_{22}, \tag{33}$$

so any attempt to retrieve w_2 will add new interference terms. Since we already have 3 dimensions of interference, the new interference added due to the noise protecting w_2 must align completely within the existing interference. This will be accomplished by cross-alignment, i.e., introducing additional structure across the storage and query codes for the different symbols to be retrieved. In particular, we will use the query vector $\overrightarrow{\gamma} \circ \left(\overrightarrow{1} + \overrightarrow{\beta} z_1'\right)$ to multiply with the stored variables containing w_1 (i.e., $\overrightarrow{1} w_1 + \overrightarrow{\beta} z_{11} + \overrightarrow{\beta^2} z_{12}$) and the query vector $\overrightarrow{\beta} \circ \left(\overrightarrow{1} + \overrightarrow{\gamma} z_2'\right)$ to multiply with the stored variables containing w_2 (i.e., $\overrightarrow{1} w_2 + \overrightarrow{\gamma} z_{21} + \overrightarrow{\gamma^2} z_{22}$). The sum of the two multiplications is returned as the answer. Note that Hadamard products are commutative and associative. The answers from

[6] For $T > 1$ our XSTPIR schemes are not symmetrically secure.

[7] Note that since $X < N$ by definition, and $T = 1$ for XSPIR, the condition $N \leq X + T$ is equivalent to $N = X + 1$.

the 5 servers are now expressed as follows.

$$\begin{aligned}&\overrightarrow{\gamma}\circ\left(\overrightarrow{1}w_1+\overrightarrow{\beta}z_{11}+\overrightarrow{\beta^2}z_{12}\right)\circ\left(\overrightarrow{1}+\overrightarrow{\beta}z_1'\right)\\&\quad+\overrightarrow{\beta}\circ(\overrightarrow{1}w_2+\overrightarrow{\gamma}z_{21}+\overrightarrow{\gamma^2}z_{22})\circ\left(\overrightarrow{1}+\overrightarrow{\gamma}z_2'\right)\quad(34)\\&=\overrightarrow{\gamma}w_1+\overrightarrow{\beta}w_2+\overrightarrow{\beta}\circ\overrightarrow{\gamma}\left(w_1z_1'+z_{11}+w_2z_2'+z_{21}\right)\\&\quad+\overrightarrow{\beta^2}\circ\overrightarrow{\gamma}(z_{11}z_1'+z_{12})+\overrightarrow{\beta}\circ\overrightarrow{\gamma^2}(z_{21}z_2'+z_{22})\\&\quad+\overrightarrow{\beta^3}\circ\overrightarrow{\gamma}z_{12}z_1'+\overrightarrow{\beta}\circ\overrightarrow{\gamma^3}z_{22}z_2'.\quad(35)\end{aligned}$$

Note that we cannot choose $\overrightarrow{\beta}=\overrightarrow{\gamma}$, because the two desired symbols (w_1, w_2) must not align in the same dimension. Also note that by cross-multiplying the first set of answers with $\overrightarrow{\gamma}$ and the second with $\overrightarrow{\beta}$ we have achieved *cross alignment* of 4 terms along $\overrightarrow{\beta}\circ\overrightarrow{\gamma}$. However, we now have 5 dimensions occupied by interference, along the 5 vectors, $\overrightarrow{\beta}\circ\overrightarrow{\gamma}, \overrightarrow{\beta^2}\circ\overrightarrow{\gamma}, \overrightarrow{\beta}\circ\overrightarrow{\gamma^2}, \overrightarrow{\beta^3}\circ\overrightarrow{\gamma}, \overrightarrow{\beta}\circ\overrightarrow{\gamma^3}$. Since the overall space is only 5 dimensional and we need two dimensions for desired symbols, we need to restrict interference to no more than 3 dimensions. Surprisingly, it is possible to do this by cross *subspace* alignment as we show next. Let us introduce a structural relationship between β and γ. In particular, let us set,

$$\overrightarrow{\beta}=\overrightarrow{1+\alpha}\quad(36)$$

$$\overrightarrow{\gamma}=\overrightarrow{2+\alpha}\quad(37)$$

so that the answers from the 5 servers are now expressed as,

$$\left(\overrightarrow{2+\alpha}\right)w_1+\left(\overrightarrow{1+\alpha}\right)w_2+\left(\overrightarrow{1+\alpha}\right)\circ\left(\overrightarrow{2+\alpha}\right)I\quad(38)$$

where the interference I is

$$\begin{aligned}I=&\overrightarrow{1}(w_1z_1'+z_{11}+w_2z_2'+z_{21})\\&+\left(\overrightarrow{1+\alpha}\right)(z_{11}z_1'+z_{12})+\left(\overrightarrow{2+\alpha}\right)(z_{21}z_1'+z_{22})\\&+\left(\overrightarrow{1+2\alpha+\alpha^2}\right)z_{12}z_1'+\left(\overrightarrow{4+4\alpha+\alpha^2}\right)z_{22}z_2'.\quad(39)\end{aligned}$$

Note that there are still 5 interference vectors, no two of which align directly with each other. However, the 5 interference vectors align into a 3 dimensional subspace of the 5 dimensional vector space. This is what we mean by *cross subspace alignment* and it is essential to this work. To see explicitly how the interference aligns into a 3 dimensional subspace, we can rewrite I as,

$$\begin{aligned}I=&\ \overrightarrow{1}(w_1z_1'+z_{11}+w_2z_2'+z_{21}+z_{11}z_1'\\&\qquad+z_{12}+2z_{21}z_1'+2z_{22}+z_{12}z_1'+4z_{22}z_2')\\&+\overrightarrow{\alpha}(z_{11}z_1'+z_{12}+z_{21}z_1'+z_{22}+2z_{12}z_1'+4z_{22}z_2')\\&+\overrightarrow{\alpha^2}(z_{12}z_1'+z_{22}z_2').\quad(40)\end{aligned}$$

Thus, due to cross subspace alignment, all of I aligns within a 3 dimensional space, leaving the remaining 2 dimensions interference-free for the desired symbols. Exactly the same alignments apply to all messages as explained in the formal descriptions of the schemes provided in this paper.

IV. Proof of Theorem 1

Let us start with two useful lemmas. The first one shows that the desired message index is independent of the messages, stored variables, queries and answers.

Lemma 1: For all $k, k' \in [1:K], \forall \mathcal{T} \in [1:N], |\mathcal{T}| = T$, we have

$$\begin{aligned}&(Q_{\mathcal{T}}^{[k]}, A_{\mathcal{T}}^{[k]}, S_{[1:N]}, W_1, \cdots, W_K)\\&\sim(Q_{\mathcal{T}}^{[k']}, A_{\mathcal{T}}^{[k']}, S_{[1:N]}, W_1, \cdots, W_K)\quad(41)\end{aligned}$$

Proof: Since $W_1, \cdots, W_K$ is a function of $S_{[1:N]}$ and $A_{\mathcal{T}}^{[\theta]}$ is a function of $(Q_{\mathcal{T}}^{[\theta]}, S_{\mathcal{T}})$ (refer to (11)), it suffices to prove $I(\theta; Q_{\mathcal{T}}^{[\theta]}, S_{[1:N]}) = 0$. From (9), we have

$$\begin{aligned}&I(Q_{[1:N]}^{[\theta]}, \theta; S_{[1:N]}) = 0 &(42)\\\Rightarrow\ &I(Q_{\mathcal{T}}^{[\theta]}, \theta; S_{[1:N]}) = 0 &(43)\\\Rightarrow\ &I(Q_{\mathcal{T}}^{[\theta]}; S_{[1:N]}) = I(Q_{\mathcal{T}}^{[\theta]}; S_{[1:N]}|\theta) = 0 &(44)\end{aligned}$$

Next, we have,

$$\begin{aligned}I(\theta; Q_{\mathcal{T}}^{[\theta]}, S_{[1:N]}) &\overset{(9)}{=} I(\theta; Q_{\mathcal{T}}^{[\theta]}|S_{[1:N]}) &(45)\\&= H(Q_{\mathcal{T}}^{[\theta]}|S_{[1:N]}) - H(Q_{\mathcal{T}}^{[\theta]}|S_{[1:N]}, \theta) &(46)\\&\overset{(44)}{=} H(Q_{\mathcal{T}}^{[\theta]}) - H(Q_{\mathcal{T}}^{[\theta]}|\theta) &(47)\\&\overset{(10)}{=} 0 &(48)\end{aligned}$$

□

The second lemma is a statement of conditional independence of answers from one set of servers from the queries to the rest of the servers.

Lemma 2: For all $\mathcal{T}, \mathcal{X} \subset [1:N], \forall k \in [1:K], \forall \mathcal{K} \in [1:K]$, we have

$$H(A_{\mathcal{T}}^{[k]}|S_{\mathcal{X}}, Q_{[1:N]}^{[k]}, W_{\mathcal{K}}) = H(A_{\mathcal{T}}^{[k]}|S_{\mathcal{X}}, Q_{\mathcal{T}}^{[k]}, W_{\mathcal{K}})\quad(49)$$

Proof: It suffices to prove that $I(A_{\mathcal{T}}^{[k]}; Q_{[1:N]}^{[k]}|S_{\mathcal{X}}, Q_{\mathcal{T}}^{[k]}, W_{\mathcal{K}}) = 0$. This proof is presented as follows.

$$\begin{aligned}&I(A_{\mathcal{T}}^{[k]}; Q_{[1:N]}^{[k]}|S_{\mathcal{X}}, Q_{\mathcal{T}}^{[k]}, W_{\mathcal{K}})\\&\leq I(A_{\mathcal{T}}^{[k]}, S_{\mathcal{X}}, W_{\mathcal{K}}; Q_{[1:N]}^{[k]}|Q_{\mathcal{T}}^{[k]}) &(50)\\&\leq I(A_{\mathcal{T}}^{[k]}, S_{[1:N]}, W_{\mathcal{K}}; Q_{[1:N]}^{[k]}|Q_{\mathcal{T}}^{[k]}) &(51)\\&\overset{(8)(11)}{=} I(S_{[1:N]}; Q_{[1:N]}^{[k]}|Q_{\mathcal{T}}^{[k]}) &(52)\\&\overset{(9)}{=} 0 &(53)\end{aligned}$$

□

The next lemma formalizes the intuition that because of the security constraint, the answers from any X servers are, in some sense, not very useful. Specifically, after conditioning on the information contained in any X servers, the answers from the remaining $N-X$ servers must still contain at least L more bits than the interference that is included in those answers. For a set $\mathcal{X}$, its complement set is denoted as $\overline{\mathcal{X}}$, i.e., $\overline{\mathcal{X}} = \{n | n \in [1:N], n \notin \mathcal{X}\}$. We use D_n to denote the expected number of symbols downloaded from Server n.

Lemma 3: For all $\mathcal{X} \in [1:N], |\mathcal{X}| = X$, we have

$$L \leq \sum_{n \in \overline{\mathcal{X}}} D_n - H(A^{[1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[1]}_{[1:N]}, W_1) \tag{54}$$

Proof:

$$L = H(W_1) \overset{(12)}{=} I(W_1; A^{[1]}_{[1:N]} | Q^{[1]}_{[1:N]}) \tag{55}$$

$$\leq I(W_1; A^{[1]}_{[1:N]}, S_{\mathcal{X}} | Q^{[1]}_{[1:N]}) \tag{56}$$

$$= I(W_1; S_{\mathcal{X}} | Q^{[1]}_{[1:N]}) + I(W_1; A^{[1]}_{\mathcal{X}}, A^{[1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[1]}_{[1:N]}) \tag{57}$$

$$\overset{(11)}{=} I(W_1; S_{\mathcal{X}} | Q^{[1]}_{[1:N]}) + I(W_1; A^{[1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[1]}_{[1:N]}) \tag{58}$$

$$\overset{(9)}{=} I(W_1, Q^{[1]}_{[1:N]}; S_{\mathcal{X}}) + I(W_1; A^{[1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[1]}_{[1:N]}) \tag{59}$$

$$\overset{(7)}{=} I(Q^{[1]}_{[1:N]}; S_{\mathcal{X}} | W_1) + I(W_1; A^{[1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[1]}_{[1:N]}) \tag{60}$$

$$\leq I(Q^{[1]}_{[1:N]}; S_{\mathcal{X}}, W_1) + I(W_1; A^{[1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[1]}_{[1:N]}) \tag{61}$$

$$\overset{(9)}{=} I(W_1; A^{[1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[1]}_{[1:N]}) \tag{62}$$

$$\leq \sum_{n \in \overline{\mathcal{X}}} D_n - H(A^{[1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[1]}_{[1:N]}, W_1) \tag{63}$$

□

We may interpret the second term of the RHS of (54) as the interference term. To bound it, we need the following recursive relation, stated in a lemma.

Lemma 4: For all $\mathcal{X} \in [1:N], |\mathcal{X}| = X$ and for all $k \in [1:K]$, we have

$$H(A^{[k]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[k]}_{[1:N]}, W_{[1:k]}) \geq \frac{T}{N-X}\left(L + H(A^{[k+1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]}, W_{[1:k+1]})\right), \text{ if } N > X + T. \tag{64}$$

$$H(A^{[k]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[k]}_{[1:N]}, W_{[1:k]}) \geq L + H(A^{[k+1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]}, W_{[1:k+1]}), \text{ if } N \leq X + T. \tag{65}$$

Proof: First consider $N > X + T$. Consider any set $\mathcal{T} \subset \overline{\mathcal{X}}, |\mathcal{T}| = T$.

$$H(A^{[k]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[k]}_{[1:N]}, W_{[1:k]})$$

$$\geq H(A^{[k]}_{\mathcal{T}} | S_{\mathcal{X}}, Q^{[k]}_{[1:N]}, W_{[1:k]}) \tag{66}$$

$$\overset{(49)}{=} H(A^{[k]}_{\mathcal{T}} | S_{\mathcal{X}}, Q^{[k]}_{\mathcal{T}}, W_{[1:k]}) \tag{67}$$

$$\overset{(41)}{=} H(A^{[k+1]}_{\mathcal{T}} | S_{\mathcal{X}}, Q^{[k+1]}_{\mathcal{T}}, W_{[1:k]}) \tag{68}$$

$$\overset{(49)}{=} H(A^{[k+1]}_{\mathcal{T}} | S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]}, W_{[1:k]}) \tag{69}$$

Averaging (69) over all choices of $\mathcal{T}$ and applying Han's inequality, we have

$$H(A^{[k]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[k]}_{[1:N]}, W_{[1:k]})$$

$$\geq \frac{T}{N-X} H(A^{[k+1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]}, W_{[1:k]}) \tag{70}$$

$$\overset{(11)(12)}{=} \frac{T}{N-X} H(A^{[k+1]}_{\overline{\mathcal{X}}}, W_{k+1} | S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]}, W_{[1:k]}) \tag{71}$$

$$= \frac{T}{N-X}\Big(H(W_{k+1} | S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]}, W_{[1:k]}) + H(A^{[k+1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]}, W_{[1:k+1]})\Big) \tag{72}$$

$$= \frac{T}{N-X}\left(L + H(A^{[k+1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]}, W_{[1:k+1]})\right) \tag{73}$$

where the last step uses $L = H(W_{k+1})$ and $I(W_{k+1}; S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]}, W_{[1:k]}) = 0$, proved as follows.

$$I(W_{k+1}; S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]}, W_{[1:k]})$$

$$\overset{(5)(6)}{=} I(W_{k+1}; S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]} \mid W_{[1:k]}) \tag{74}$$

$$\leq I(W_{[1:k+1]}; S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]}) \tag{75}$$

$$\overset{(7)}{=} I(W_{[1:k+1]}; Q^{[k+1]}_{[1:N]} | S_{\mathcal{X}}) \tag{76}$$

$$\leq I(W_{[1:k+1]}, S_{\mathcal{X}}; Q^{[k+1]}_{[1:N]}) \tag{77}$$

$$\leq I(S_{[1:N]}; Q^{[k+1]}_{[1:N]}) \tag{78}$$

$$\overset{(9)}{=} 0 \tag{79}$$

Next, consider $N \leq X + T$. The proof is similar to that presented above. Note that $|\mathcal{X}| =$ N-X $\leq T$.

$$H(A^{[k]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[k]}_{[1:N]}, W_{[1:k]})$$

$$\overset{(49)}{=} H(A^{[k]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[k]}_{\overline{\mathcal{X}}}, W_{[1:k]}) \tag{80}$$

$$\overset{(41)}{=} H(A^{[k+1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[k+1]}_{\overline{\mathcal{X}}}, W_{[1:k]}) \tag{81}$$

$$\overset{(49)}{=} H(A^{[k+1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]}, W_{[1:k]}) \tag{82}$$

$$\overset{(11)(12)}{=} H(A^{[k+1]}_{\overline{\mathcal{X}}}, W_{k+1} | S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]}, W_{[1:k]}) \tag{83}$$

$$= H(W_{k+1} | S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]}, W_{[1:k]}) + H(A^{[k+1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]}, W_{[1:k+1]}) \tag{84}$$

$$\overset{(79)}{=} L + H(A^{[k+1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[k+1]}_{[1:N]}, W_{[1:k+1]}) \tag{85}$$

This completes the proof of Lemma 4. □

Now let us apply Lemma 4 repeatedly for $k = 1, 2, \cdots$. When $N > X + T$, we have

$$H(A^{[1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[1]}_{[1:N]}, W_1)$$

$$\geq \frac{T}{N-X}\left(L + H(A^{[2]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[2]}_{[1:N]}, W_{[1:2]}\right) \tag{86}$$

$$\geq \frac{T}{N-X}\left(L + \frac{T}{N-X}\left(L + H(A^{[3]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[3]}_{[1:N]}, W_{[1:3]}\right)\right) \tag{87}$$

$$\geq \cdots \tag{88}$$

$$\geq L\left(\frac{T}{N-X} + \left(\frac{T}{N-X}\right)^2 + \cdots + \left(\frac{T}{N-X}\right)^{K-1}\right) \tag{89}$$

Similarly, when $N \leq X + T$, we have

$$H(A^{[1]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[1]}_{[1:N]}, W_1) \geq L + H(A^{[2]}_{\overline{\mathcal{X}}} | S_{\mathcal{X}}, Q^{[2]}_{[1:N]}, W_{[1:2]}) \tag{90}$$

$$\geq \cdots \tag{91}$$

$$\geq L(K-1) \tag{92}$$

Substituting (89), (92) into (54), we have

$$L \leq \sum_{n\in\overline{\mathcal{X}}} D_n - L\left(\frac{T}{N-X} + \left(\frac{T}{N-X}\right)^2 + \cdots + \left(\frac{T}{N-X}\right)^{K-1}\right), \text{ if } N > X+T. \tag{93}$$

$$L \leq \sum_{n\in\overline{\mathcal{X}}} D_n - L(K-1), \text{ if } N \leq X+T. \tag{94}$$

Averaging over all $\mathcal{X}$, we have

$$L \leq \left(\frac{N-X}{N}\right) D - L\left(\frac{T}{N-X} + \left(\frac{T}{N-X}\right)^2 + \cdots + \left(\frac{T}{N-X}\right)^{K-1}\right), \text{ if } N > X+T. \tag{95}$$

$$L \leq \left(\frac{N-X}{N}\right) D - L(K-1), \text{ if } N \leq X+T. \tag{96}$$

Finally since the rate is defined as $R = L/D$, we arrive at the final bound.

$$R \leq \frac{N-X}{N}\left(1 + \frac{T}{N-X} + \left(\frac{T}{N-X}\right)^2 + \cdots + \left(\frac{T}{N-X}\right)^{K-1}\right)^{-1}, \text{ if } N > X+T. \tag{97}$$

$$R \leq \frac{N-X}{N} \times \frac{1}{K}, \text{ if } N \leq X+T. \tag{98}$$

Thus

$$C_{\text{XSTPIR}}(N,K,X,T) \leq \left(\frac{N-X}{N}\right) C_{\text{TPIR}}(N-X,K,T), \tag{99}$$

and the proof of Theorem 1 is complete. □

V. Proof of Theorem 2

Let each message consist of $L = N - X$ symbols in $\mathbb{F}_q$, $q \geq N$, and append X instances of 0 symbols, to create artificial messages of length N,

$$\bar{W}_k = (W_{k1}, W_{k_2}, \cdots, W_{k(N-X)}, \underbrace{0, 0, \cdots, 0}_{X}), \quad \forall k \in [1:K]. \tag{100}$$

Corresponding to each message W_k, let $Z_k = (Z_{k1}, Z_{k2}, \cdots, Z_{kX}) \in \mathbb{F}_q^X$ be X independent uniform noise symbols, to be used for X-security. Let Z_k be encoded with an (N, X) MDS code to produce $\bar{Z}_k \in \mathbb{F}_q^N$. For each $k \in [1:K]$ and $n \in [1:N]$, the n^{th} server stores the n^{th} symbol of $\bar{W}_k + \bar{Z}_k$. Thus, each server stores a total of K symbols. The MDS property of $\bar{Z}_k$ ensures that the data storage is X-secure. Retrieval is trivial — in order to retrieve the desired message W_θ, the user simply downloads everything from all servers. Since the queries do not depend on the desired message, the scheme is N-private, so it is also T-private. The rate achieved is $\frac{N-X}{NK}$ which matches the capacity for this setting. □

VI. Proof of Theorem 3

Let us present an XSTPIR scheme for arbitrary X, T, N, K, that is asymptotically optimal (as $K \to \infty$). The asymptotic capacity is zero for $N \leq X + T$, so we only need to consider $N > X + T$. Throughout this scheme we will set

$$L = N - X - T \tag{101}$$

and we will use the compact notation,

$$\Delta = \prod_{i=1}^{L}(i+\alpha). \tag{102}$$

Δ_n will represent the value of Δ when α is replaced with α_n.

Each message $W_k, k \in [1:K]$, consists of $L = N - X - T$ symbols, $W_k = (W_{k1}, W_{k2}, \cdots, W_{kL})$ from a finite field $\mathbb{F}_q$. The field $\mathbb{F}_q$ is assumed to have size[8] $q \geq L + N$, and characteristic greater than $L-1$. For the design of this scheme, we will need constants $\alpha_n, n \in [1:N]$ that are distinct elements of $\mathbb{G}$,

$$\mathbb{G} = \{\alpha \in \mathbb{F}_q : \alpha + i \neq 0, \ \forall i \in [1:L]\}. \tag{103}$$

Such $\alpha_n, n \in [1:N]$ must exist because $q \geq L + N$. These constants will be globally known. In the following description of the scheme, we will explain explicitly how the values of these constants are chosen. For now, let us note that because the characteristic of the field is assumed to be greater than $L - 1$, the values $\alpha + 1, \alpha + 2, \cdots, \alpha + L$ are distinct for any $\alpha \in \mathbb{F}_q$.

Let us split the messages into L vectors, so that $\mathbf{W}_l = (W_{1l}, W_{2l}, \cdots, W_{Kl})$, $l \in [1:L]$, contains the l^{th} symbol of every message. Let $\mathbf{Z}_{lx}, l \in [1:L], x \in [1:X]$, be independent uniformly random noise vectors from $\mathbb{F}_q^{1\times K}$, that are used to guarantee security. Similarly, let $\mathbf{Z}'_{lt}, l \in [1:L], t \in [1:T]$, be independent uniformly random noise vectors from $\mathbb{F}_q^{K\times 1}$, that are used to guarantee privacy. The independence between noise vectors, messages, and the user's desired message index θ is specified as follows.

$$\begin{aligned}&H\Big((\mathbf{W}_l)_{l\in[1:L]}, (\mathbf{Z}_{lx})_{l\in[1:L],x\in[1:X]}, (\mathbf{Z}'_{lt})_{l\in[1:L],t\in[1:T]}, \theta\Big)\\ =&H((\mathbf{W}_l)_{l\in[1:L]}) + H(\theta) + KL(X+T)\end{aligned} \tag{104}$$

in q-ary units. Let $\mathbf{Q}_\theta$ represent[9] the θ^{th} column of the $K \times K$ identity matrix, so it contains a 1 in the θ^{th} position and zeros everywhere else. Note that

$$\begin{aligned}&(\mathbf{W}_1\mathbf{Q}_\theta, \mathbf{W}_2\mathbf{Q}_\theta, \cdots, \mathbf{W}_L\mathbf{Q}_\theta)\\ =&(W_{\theta 1}, W_{\theta 2}, \cdots, W_{\theta L}) = W_\theta\end{aligned} \tag{105}$$

[8] In other words, we set $q = p^n$ for a prime number p and an integer $n \geq 1$ such that $p^n \geq L + N$ and $p \geq L$. While this makes the scheme more general, let us note that for simplicity it may be desirable to choose $n = 1$ and $q = p \geq L + N$. On the other hand, the general scheme is useful for extensions of this work, say to private computation (see Footnote 9), where the choice of field may be fixed by the functions that need to be computed.

[9] Note that the XSTPIR scheme described in this section works even if $\mathbf{Q}_\theta$ is an arbitrary vector, i.e., if instead of retrieving one of the K messages, the user wishes to compute an arbitrary linear function of the K messages over $\mathbb{F}_q$. Thus, the scheme automatically settles the asymptotic capacity of the natural X-secure and T-private generalization of the linear private computation problem introduced in [33] (also known as linear private function retrieval [34]).

is the message desired by the user. A succinct summary of the storage at each server, the queries, and a partitioning of signal and interference dimensions contained in the answers from each server, is provided below.

	Server 'n' (Replace α, Δ with α_n, Δ_n)
Storage	$\mathbf{W}_1 + (1+\alpha)\mathbf{Z}_{11} + \cdots + (1+\alpha)^X \mathbf{Z}_{1X},$
(S_n)	$\mathbf{W}_2 + (2+\alpha)\mathbf{Z}_{21} + \cdots + (2+\alpha)^X \mathbf{Z}_{2X},$
	$\vdots$
	$\mathbf{W}_L + (L+\alpha)\mathbf{Z}_{L1} + \cdots + (L+\alpha)^X \mathbf{Z}_{LX}$
Query	$\frac{\Delta}{1+\alpha}\Big(\mathbf{Q}_\theta + (1+\alpha)\mathbf{Z}'_{11} + \cdots + (1+\alpha)^T \mathbf{Z}'_{1T}\Big),$
($Q_n^{[\theta]}$)	$\frac{\Delta}{2+\alpha}\Big(\mathbf{Q}_\theta + (2+\alpha)\mathbf{Z}'_{21} + \cdots + (2+\alpha)^T \mathbf{Z}'_{2T}\Big),$
	$\vdots$
	$\frac{\Delta}{L+\alpha}\Big(\mathbf{Q}_\theta + (L+\alpha)\mathbf{Z}'_{L1} + \cdots + (L+\alpha)^T \mathbf{Z}'_{LT}\Big)$
	Desired symbols appear along vectors
	$\overrightarrow{\Delta} \circ \Big(\overrightarrow{(1+\alpha)^{-1}}, \overrightarrow{(2+\alpha)^{-1}}, \cdots, \overrightarrow{(L+\alpha)^{-1}}\Big)$
	Interference appears along vectors
	$\overrightarrow{\Delta} \circ \Big(\overrightarrow{1}, \overrightarrow{(1+\alpha)}, \cdots, \overrightarrow{(1+\alpha)^{X+T-1}},$
	$\overrightarrow{(2+\alpha)}, \cdots, \overrightarrow{(2+\alpha)^{X+T-1}}, \cdots,$
	$\cdots, \overrightarrow{(L+\alpha)}, \cdots, \overrightarrow{(L+\alpha)^{X+T-1}}\Big)$

Initially, the user knows only his desired message index θ and the noise terms $\mathbf{Z}'_{lt}, l \in [1:L], t \in [1:T]$, all of which are privately generated by the user. Each server $n \in [1:N]$ knows only its stored information S_n. The storage S_n at Server n may be viewed as a $1 \times LK$ row vector formed by concatenating the L row vectors, $\mathbf{W}_l + \sum_{x=1}^{X}(l+\alpha_n)^x \mathbf{Z}_{lx}$, $l \in [1:L]$. Similarly, the query $Q_n^{[\theta]}$ may be viewed as an $LK \times 1$ column vector formed by concatenating the L column vectors, $\frac{\Delta_n}{l+\alpha_n}\Big(\mathbf{Q}_\theta + \sum_{t=1}^{T}(l+\alpha_n)^t \mathbf{Z}'_{lt}\Big)$, $l \in [1:L]$.

Upon receiving the query $Q_n^{[\theta]}$ from the user, Server n responds with the answer $A_n^{[\theta]}$ that is exactly one symbol in $\mathbb{F}_q$, found by multiplying S_n with $Q_n^{[\theta]}$.

$$A_n^{[\theta]} = S_n Q_n^{[\theta]}. \tag{106}$$

This produces a single equation in a total of $L(X+1)(T+1)$ terms. Out of these, L terms are desired message symbols $\mathbf{W}_l\mathbf{Q}_\theta$, $l \in [1:L]$, and the remaining $L(X+1)(T+1)-L$ terms are undesired, or interference terms. The interference terms include LT terms of the type $\mathbf{W}_l\mathbf{Z}'_{lt}$, LX terms of the type $\mathbf{Z}_{lx}\mathbf{Q}_\theta$, and LXT terms of the type $\mathbf{Z}_{lx}\mathbf{Z}'_{lt}$. The user obtains one such equation from each server, for a total of N equations, from which he must be able to retrieve his L desired symbols. The key to this is the alignment of $L(X+1)(T+1)-L$ interference terms into $N-L$ dimensions, leaving L dimensions free from interference from which the L desired symbols can be decoded.

First let us identify the desired signal dimensions, i.e., the vectors along which desired symbols are seen by the user. Each answer $A_n^{[\theta]}$ contains the desired symbols $\frac{\Delta_n}{l+\alpha_n}\mathbf{W}_l\mathbf{Q}_\theta = \frac{\Delta_n}{l+\alpha_n}W_{\theta l}$, $l \in [1:L]$. These L desired symbols appear along the following L vectors.

$$\begin{bmatrix} \frac{\Delta_1}{1+\alpha_1} \\ \frac{\Delta_2}{1+\alpha_2} \\ \vdots \\ \frac{\Delta_N}{1+\alpha_N} \end{bmatrix}, \begin{bmatrix} \frac{\Delta_1}{2+\alpha_1} \\ \frac{\Delta_2}{2+\alpha_2} \\ \vdots \\ \frac{\Delta_N}{2+\alpha_N} \end{bmatrix}, \cdots, \begin{bmatrix} \frac{\Delta_1}{L+\alpha_1} \\ \frac{\Delta_2}{L+\alpha_2} \\ \vdots \\ \frac{\Delta_N}{L+\alpha_N} \end{bmatrix}$$
$$\triangleq \overrightarrow{\Delta} \circ \Big(\overrightarrow{(1+\alpha)^{-1}}, \overrightarrow{(2+\alpha)^{-1}}, \cdots, \overrightarrow{(L+\alpha)^{-1}}\Big). \tag{107}$$

Recall that $\circ$ represents the Hadamard product. Similarly, the vectors along which interference symbols appear are identified as follows.

$$\overrightarrow{\Delta} \circ \Big(\overrightarrow{1}, \overrightarrow{(1+\alpha)}, \cdots, \overrightarrow{(1+\alpha)^{X+T-1}}, \overrightarrow{(2+\alpha)}, \cdots, \overrightarrow{(2+\alpha)^{X+T-1}}, \cdots, \cdots, \overrightarrow{(L+\alpha)}, \cdots, \overrightarrow{(L+\alpha)^{X+T-1}}\Big). \tag{108}$$

Thus, the vector of answers from all N servers can be expressed as

$$\overrightarrow{A^{[\theta]}} = \sum_{l=1}^{L} W_{\theta l}\overrightarrow{\Delta} \circ \overrightarrow{(l+\alpha)^{-1}} + \sum_{l=1}^{L}\sum_{i=0}^{X+T-1} \overrightarrow{\Delta} \circ \overrightarrow{(l+\alpha)^i} I_{li} \tag{109}$$

for some interference terms I_{li} that are sums of various $\mathbf{W}_l\mathbf{Z}'_{lt}$, $\mathbf{Z}_{lx}\mathbf{Q}_\theta$, and $\mathbf{Z}_{lx}\mathbf{Z}'_{lt}$ terms. The exact form of I_{li} terms is not important for our analysis. Using binomial expansion to write each $\overrightarrow{(l+\alpha)^i}$ vector as $\sum_{j=0}^{i}\binom{i}{j}l^j\overrightarrow{\alpha^{i-j}}$, and grouping terms by the vectors $\overrightarrow{\alpha}^i$, we can write,

$$\overrightarrow{A^{[\theta]}} = \sum_{l=1}^{L} W_{\theta l}\overrightarrow{\Delta} \circ \overrightarrow{(l+\alpha)^{-1}} + \sum_{i=0}^{X+T-1} \overrightarrow{\Delta} \circ \overrightarrow{\alpha}^i I'_i. \tag{110}$$

Thus, all interference is aligned within the subspace spanned by vectors $\overrightarrow{\Delta}$, $\overrightarrow{\Delta} \circ \overrightarrow{\alpha}$, $\ldots$, $\overrightarrow{\Delta} \circ \overrightarrow{\alpha^{X+T-1}}$. As explained in Section III-B.3, this is because of *cross subspace alignment*.

In matrix notation, we have,

$$\overrightarrow{A^{[\theta]}} = \begin{bmatrix} A_1^{[\theta]} \\ A_2^{[\theta]} \\ \vdots \\ A_N^{[\theta]} \end{bmatrix} = M_N \begin{bmatrix} W_{\theta 1} \\ \vdots \\ W_{\theta L} \\ I'_0 \\ \vdots \\ I'_{(X+T-1)} \end{bmatrix} \tag{111}$$

where the $N \times N$ square matrix (note that $L+X+T=N$)

$$M_N = \begin{bmatrix} \frac{\Delta_1}{1+\alpha_1} & \cdots & \frac{\Delta_1}{L+\alpha_1} & \Delta_1 & \Delta_1\alpha_1 & \cdots & \Delta_1\alpha_1^{X+T-1} \\ \frac{\Delta_2}{1+\alpha_2} & \cdots & \frac{\Delta_2}{L+\alpha_2} & \Delta_2 & \Delta_2\alpha_2 & \cdots & \Delta_2\alpha_2^{X+T-1} \\ \vdots & & & & & & \\ \frac{\Delta_N}{1+\alpha_N} & \cdots & \frac{\Delta_N}{L+\alpha_N} & \Delta_N & \Delta_N\alpha_N & \cdots & \Delta_N\alpha_N^{X+T-1} \end{bmatrix} \tag{112}$$

$$=\Big[\overrightarrow{\Delta}\circ\overrightarrow{(1+\alpha)^{-1}}\;\cdots\;\overrightarrow{\Delta}\circ\overrightarrow{(L+\alpha)^{-1}}$$

$$\overrightarrow{\Delta}\;\;\overrightarrow{\Delta}\circ\overrightarrow{\alpha}\;\cdots\;\overrightarrow{\Delta}\circ\overrightarrow{\alpha^{X+T-1}}\Big] \tag{113}$$

is called the decoding matrix. Evidently, if the decoding matrix is invertible, then the user can recover his L desired message symbols. We show that if $\alpha_n, n \in [1:N]$ are distinct elements of $\mathbb{G}$, then M_N is invertible. This result is stated in the following lemma. Note that in our design, we have chosen α_n as distinct elements, so Lemma 5 guarantees that the scheme satisfies the correctness constraint. Fixing distinct values of $\alpha_1, \cdots, \alpha_N$ completes the design of the scheme.

Lemma 5: The decoding matrix M_N is invertible if all $\alpha_n, n \in [1:N]$ are distinct.

Proof: To set up the proof by contradiction, suppose on the contrary that M_N is singular. Then there must exist $c_n \in \mathbb{F}_q, n \in [1:N]$, at least one of which is non-zero, such that

$$c_1\overrightarrow{\Delta}\circ\overrightarrow{(1+\alpha)^{-1}}+\cdots+c_L\overrightarrow{\Delta}\circ\overrightarrow{(L+\alpha)^{-1}}$$

$$+c_{L+1}\overrightarrow{\Delta}+c_{L+2}\overrightarrow{\Delta}\circ\overrightarrow{\alpha}+\cdots+c_N\overrightarrow{\Delta}\circ\overrightarrow{\alpha^{X+T-1}}=\overrightarrow{0} \tag{114}$$

where $\overrightarrow{0}$ is the vector whose elements are all 0. Now consider n-th row of (114).

$$c_1\frac{\Delta_n}{1+\alpha_n}+\cdots+c_L\frac{\Delta_n}{L+\alpha_n}$$

$$+c_{L+1}\Delta_n+c_{L+2}\Delta_n\alpha_n+\cdots+c_N\Delta_n\alpha_n^{X+T-1}=0. \tag{115}$$

From (102) and (103), we know that $\Delta_n \neq 0$. Then α_n must be the root of the following polynomial

$$g(\alpha)=\sum_{i=1}^{L}c_i\left(\frac{\Delta}{i+\alpha}\right)+\sum_{i=L+1}^{N}c_i\Delta\alpha^{i-(L+1)} \tag{116}$$

Note that Δ (as a function of α) has order L and $i+\alpha$ is a factor of Δ (refer to (102)), so $g(\alpha)$ has order *at most* $N-1$. If $g(\alpha)$ is a non-zero polynomial, then it can have at most $N-1$ roots over $\mathbb{F}_q$. Now $\alpha_n, n \in [1:N]$ are N distinct roots of $g(\alpha)$, thus $g(\alpha)$ must be the zero polynomial, i.e., the coefficients of all monomials in $g(\alpha)$ must be zero. The coefficient of α^{N-1} is c_N so we must have $c_N = 0$. Then, the remaining coefficient of α^{N-2} is c_{N-1}, so we must have $c_{N-1} = 0$. Similarly, we find $c_{L+1} = c_{L+2} = \cdots = c_N = 0$, leaving us with

$$g(\alpha)=\sum_{i=1}^{L}c_i\left(\frac{\Delta}{i+\alpha}\right). \tag{117}$$

Now, if this $g(\alpha)$ is the zero polynomial, then it must be zero for every $\alpha \in \mathbb{F}_q$. Choosing α such[10] that $(i+\alpha) = 0$, gives us $c_i = 0$ for every $i \in [1:L]$. Thus, we have $c_1 = c_2 = \cdots = c_N = 0$. This is a contradiction since we assumed that at least one of $c_n, n \in [1:N]$ is non-zero. Thus, the proof is complete. □

Now consider the security guarantee. For any X colluding servers, $i_1, i_2, \cdots, i_X$, the X observations, $U_{kl1}, \cdots, U_{klX}$, of each message symbol $W_{kl}, k \in [1:K], l \in [1:L]$, are protected by noise terms as follows.

$$\begin{bmatrix}U_{kl1}\\ \vdots\\ U_{klX}\end{bmatrix}=\begin{bmatrix}W_{kl}\\ \vdots\\ W_{kl}\end{bmatrix}+\underbrace{\begin{bmatrix}l+\alpha_{i_1} & (l+\alpha_{i_1})^2 & \cdots & (l+\alpha_{i_1})^X\\ l+\alpha_{i_2} & (l+\alpha_{i_2})^2 & \cdots & (l+\alpha_{i_2})^X\\ \vdots & \vdots & & \vdots\\ l+\alpha_{i_X} & (l+\alpha_{i_X})^2 & \cdots & (l+\alpha_{i_X})^X\end{bmatrix}}_{P}\underbrace{\begin{bmatrix}\mathbf{Z}_{l1}(k)\\ \vdots\\ \mathbf{Z}_{lX}(k)\end{bmatrix}}_{Z} \tag{118}$$

$$=W_{kl}\underbrace{\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}}_{\mathbf{1}}+\begin{bmatrix}l+\alpha_{i_1} & 0 & \cdots & 0\\ 0 & l+\alpha_{i_2} & \cdots & 0\\ & & & \\ 0 & 0 & \ddots & 0\\ 0 & 0 & \cdots & l+\alpha_{i_X}\end{bmatrix}$$

$$\begin{bmatrix}1 & l+\alpha_{i_1} & \cdots & (l+\alpha_{i_1})^{X-1}\\ 1 & l+\alpha_{i_2} & \cdots & (l+\alpha_{i_2})^{X-1}\\ \vdots & \vdots & & \vdots\\ 1 & l+\alpha_{i_X} & \cdots & (l+\alpha_{i_X})^{X-1}\end{bmatrix}\begin{bmatrix}\mathbf{Z}_{l1}(k)\\ \vdots\\ \mathbf{Z}_{lX}(k)\end{bmatrix}. \tag{119}$$

where $\mathbf{Z}_{lx}(k)$ is the k^{th} element of the vector $\mathbf{Z}_{lx}$. Note that P is a product of a diagonal matrix which is invertible because $(l+\alpha_{i_j})$ are non-zero, and a Vandermonde matrix which is invertible because $(l+\alpha_{i_j})$ are distinct. Therefore, P is invertible, and the observations are independent of the message symbols as shown below.

$$\begin{aligned}&I(W_{kl};(U_{klx})_{x\in[1:X]})\\ =&I(W_{kl};W_{kl}\mathbf{1}+PZ)\\ =&I(W_{kl};W_{kl}P^{-1}\mathbf{1}+Z)=I(W_{kl};Z)=0.\end{aligned} \tag{120}$$

Furthermore, since the noise terms protecting each message symbol $W_{kl}, k \in [1:K], l \in [1:L]$, i.e., $\mathbf{Z}_{lx}(k), x \in [1:X]$ are independent across (k, l, x), security is preserved for all data.

The noise terms protecting each *query* also have the same structure and independence properties by design. Therefore, it follows from the same reasoning that user's privacy is protected from any T colluding servers.

Finally, note that the user is able to retrieve $L = N - X - T$ desired q-ary symbols by downloading N q-ary symbols, one from each server. The rate achieved is $L/N = 1-(X+T)/N$, which is the asymptotic capacity for this general setting. This completes the proof of Theorem 3. □

A. Example: (X = 1) Secure, (T = 1) Private Scheme With N = 5 Servers

Each message consists of $L = 3$ symbols from a finite field $\mathbb{F}_q$, $q \geq N + L = 8$, and characteristic greater than 2. For this

[10] Note that $\frac{\Delta}{i+\alpha}$ is simply a compact notation for $\prod_{l\in[1:L],l\neq i}(l+\alpha)$, i.e., it only means that the $(i+\alpha)$ factor is eliminated from Δ, so there is no 'division by 0' when we set $i+\alpha=0$ in $\frac{\Delta}{i+\alpha}$.

setting, $\Delta=(1+\alpha)(2+\alpha)(3+\alpha)$.

	Server 'n' (Replace α, Δ with α_n, Δ_n)
Storage (S_n)	$\mathbf{W}_1+(1+\alpha)\mathbf{Z}_1$, $\mathbf{W}_2+(2+\alpha)\mathbf{Z}_2$, $\mathbf{W}_3+(3+\alpha)\mathbf{Z}_3$
Query $(Q_n^{[\theta]})$	$\frac{\Delta}{1+\alpha}\Big(\mathbf{Q}_\theta+(1+\alpha)\mathbf{Z}_1'\Big)$, $\frac{\Delta}{2+\alpha}\Big(\mathbf{Q}_\theta+(2+\alpha)\mathbf{Z}_2'\Big)$, $\frac{\Delta}{3+\alpha}\Big(\mathbf{Q}_\theta+(3+\alpha)\mathbf{Z}_3'\Big)$
Desired symbols appear along vectors $\overrightarrow{\Delta}\circ\Big(\overrightarrow{(1+\alpha)^{-1}},\overrightarrow{(2+\alpha)^{-1}},\overrightarrow{(3+\alpha)^{-1}}\Big)$ Interference symbols appear along vectors $\overrightarrow{\Delta}\circ\Big(\overrightarrow{1},\overrightarrow{1+\alpha},\overrightarrow{2+\alpha},\overrightarrow{3+\alpha}\Big)$	

The answers from all $N=5$ servers may be written explicitly as,

$$\begin{aligned}
&\overrightarrow{A^{[\theta]}}\\
&=\overrightarrow{\Delta}\circ\overrightarrow{(1+\alpha)^{-1}}\mathbf{W}_1\mathbf{Q}_\theta+\overrightarrow{\Delta}\circ\overrightarrow{(2+\alpha)^{-1}}\mathbf{W}_2\mathbf{Q}_\theta\\
&\quad+\overrightarrow{\Delta}\circ\overrightarrow{(3+\alpha)^{-1}}\mathbf{W}_3\mathbf{Q}_\theta\\
&\quad+\overrightarrow{\Delta}\underbrace{\big(\mathbf{W}_1\mathbf{Z}_1'+\mathbf{W}_2\mathbf{Z}_2'+\mathbf{W}_3\mathbf{Z}_3'+\mathbf{Z}_1\mathbf{Q}_\theta+\mathbf{Z}_2\mathbf{Q}_\theta+\mathbf{Z}_3\mathbf{Q}_\theta\big)}_{I_{10}+I_{20}+I_{30}}\\
&\quad+\overrightarrow{\Delta}\circ\overrightarrow{(1+\alpha)}\underbrace{(\mathbf{Z}_1\mathbf{Z}_1')}_{I_{11}}+\overrightarrow{\Delta}\circ\overrightarrow{(2+\alpha)}\underbrace{(\mathbf{Z}_2\mathbf{Z}_2')}_{I_{21}}\\
&\quad+\overrightarrow{\Delta}\circ\overrightarrow{(3+\alpha)}\underbrace{(\mathbf{Z}_3\mathbf{Z}_3')}_{I_{31}} \qquad (121)\\
&=\overrightarrow{\Delta}\circ\overrightarrow{(1+\alpha)^{-1}}W_{\theta1}+\overrightarrow{\Delta}\circ\overrightarrow{(2+\alpha)^{-1}}W_{\theta2}\\
&\quad+\overrightarrow{\Delta}\circ\overrightarrow{(3+\alpha)^{-1}}W_{\theta3}\\
&\quad+\overrightarrow{\Delta}(I_{10}+I_{20}+I_{30}+I_{11}+2I_{21}+3I_{31})\\
&\quad+\overrightarrow{\Delta}\circ\overrightarrow{\alpha}(I_{11}+I_{21}+I_{31}). \qquad (122)
\end{aligned}$$

Privacy and security are guaranteed since $1+\alpha_n\neq 0,\forall n\in[1:5]$, the messages and queries are hidden behind the noise.

Interference terms align into the space spanned by the two vectors, $\overrightarrow{\Delta},\overrightarrow{\Delta}\circ\overrightarrow{\alpha}$, while the 3 symbols of the desired message appear along $\overrightarrow{\Delta}\circ\overrightarrow{(1+\alpha)^{-1}},\overrightarrow{\Delta}\circ\overrightarrow{(2+\alpha)^{-1}},\overrightarrow{\Delta}\circ\overrightarrow{(3+\alpha)^{-1}}$. Independence of the 3 desired signal dimensions from the two interference dimensions is trivially verified, because the highest exponent of α along desired signal dimensions is 2, but each interference dimension has an α^3 term (contributed by Δ). Independence of the 3 desired signal dimensions among themselves is also easily verified, because for

$$c_1(2+\alpha)(3+\alpha)+c_2(1+\alpha)(3+\alpha)+c_3(1+\alpha)(2+\alpha) \qquad (123)$$

to be the zero polynomial it must be zero everywhere, but in that case, setting $\alpha+i=0$ for $i=1,2,3$, leads us to $c_1=c_2=c_3=0$, thus proving their independence. The rate achieved is $3/5$, which matches the asymptotic capacity for this setting.

B. Example: ($X=2$) Secure, ($T=1$) Private Scheme With $N=4$ Servers

Each message consists of $L=1$ symbol from a finite field $\mathbb{F}_q$, $q\geq N+L=5$. $\Delta=(1+\alpha)$.

	Server 'n' (Replace α, Δ with α_n, Δ_n)
Storage (S_n)	$\mathbf{W}_1+(1+\alpha)\mathbf{Z}_{11}+(1+\alpha)^2\mathbf{Z}_{12}$
Query $(Q_n^{[\theta]})$	$\mathbf{Q}_\theta+(1+\alpha)\mathbf{Z}_1'$
Desired symbols appear along vector $\overrightarrow{1}$ Interference symbols appear along vectors $\overrightarrow{\Delta}\circ\Big(\overrightarrow{1},\overrightarrow{1+\alpha},\overrightarrow{(1+\alpha)^2}\Big)$	

The answers from all $N=4$ servers may be written explicitly as,

$$\begin{aligned}
&\overrightarrow{A^{[\theta]}}\\
&=\overrightarrow{1}\mathbf{W}_1\mathbf{Q}_\theta+\overrightarrow{(1+\alpha)}\big(\mathbf{W}_1\mathbf{Z}_1'+\mathbf{Z}_{11}\mathbf{Q}_\theta\big)\\
&\quad+\overrightarrow{(1+\alpha)^2}\big(\mathbf{Z}_{11}\mathbf{Z}_1'+\mathbf{Z}_{12}\mathbf{Q}_\theta\big)+\overrightarrow{(1+\alpha)^3}\mathbf{Z}_{12}\mathbf{Z}_1' \qquad (124)\\
&=\overrightarrow{1}W_{\theta1}+\overrightarrow{\Delta}\circ\Bigg(\overrightarrow{1}\underbrace{\big(\mathbf{W}_1\mathbf{Z}_1'+\mathbf{Z}_{11}\mathbf{Q}_\theta\big)}_{I_{10}}\\
&\qquad+\overrightarrow{(1+\alpha)}\underbrace{\big(\mathbf{Z}_{11}\mathbf{Z}_1'+\mathbf{Z}_{12}\mathbf{Q}_\theta\big)}_{I_{11}}+\overrightarrow{(1+\alpha)^2}\underbrace{\mathbf{Z}_{12}\mathbf{Z}_1'}_{I_{12}}\Bigg) \qquad (125)\\
&=\overrightarrow{1}W_{\theta1}+\overrightarrow{\Delta}\underbrace{(I_{10}+I_{11}+I_{12})}_{I_0'}+\overrightarrow{\Delta}\circ\overrightarrow{\alpha}\underbrace{(I_{11}+2I_{12})}_{I_1'}\\
&\quad+\overrightarrow{\Delta}\circ\overrightarrow{\alpha^2}\underbrace{(I_{12})}_{I_2'}. \qquad (126)
\end{aligned}$$

Interference aligns in the space spanned by the three vectors, $\overrightarrow{\Delta},\overrightarrow{\Delta}\circ\overrightarrow{\alpha},\overrightarrow{\Delta}\circ\overrightarrow{\alpha^2}$, while the desired symbol appears along the vector of all ones. The independence of these directions is easily established. Privacy is guaranteed because $1+\alpha_n\neq 0$, $\forall n\in[1:4]$, so the queries are hidden behind random noise. Security is guaranteed because for any $X=2$ colluding servers, i and j, the independent noise protecting each message W_k, $k\in[1:K]$,

$$\begin{aligned}
&\underbrace{\begin{bmatrix}1+\alpha_i & (1+\alpha_i)^2\\ 1+\alpha_j & (1+\alpha_j)^2\end{bmatrix}}_{P_{ij}}\begin{bmatrix}\mathbf{Z}_{11}(k)\\ \mathbf{Z}_{12}(k)\end{bmatrix}\\
&=\begin{bmatrix}1+\alpha_i & 0\\ 0 & 1+\alpha_j\end{bmatrix}\begin{bmatrix}1 & (1+\alpha_i)\\ 1 & (1+\alpha_j)\end{bmatrix}\begin{bmatrix}\mathbf{Z}_{11}(k)\\ \mathbf{Z}_{12}(k)\end{bmatrix} \qquad (127)
\end{aligned}$$

spans $X=2$ dimensions, because P_{ij} is invertible for distinct and non-zero values of $(1+\alpha_i),(1+\alpha_j)$. The rate achieved is $1/4$ which matches the asymptotic capacity for this setting.

C. Example: $(X = 1)$ Secure, $(T = 2)$ Private Scheme With $N = 5$ Servers

Each message consists of $L = N - X - T = 2$ symbols from $\mathbb{F}_q$, $q \geq 7$.

$$\Delta = (1+\alpha)(2+\alpha). \tag{128}$$

	Server 'n' (Replace α, Δ with α_n, Δ_n)
Storage	$\mathbf{W}_1 + (1+\alpha)\mathbf{Z}_1,$
(S_n)	$\mathbf{W}_2 + (2+\alpha)\mathbf{Z}_2$
Query	$\frac{\Delta}{1+\alpha}\Big(\mathbf{Q}_\theta + (1+\alpha)\mathbf{Z}'_{11} + (1+\alpha)^2\mathbf{Z}'_{12}\Big),$
$(Q_n^{[\theta]})$	$\frac{\Delta}{2+\alpha}\Big(\mathbf{Q}_\theta + (2+\alpha)\mathbf{Z}'_{21} + (2+\alpha)^2\mathbf{Z}'_{22}\Big)$
	Desired symbols appear along vectors
	$\overrightarrow{\Delta} \circ \Big(\overrightarrow{(1+\alpha)^{-1}}, \overrightarrow{(2+\alpha)^{-1}}\Big)$
	Interference symbols appear along vectors
	$\overrightarrow{\Delta} \circ \Big(\overrightarrow{1}, \overrightarrow{1+\alpha}, \overrightarrow{(1+\alpha)^2}, \overrightarrow{2+\alpha}, \overrightarrow{(2+\alpha)^2}\Big)$

The answers from all $N = 5$ servers may be written explicitly as,

$$\begin{aligned}
&\overrightarrow{A^{[\theta]}} \\
&= \overrightarrow{\Delta} \circ \overrightarrow{(1+\alpha)^{-1}}\mathbf{W}_1\mathbf{Q}_\theta + \overrightarrow{\Delta} \circ \overrightarrow{(2+\alpha)^{-1}}\mathbf{W}_2\mathbf{Q}_\theta \qquad (129)\\
&\quad + \overrightarrow{\Delta}\underbrace{\left(\mathbf{W}_1\mathbf{Z}'_{11} + \mathbf{W}_2\mathbf{Z}'_{21} + \mathbf{Z}_1\mathbf{Q}_\theta + \mathbf{Z}_2\mathbf{Q}_\theta\right)}_{I_{10}+I_{20}} \\
&\quad + \overrightarrow{\Delta} \circ \overrightarrow{(2+\alpha)^2}\underbrace{\mathbf{Z}_2\mathbf{Z}'_{22}}_{I_{22}} \\
&\quad + \overrightarrow{\Delta} \circ \overrightarrow{(1+\alpha)}\underbrace{\left(\mathbf{Z}_1\mathbf{Z}'_{11} + \mathbf{W}_1\mathbf{Z}'_{12}\right)}_{I_{11}} \\
&\quad + \overrightarrow{\Delta} \circ \overrightarrow{(2+\alpha)}\underbrace{\left(\mathbf{Z}_2\mathbf{Z}'_{21} + \mathbf{W}_2\mathbf{Z}'_{22}\right)}_{I_{21}} + \overrightarrow{\Delta} \circ \overrightarrow{(1+\alpha)^2}\underbrace{\mathbf{Z}_1\mathbf{Z}'_{12}}_{I_{12}} \\
&= \overrightarrow{\Delta} \circ \overrightarrow{(1+\alpha)^{-1}}W_{\theta 1} + \overrightarrow{\Delta} \circ \overrightarrow{(2+\alpha)^{-1}}W_{\theta 2} \\
&\quad + \overrightarrow{\Delta}(I_{10} + I_{20} + I_{11} + 2I_{21} + I_{12} + 4I_{22}) \\
&\quad + \overrightarrow{\Delta} \circ \overrightarrow{\alpha}(I_{11} + I_{21} + 2I_{12} + 4I_{22}) + \overrightarrow{\Delta} \circ \overrightarrow{\alpha^2}(I_{12} + I_{22}). \qquad (130)
\end{aligned}$$

Thus, interference aligns into the space spanned by the 3 vectors: $\overrightarrow{\Delta}, \overrightarrow{\Delta} \circ \overrightarrow{\alpha}, \overrightarrow{\Delta} \circ \overrightarrow{\alpha^2}$, while the 2 desired symbols appear along $\overrightarrow{\Delta} \circ \overrightarrow{(1+\alpha)^{-1}}$, $\overrightarrow{\Delta} \circ \overrightarrow{(2+\alpha)^{-1}}$. Note that the highest exponent of α along a desired signal dimension is 1, but every interference dimension contains α^2 (contributed by Δ), so the desired signals are independent of the interference. The independence of desired signals among themselves is also easily verified because if $c_1\frac{\Delta}{1+\alpha} + c_2\frac{\Delta}{2+\alpha} = c_1(2+\alpha) + c_2(1+\alpha)$ is the zero polynomial, then by substituting $i + \alpha = 0$ for $i = 1, 2$ we find that we must have $c_1 = c_2 = 0$. Privacy and security are guaranteed by the MDS coded independent noise terms mixed with the message and query symbols. The rate achieved is $2/5$, which matches the asymptotic capacity for this setting.

D. Example: $(X = 2)$ Secure, $(T = 2)$ Private Scheme With $N = 7$ Servers

Each message consists of $L = 3$ symbols from a finite field $\mathbb{F}_q$, of size $q \geq 10$ and characteristic greater than 2.

$$\Delta = (1+\alpha)(2+\alpha)(3+\alpha).$$

	Server 'n' (Replace α, Δ with α_n, Δ_n)
Storage	$\mathbf{W}_1 + (1+\alpha)\mathbf{Z}_{11} + (1+\alpha)^2\mathbf{Z}_{12},$
(S_n)	$\mathbf{W}_2 + (2+\alpha)\mathbf{Z}_{21} + (2+\alpha)^2\mathbf{Z}_{22},$
	$\mathbf{W}_3 + (3+\alpha)\mathbf{Z}_{31} + (3+\alpha)^2\mathbf{Z}_{32}$
Query	$\frac{\Delta}{1+\alpha}\Big(\mathbf{Q}_\theta + (1+\alpha)\mathbf{Z}'_{11} + (1+\alpha)^2\mathbf{Z}'_{12}\Big),$
$(Q_n^{[\theta]})$	$\frac{\Delta}{2+\alpha}\Big(\mathbf{Q}_\theta + (2+\alpha)\mathbf{Z}'_{21} + (2+\alpha)^2\mathbf{Z}'_{22}\Big),$
	$\frac{\Delta}{3+\alpha}\Big(\mathbf{Q}_\theta + (3+\alpha)\mathbf{Z}'_{31} + (3+\alpha)^2\mathbf{Z}'_{32}\Big)$
	Desired symbols appear along vectors
	$\overrightarrow{\Delta} \circ \Big(\overrightarrow{(1+\alpha)^{-1}}, \overrightarrow{(2+\alpha)^{-1}}, \overrightarrow{(3+\alpha)^{-1}}\Big)$
	Interference symbols appear along vectors
	$\overrightarrow{\Delta} \circ \Big(\overrightarrow{1}, \overrightarrow{1+\alpha}, \overrightarrow{(1+\alpha)^2}, \overrightarrow{(1+\alpha)^3},$
	$\overrightarrow{2+\alpha}, \overrightarrow{(2+\alpha)^2}, \overrightarrow{(2+\alpha)^3},$
	$\overrightarrow{3+\alpha}, \overrightarrow{(3+\alpha)^2}, \overrightarrow{(3+\alpha)^3}\Big)$

Interference aligns into the space spanned by the 4 vectors: $\overrightarrow{\Delta}, \overrightarrow{\Delta} \circ \overrightarrow{\alpha}, \overrightarrow{\Delta} \circ \overrightarrow{\alpha^2}, \overrightarrow{\Delta} \circ \overrightarrow{\alpha^3}$. Independence of desired signals from interference is trivially verified – highest exponent of α along any desired signal dimension is 2, but each interference dimension has an α^3 term (contributed by Δ). The desired signal dimensions are easily verified to be linearly independent among themselves because in order for

$$c_1(2+\alpha)(3+\alpha) + c_2(1+\alpha)(3+\alpha) + c_3(1+\alpha)(2+\alpha) \tag{131}$$

to be the zero polynomial it must be zero everywhere, but in that case, setting $\alpha + i = 0$ for $i = 1, 2, 3$ leads us to $c_1 = c_2 = c_3 = 0$. Privacy and security are guaranteed by the MDS coded independent noise terms mixed with the message and query symbols. The rate achieved is $3/7$, which matches the asymptotic capacity for this setting.

VII. Proof of Theorem 4

In Section VI we presented an XSTPIR scheme for arbitrary X, T, N, K that achieves capacity as $K \to \infty$. Since the scheme also works for any K, a natural starting point for finite K settings is to apply the same scheme. A key insight here is that the rate achieved by the scheme improves as K decreases. Let us elaborate. Note that the query $Q_n^{[\theta]}$ that is sent to each server is uniformly distributed in $\mathbb{F}_q^{LK}$. Therefore, with probability $\frac{1}{q^{LK}}$, the query vector is the all zero vector. Whenever this happens, no download is needed from the server. Thus, the average download is reduced by the factor $(1 - 1/q^{LK})$ and the rate achieved is expressed as follows.

Lemma 6: The asymptotically capacity achieving XSTPIR scheme of Section VI achieves the rate

$$R = \left(1 - \frac{1}{q^{KL}}\right)^{-1}\left(1 - \left(\frac{X+T}{N}\right)\right) \tag{132}$$

for arbitrary X, L, K, N values, where $N > X + T$.

Note that $N \leq X + T$ is excluded as the degenerate setting where we already know the capacity for all parameters, according to Theorem 2. Remarkably, the rate in Lemma 6 depends on the message size L and the field size q used by the scheme. As presented, the scheme uses $q \geq L + N$ and $L = N - X - T$. So the achieved rate for finite K becomes

$$R = \left(1 - \frac{1}{(2N - X - T)^{K(N-X-T)}}\right)^{-1}\left(1 - \left(\frac{X+T}{N}\right)\right). \tag{133}$$

Consider the simplest non-trivial setting of interest, i.e., the setting for Theorem 4, where $T = X = 1$, $N = 3$ and K is arbitrary. The scheme of Section VI uses $L = 1, q \geq 4$, so the rate achieved for arbitrary K is

$$R = \frac{1}{3}\left(1 - \frac{1}{4^K}\right)^{-1}. \tag{134}$$

However, note that if the field size could be reduced to $q = 2$, then the rate achieved by the scheme would become

$$\begin{aligned}\frac{1}{3}\left(1 - \frac{1}{2^K}\right)^{-1} &= \frac{2}{3}\left(\frac{1 - \frac{1}{2}}{1 - \frac{1}{2^K}}\right)\\ &= \frac{2}{3}\left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{K-1}}\right)^{-1}\end{aligned}$$

which matches the capacity upper bound from Theorem 1. Surprisingly, this can be done with some modification to the structure of the scheme, as explained below.

Suppose each message $W_k, k \in [1:K]$ consists of $L = 1$ symbol (bit) from $\mathbb{F}_2$. Let $\mathbf{W} = (W_1, W_2, \cdots, W_K)$ be a random row vector in $\mathbb{F}_2^{1\times K}$, containing all messages. Let $\mathbf{Z}$ and $\mathbf{Z}'$ be uniformly random noise vectors from $\mathbb{F}_2^{1\times K}$ and $\mathbb{F}_2^{K\times 1}$, that are used to guarantee data security and user privacy, respectively. The noise vectors are independent of each other and of the message vector and θ, i.e., $H(\mathbf{W}, \mathbf{Z}, \mathbf{Z}', \theta) = H(\mathbf{W}) + H(\mathbf{Z}) + H(\mathbf{Z}') + H(\theta)$. Let $\mathbf{Q}_\theta$ represent the θ^{th} column of $\mathbf{I}_K$ (the $K \times K$ identity matrix). Note that $\mathbf{WQ}_\theta = W_\theta$ is the message desired by the user. The storage at the servers, the queries and the answers are listed below.

	Server 1	Server 2	Server 3
Storage S_n	$\mathbf{W} + \mathbf{Z}$	$\mathbf{W} + \mathbf{ZB}$	$\mathbf{Z}$
Query $Q_n^{[\theta]}$	$\mathbf{Z}'$	$\mathbf{Q}_\theta + \mathbf{Z}'$	$(\mathbf{I}_K + \mathbf{B})\mathbf{Z}'$ $+\mathbf{BQ}_\theta$
Answer $A_n^{[\theta]}$	$\mathbf{WZ}' + \mathbf{ZZ}'$	$\mathbf{WQ}_\theta + \mathbf{WZ}'$ $+\mathbf{ZBZ}' + \mathbf{ZBQ}_\theta$	$\mathbf{ZZ}' + \mathbf{ZBZ}'$ $+\mathbf{ZBQ}_\theta$

where $\mathbf{B}$ is a $K \times K$ deterministic binary matrix such that $\mathbf{B}$ and $\mathbf{I}_K + \mathbf{B}$ are both full rank. Any such choice of $\mathbf{B}$ will work for our scheme. The existence of such $\mathbf{B}$ is established in the following lemma whose proof appears in Appendix A.

Lemma 7: For all $K \geq 2$, there exists a matrix $\mathbf{B} \in \mathbb{F}_2^{K\times K}$ such that $\mathbf{B}$ and $\mathbf{I}_K + \mathbf{B}$ are both invertible.

Now, let us check the correctness, security and privacy of this scheme. The scheme is obviously correct because by adding the three answers shown in the table above, the user recovers W_θ. It is obviously secure because $\mathbf{B}$ is invertible, so $\mathbf{ZB} \sim \mathbf{Z}$, is still uniform noise independent of $\mathbf{W}$. And similarly, it is also obviously private, because $\mathbf{I}_K + \mathbf{B}$ is also invertible, so $(\mathbf{I}_K + \mathbf{B})\mathbf{Z}' \sim \mathbf{Z}'$ is still uniform noise independent of $\mathbf{BQ}_\theta$. Thus, surprisingly, we have achieved the capacity of XSTPIR for arbitrary K, when $X = T = 1$ and $N = 3$, completing the proof of Theorem 4. □

VIII. Conclusion

The XSTPIR problem is timely due to the growing importance of privacy and security concerns in modern information storage and retrieval systems. It is a conceptually rich topic that reveals new insights into alignment of noise terms, dependence of coding and query structures, cost of symmetric security, significance of field size for the rate of information retrieval, etc. As indicated by various open problems identified here, XSTPIR is a fertile research avenue for future work. In particular, the capacity characterization for arbitrary K could reveal fundamentally new schemes for PIR. Especially intriguing would be the role that field size might play in such a result. Capacity of Sym-XSTPIR is another promising open problem. XSTPIR with constraints on the amount of storage per server, coded storage, multi-message retrieval are other open problems that merit investigation.

Appendix A
Proof of Lemma 7

Let $\mathbf{J}_k$ denote the $k \times k$ anti-diagonal identity matrix, and let $\mathbf{0}_{k_1\times k_2}$ denote the $k_1 \times k_2$ matrix where all elements are equal to 0 (when $k_1 = k_2$, this notation is further simplified to $\mathbf{0}_{k_1}$). Define

$$\mathbf{I}'_k = \begin{bmatrix} \mathbf{I}_k & \mathbf{0}_{k\times 1} \\ \mathbf{0}_{1\times k} & 0 \end{bmatrix}. \tag{135}$$

Choose $\mathbf{B}$ as follows.

$$\mathbf{B} = \begin{cases} \begin{bmatrix} \mathbf{I}_{\frac{K}{2}} & \mathbf{J}_{\frac{K}{2}} \\ \mathbf{J}_{\frac{K}{2}} & \mathbf{0}_{\frac{K}{2}} \end{bmatrix}, & \text{if } K \text{ is even,} \\ \left[\begin{array}{c|c} \mathbf{J}_{\frac{K+1}{2}} + \mathbf{I}'_{\frac{K-1}{2}} + \mathbf{I}_{\frac{K+1}{2}} & \begin{matrix} \mathbf{J}_{\frac{K-1}{2}} \\ \mathbf{0}_{1\times\frac{K-1}{2}} \end{matrix} \\ \hdashline \mathbf{J}_{\frac{K-1}{2}} \quad \mathbf{0}_{\frac{K-1}{2}\times 1} & \mathbf{0}_{\frac{K-1}{2}} \end{array}\right], & \text{if } K \text{ is odd.} \end{cases} \tag{136}$$

For example,

when $K = 4$: $$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

and when $K = 5$:
$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{137}$$

Let us show that $\mathbf{B}$ and $\mathbf{I}_K + \mathbf{B}$ are both invertible.

First, consider $\mathbf{B}$. Regardless of whether K is even or odd, $\mathbf{B}$ is an upper anti-triangular matrix where all anti-diagonal elements are 1 so that $\det(\mathbf{B}) = 1$ and $\mathbf{B}$ has full rank.

Next, consider $\mathbf{I}_K + \mathbf{B}$.

When K is even:
$$\begin{aligned} \mathbf{I}_K + \mathbf{B} &= \begin{bmatrix} \mathbf{I}_{\frac{K}{2}} & \mathbf{0}_{\frac{K}{2}} \\ \mathbf{0}_{\frac{K}{2}} & \mathbf{I}_{\frac{K}{2}} \end{bmatrix} + \begin{bmatrix} \mathbf{I}_{\frac{K}{2}} & \mathbf{J}_{\frac{K}{2}} \\ \mathbf{J}_{\frac{K}{2}} & \mathbf{0}_{\frac{K}{2}} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{0}_{\frac{K}{2}} & \mathbf{J}_{\frac{K}{2}} \\ \mathbf{J}_{\frac{K}{2}} & \mathbf{I}_{\frac{K}{2}} \end{bmatrix} && (138) \\ \Rightarrow \quad & \det(\mathbf{I}_K + \mathbf{B}) = 1. && (139) \end{aligned}$$

When K is odd:
$$\begin{aligned} \mathbf{I}_K + \mathbf{B} &= \left[\begin{array}{c|c} \mathbf{I}_{\frac{K+1}{2}} & \mathbf{0}_{\frac{K+1}{2} \times \frac{K-1}{2}} \\ \hdashline \mathbf{0}_{\frac{K-1}{2} \times \frac{K+1}{2}} & \mathbf{I}_{\frac{K-1}{2}} \end{array}\right] \\ &\quad + \left[\begin{array}{c|c} \mathbf{J}_{\frac{K+1}{2}} + \mathbf{I}'_{\frac{K-1}{2}} + \mathbf{I}_{\frac{K+1}{2}} & \begin{matrix} \mathbf{J}_{\frac{K-1}{2}} \\ \mathbf{0}_{1 \times \frac{K-1}{2}} \end{matrix} \\ \hdashline \mathbf{J}_{\frac{K-1}{2}} \quad \mathbf{0}_{\frac{K-1}{2} \times 1} & \mathbf{0}_{\frac{K-1}{2}} \end{array}\right] && (140) \\ &= \left[\begin{array}{c|c} \mathbf{J}_{\frac{K+1}{2}} + \mathbf{I}'_{\frac{K-1}{2}} & \begin{matrix} \mathbf{J}_{\frac{K-1}{2}} \\ \mathbf{0}_{1 \times \frac{K-1}{2}} \end{matrix} \\ \hdashline \mathbf{J}_{\frac{K-1}{2}} \quad \mathbf{0}_{\frac{K-1}{2} \times 1} & \mathbf{I}_{\frac{K-1}{2}} \end{array}\right] && (141) \end{aligned}$$

$$\begin{aligned} \Rightarrow \quad & \det(\mathbf{I}_K + \mathbf{B}) \\ &= \det\Bigg(\mathbf{J}_{\frac{K+1}{2}} + \mathbf{I}'_{\frac{K-1}{2}} \\ &\qquad + \begin{bmatrix} \mathbf{J}_{\frac{K-1}{2}} \\ \mathbf{0}_{1 \times \frac{K-1}{2}} \end{bmatrix} \mathbf{I}^{-1}_{\frac{K-1}{2}} \begin{bmatrix} \mathbf{J}_{\frac{K-1}{2}} & \mathbf{0}_{\frac{K-1}{2} \times 1} \end{bmatrix} \Bigg) && (142) \\ &= \det\left(\mathbf{J}_{\frac{K+1}{2}} + \mathbf{I}'_{\frac{K-1}{2}} + \mathbf{I}'_{\frac{K-1}{2}}\right) = \det\left(\mathbf{J}_{\frac{K+1}{2}}\right) = 1 && (143) \end{aligned}$$

where (142) follows from the following formula on the determinant of a block matrix that is made up of matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ with proper dimensions and $\mathbf{D}$ is invertible.
$$\det\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \det(\mathbf{D}) \det\left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right). \tag{144}$$

The proof is thus complete. □

Appendix B
Proof of Corollaries 1, 2, 3, 4

The proof of Corollary 1 is trivial because imposing the symmetric security constraint cannot increase capacity.

A. Proof of Corollary 2

To prove Corollary 2 we provide a scheme as follows. Each message $W_k, k \in [1 : K]$, consists of $L = 1$ symbol from some finite field $\mathbb{F}_q$. Let $Z_{x,k,m}, x \in [1 : X], k \in [1 : K], m \in [1 : K]$ be independent uniform noise symbols from $\mathbb{F}_q$. The subscript, m, in $Z_{x,k,m}$ is interpreted modulo K, i.e., $Z_{x,k,m} = Z_{x,k,m+K}$. The storage at each server is specified as,
$$S_n = \{Z_{n,k,m}, k \in [1 : K], m \in [1 : K]\}, n \in [1 : X], \tag{145}$$
$$S_n = \left\{W_k + \sum_{x=1}^{X} Z_{x,k,m}, k \in [1 : K], m \in [1 : K]\right\}, n = N. \tag{146}$$

The queries from each server are specified as,
$$Q_n^{[\theta]} : \text{Ask for } \{Z_{n,k,m_o}, k \in [1 : K]\}, n \in [1 : X], \tag{147}$$
$$Q_n^{[\theta]} : \text{Ask for } \{W_k + \sum_{x=1}^{X} Z_{x,k,m_o-\theta+k}, k \in [1 : K]\}, n = N, \tag{148}$$

where m_o is chosen privately and uniformly randomly by the user from $[1 : K]$. Thus, in order to retrieve 1 desired message symbol, the user downloads a total of KN symbols from all servers. The scheme is X-secure because each message symbol is protected by independent uniform noise terms. It is correct because for $k = \theta$ the download from Server N, contains the symbol $W_\theta + \sum_{x=1}^{X} Z_{x,\theta,m_o}$ and the downloads from the first X servers include all the noise terms Z_{x,θ,m_o}. The scheme is private because m_o is chosen uniformly and privately by the user. It satisfies symmetric security because all the undesired message symbols $W_k, k \neq \theta$, contained in the answers are protected by noise terms $Z_{x,k,m_o-\theta+k}$ and these noise terms are independent of the noise terms downloaded from servers $n \in [1 : X]$ because $m_o - \theta + k \neq m_o$ when $k \neq \theta$. The rate achieved is $\frac{1}{KN}$, which is the capacity for this setting. □

Note that in the Sym-XSPIR scheme described above, each server stores K^2 symbols, when the total data is only $KL = K$ symbols. Thus, this Sym-XSPIR scheme takes advantage of unconstrained storage when K is large, more so than the XSTPIR schemes which store no more than KL symbols at each server.

B. Proof of Corollary 3

To prove Corollary 3, we show that the scheme presented in Section VI automatically guarantees symmetric security when $T = 1$. Define
$$\mathbf{W}_i^c = \{\mathbf{W}_l, l \in [1 : L], l \neq i\} \tag{149}$$
$$\mathbf{Z}_{ij}^c = \{\mathbf{Z}_{lx}, l \in [1 : L], x \in [1 : X], (l, x) \neq (i, j)\}. \tag{150}$$

We need to prove that beyond the information that the user must have, i.e., $W_\theta, Q_{[1:N]}^{[\theta]}, \theta$, he cannot learn anything about the messages $\mathbf{W}_{[1:L]}$ from the answers $A_{[1:N]}^{[\theta]}$.
$$\begin{aligned} & I\left(\mathbf{W}_{[1:L]}; A_{[1:N]}^{[\theta]} \mid W_\theta, Q_{[1:N]}^{[\theta]}, \theta\right) \\ &= \sum_{l \in [1:L]} I\left(\mathbf{W}_l; A_{[1:N]}^{[\theta]} \mid \mathbf{W}_{[1:l-1]}, W_\theta, Q_{[1:N]}^{[\theta]}, \theta\right) && (151) \\ &\leq \sum_{l \in [1:L]} I\left(\mathbf{W}_l; A_{[1:N]}^{[\theta]} \mid \mathbf{W}_l^c, W_\theta, Q_{[1:N]}^{[\theta]}, \theta\right) && (152) \\ &\leq \sum_{l \in [1:L]} I\left(\mathbf{W}_l; A_{[1:N]}^{[\theta]} \mid \mathbf{Z}_{l1}^c, \mathbf{W}_l^c, W_\theta, Q_{[1:N]}^{[\theta]}, \theta\right) && (153) \end{aligned}$$

where we repeatedly used the fact that $I(A; B \mid C) \leq I(A; B \mid C, D)$ if $I(A; D \mid C) = 0$ and the facts that

$$I(\mathbf{W}_l; \mathbf{W}_{[l+1:L]} \mid W_{[1:l-1]}, W_\theta, Q^{[\theta]}_{[1:N]}, \theta) = 0 \quad (154)$$

$$I\left(\mathbf{W}_l; \mathbf{Z}^c_{l1} \mid \mathbf{W}^c_l, W_\theta, Q^{[\theta]}_{[1:N]}, \theta\right) = 0 \quad (155)$$

that follow from the independence of messages, queries, and the noise terms, by construction of the scheme in Section VI. To prove Corollary 3 it suffices to show that each of the terms in the summation is zero. Without loss of generality, let us consider $l = 1$. Because of the conditioning on $\mathbf{Z}^c_{11}, \mathbf{W}^c_1, W_\theta, Q^{[\theta]}_{[1:N]}, \theta$, we can subtract the contributions from these terms, whose values are fixed, from $A^{[\theta]}_n$, leaving us with only

$$\begin{aligned}
&A'^{[\theta]}_n \\
&= (\mathbf{W}_1 + (1+\alpha_n)\mathbf{Z}_{11})\left(\frac{\Delta_n}{1+\alpha_n}\right)(\mathbf{Q}_\theta + (1+\alpha_n)\mathbf{Z}'_1) \quad (156)\\
&= \left(\frac{\Delta_n}{1+\alpha_n}\right)\mathbf{W}_1\mathbf{Q}_\theta \\
&\quad + \Delta_n(\mathbf{W}_1\mathbf{Z}'_1 + \mathbf{Z}_{11}\mathbf{Q}_\theta) + \Delta_n(1+\alpha_n)\mathbf{Z}_{11}\mathbf{Z}'_1 \quad (157)\\
&= \left(\frac{\Delta_n}{1+\alpha_n}\right)W_{\theta 1} \\
&\quad + \Delta_n(\mathbf{W}_1\mathbf{Z}'_1 + \mathbf{Z}_{11}(\theta)) + \Delta_n(1+\alpha_n)\mathbf{Z}_{11}\mathbf{Z}'_1 \quad (158)
\end{aligned}$$

where $\mathbf{Z}_{11}(i)$ is the i^{th} element of the vector $\mathbf{Z}_{11}$. Note that $W_{\theta 1}$ is also a constant because of the conditioning on W_θ. Given $\mathbf{Z}^c_{11}, \mathbf{W}^c_1, W_\theta, Q^{[\theta]}_{[1:N]}, \theta$, the random variable $A^{[\theta]}_{[1:N]}$ is an invertible function of $A'^{[\theta]}_{[1:N]}$.

$$\begin{aligned}
&I\left(\mathbf{W}_1; A^{[\theta]}_{[1:N]} \mid \mathbf{Z}^c_{11}, \mathbf{W}^c_1, W_\theta, Q^{[\theta]}_{[1:N]}, \theta\right) \quad (159)\\
&= I\left(\mathbf{W}_1; A'^{[\theta]}_{[1:N]} \mid \mathbf{Z}^c_{11}, \mathbf{W}^c_1, W_\theta, Q^{[\theta]}_{[1:N]}, \theta\right) \quad (160)\\
&= I\left(\mathbf{W}_1; A'^{[\theta]}_{[1:N]} \mid W_{\theta 1}, Q^{[\theta]}_{[1:N]}, \theta\right) \quad (161)\\
&\leq I(\mathbf{W}_1; \mathbf{W}_1\mathbf{Z}'_1 + \mathbf{Z}_{11}(\theta), \mathbf{Z}_{11}\mathbf{Z}'_1 \mid W_{\theta 1}, \mathbf{Q}_\theta, \theta) \quad (162)\\
&\leq I(\mathbf{W}_1; \mathbf{W}_1\mathbf{Z}'_1 + \mathbf{Z}_{11}(\theta), \mathbf{Z}_{11}\mathbf{Z}'_1 \mid W_{\theta 1}, \mathbf{Q}_\theta, \mathbf{Z}'_1, \theta) \quad (163)\\
&= 0. \quad (164)
\end{aligned}$$

In (162) we used the fact that given $W_{\theta 1}$, the random variable $A'^{[\theta]}_{[1:N]}$ is a function of $\mathbf{W}_1\mathbf{Z}'_1 + \mathbf{Z}_{11}(\theta), \mathbf{Z}_{11}\mathbf{Z}'_1$ because of (158), and the fact that for any random variables A, B, C, we must have $I(A; f(B) \mid C) \leq I(A; B \mid C)$. In (163) we used the fact that conditioning on an independent random variable cannot reduce mutual information, i.e., $I(A; B \mid C) \leq I(A; B \mid C, D)$ if $I(A; D \mid C) = 0$, and the fact that $\mathbf{Z}'_1$ is independent of $\mathbf{W}_1$ after conditioning on $W_{\theta 1}, \mathbf{Q}_\theta, \theta$ by construction of the scheme as described in Section VI. The last step is justified as follows. Because of the conditioning on $\mathbf{Z}'_1$, its value is a constant for which there are only three possibilities: $\mathbf{Z}'_1$ is either the zero vector, or it is equal to $\mu\mathbf{Q}_\theta$ for some non-zero $\mu \in \mathbb{F}_q$, or it is neither zero nor equal to $\mu\mathbf{Q}_\theta$. If $\mathbf{Z}'_1$ is the zero vector, then the mutual information is automatically zero because $\mathbf{W}_1$ is eliminated entirely. If $\mathbf{Z}'_1 = \mu\mathbf{Q}_\theta$ for some non-zero μ, then $\mathbf{W}_1\mathbf{Z}'_1 = \mu W_{\theta 1}$ and the mutual information is again zero because of the conditioning on $W_{\theta 1}$. Finally, if $\mathbf{Z}'_1$ is neither zero nor a scaled version of $\mathbf{Q}_\theta$, then $\mathbf{Z}_{11}\mathbf{Z}'_1$ is a sum of uniformly random noise terms in $\mathbb{F}_q$, at least one of which is independent of $\mathbf{Z}_{11}(\theta)$ and $\mathbf{Z}'_1$. So in this case also the mutual information is zero. This completes the proof of Corollary 3. □

C. Proof of Corollary 4

The proof of Corollary 4 is presented next. Recall that in the scheme of the proof of Theorem 4, the user obtains the following three symbols from the answers,

$$\mathbf{W}\mathbf{Q}_\theta = W_\theta \quad (165)$$

$$\mathbf{W}\mathbf{Z}' + \mathbf{Z}\mathbf{Z}' \quad (166)$$

$$\mathbf{W}\mathbf{Z}' + \mathbf{Z}\mathbf{B}\mathbf{Z}' + \mathbf{Z}\mathbf{B}\mathbf{Q}_\theta. \quad (167)$$

We show that symmetric security holds, i.e., conditioned on $\mathbf{Z}'$, from these three symbols the user learns nothing about the undesired messages $W_1, \cdots, W_{\theta-1}, W_{\theta+1}, \cdots, W_K$. When $\mathbf{Z}'$ is the zero vector, the symbol $\mathbf{W}\mathbf{Z}'$ is zero as well, leaking nothing about the undesired messages. Now consider (166). If $\mathbf{Z}'$ is not the zero vector, then the symbol $\mathbf{W}\mathbf{Z}'$ is protected by an independent noise term. Similarly, consider (167) and consider three possibilities: $\mathbf{B}(\mathbf{Z}' + \mathbf{Q}_\theta)$ is either zero, or equal to $\mathbf{Z}'$, or not zero and not equal to $\mathbf{Z}'$. If $\mathbf{B}(\mathbf{Z}' + \mathbf{Q}_\theta)$ is the zero vector, then because $\mathbf{B}$ is invertible, we must have $\mathbf{Z}' = \mathbf{Q}_\theta$, so the symbol $\mathbf{W}\mathbf{Z}' = \mathbf{W}\mathbf{Q}_\theta$ is the desired message, again leaking nothing about undesired messages. If $\mathbf{B}(\mathbf{Z}' + \mathbf{Q}_\theta) = \mathbf{Z}'$ then (167) is redundant, i.e., same as (166), so it leaks no new information. Finally, if $\mathbf{B}(\mathbf{Z}' + \mathbf{Q}_\theta)$ is not zero and not equal to $\mathbf{Z}'$, then $\mathbf{Z}\mathbf{B}(\mathbf{Z}' + \mathbf{Q}_\theta)$ is independent of $\mathbf{Z}\mathbf{Z}'$, so that (167) is protected by an independent noise term. Therefore, in all cases, the user learns nothing about undesired messages, and this completes the proof of symmetric security. □

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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, Nov. 2017.

[7] 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.

[8] 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

[9] M. A. Attia, D. Kumar, and R. Tandon. (2018). "The capacity of private information retrieval from uncoded storage constrained databases." [Online]. Available: https://arxiv.org/abs/1805.04104

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

[11] 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.

[12] H. Sun and S. A. Jafar, "Multiround private information retrieval: Capacity and storage overhead," *IEEE Trans. Inf. Theory*, vol. 64, no. 8, pp. 5743–5754, Aug. 2018.

[13] C. Tian, H. Sun, and J. Chen, "A Shannon-theoretic approach to the storage-retrieval tradeoff in PIR systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1904–1908.

[14] 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

[15] 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

[16] R. Tandon. (2017). "The capacity of cache aided private information retrieval." [Online]. Available: https://arxiv.org/abs/1706.07035

[17] Y.-P. Wei, K. Banawan, and S. Ulukus, "Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 3215–3232, May 2019.

[18] 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

[19] S. P. Shariatpanahi, M. J. Siavoshani, and M. A. Maddah-Ali. (2018). "Multi-message private information retrieval with private side information." [Online]. Available: https://arxiv.org/abs/1805.11892

[20] S. Li and M. Gastpar. (2018). "Single-server multi-message private information retrieval with side information." [Online]. Available: https://arxiv.org/abs/1808.05797

[21] K. Banawan and S. Ulukus. (2018). "Private information retrieval through wiretap channel ii: Privacy meets security." [Online]. Available: https://arxiv.org/abs/1801.06171

[22] Q. Wang, H. Sun, and M. Skoglund, "The capacity of private information retrieval with eavesdroppers," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 3198–3214, May 2019.

[23] K. Banawan and S. Ulukus, "The capacity of private information retrieval from byzantine and colluding databases," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 1206–1219, Feb. 2019.

[24] Y. Zhang and G. Ge. (2017). "Private information retrieval from MDS coded databases with colluding servers under several variant models." [Online]. Available: https://arxiv.org/abs/1705.03186

[25] R. Tajeddine, O. W. Gnilke, D. Karpuk, R. Freij-Hollanti, an C. Hollanti. (2018). "Private information retrieval from coded storage systems with colluding, byzantine, and unresponsive servers." [Online]. Available: https://arxiv.org/abs/1806.08006

[26] Q. Wang, H. Sun, and M. Skoglund, "The ϵ-error capacity of symmetric PIR with byzantine adversaries," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2018, pp. 1–5.

[27] X. Yao, N. Liu, and W. Kang. (2019). "The capacity of multi-round private information retrieval from byzantine databases." [Online]. Available: https://arxiv.org/abs/1901.06907

[28] H. Sun and S. A. Jafar, "The capacity of symmetric private information retrieval," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 322–329, Jan. 2019.

[29] Q. Wang and M. Skoglund. (2017). "Linear symmetric private information retrieval for MDS coded distributed storage with colluding servers." [Online]. Available: https://arxiv.org/abs/1708.05673

[30] Q. Wang and M. Skoglund. (2017). "Secure symmetric private information retrieval from colluding databases with adversaries." [Online]. Available: https://arxiv.org/abs/1707.02152

[31] H. Yang, W. Shin, and J. Lee, "Private information retrieval for secure distributed storage systems," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 12, pp. 2953–2964, Dec. 2018.

[32] H. Sun and S. A. Jafar, "Blind interference alignment for private information retrieval," in *Proc. IEEE Int. Symp. Inf.Theory (ISIT)*, Jul. 2016, pp. 560–564.

[33] H. Sun and S. A. Jafar, "The capacity of private computation," *IEEE Trans. Inf. Theory*, to be published. doi: 10.1109/TIT.2018.2888494.

[34] M. Mirmohseni and M. A. Maddah-Ali. (2017). "Private function retrieval." [Online]. Available: https://arxiv.org/abs/1711.04677

Zhuqing Jia (S'19) received the B.E. degree in Electronic Information Engineering in 2015 from Beijing University of Posts and Telecommunications, Beijing, China. He is currently pursuing the Ph.D. degree at the University of California, Irvine, CA USA. His research interests include information theory and its applications to security, privacy and storage.

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 in 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 Paper Award, IEEE Communication Society and Information Theory Society Joint 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, an IEEE Communications Society Distinguished Lecturer for 2013-2014, and is serving as an IEEE Information Theory Society Distinguished Lecturer for 2019-20. 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, 2017 and 2018. 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.