The Capacity of Private Information Retrieval With Eavesdroppers

Qiwen Wang[iD], *Member, IEEE*, Hua Sun[iD], *Member, IEEE*, and Mikael Skoglund[iD], *Fellow, IEEE*

Abstract— We consider the problem of private information retrieval (PIR) with colluding servers and eavesdroppers (abbreviated as ETPIR). The ETPIR problem is comprised of K messages and N servers where each server stores all K messages, a user who wants to retrieve one of the K messages without revealing the desired message index to any set of T colluding servers, and an eavesdropper who can listen to the queries and answers of any E servers but is prevented from learning any information about the messages. The information theoretic capacity of ETPIR is defined to be the maximum number of desired message symbols retrieved privately per information symbol downloaded. We show that the capacity of ETPIR is $C = (1 - (E/N))(1 + (T - E/N - E) + \cdots + ((T - E/N - E))^{K-1})^{-1}$ when $E < T$, and $C = (1 - (E/N))$ when $E \geq T$. To achieve the capacity, the servers need to share a common random variable (independent of the messages), and its size must be at least $(E/N) \cdot (1/C)$ symbols per message symbol. Otherwise, with less amount of shared common randomness, ETPIR is not feasible and the capacity reduces to zero. An interesting observation is that the ETPIR capacity expression takes different forms in two regimes. When $E < T$, the capacity equals the inverse of a sum of a geometric series with K terms and decreases with K; this form is typical for capacity expressions of PIR. When $E \geq T$, the capacity does not depend on K, a typical form for capacity expressions of SPIR (symmetric PIR, which further requires data-privacy, *i.e.*, the user learns no information about other undesired messages); the capacity does not depend on T either. In addition, the ETPIR capacity result includes multiple previous PIR and SPIR capacity results as special cases.

Index Terms— Private information retrieval, capacity, colluding servers, eavesdroppers.

I. INTRODUCTION

THE problem of private information retrieval (PIR) is to retrieve one message out of K messages privately from N servers (each stores all K messages) without letting any server know the identity of the requested message. An important factor in the design of such PIR protocols is the communication efficiency. Recently, a series of works studies diverse

variations of the PIR problem from an information-theoretic perspective. Among these works, the messages are considered to be long sequences, such that the upload cost is neglected as it is diminishing when compared to the download cost. The communication efficiency is measured by the capacity, which is defined as the maximum number of desired information bits retrieved privately per information bit downloaded.

The first work in this series is [1], which finds the PIR capacity as $C_{\text{PIR}} = \left(1 + \frac{1}{N} + \cdots + \left(\frac{1}{N}\right)^{K-1}\right)^{-1}$. A significant number of following works add various elements to the PIR problem to form different variations. Among these elements, three are most relevant to this work: colluding servers, adversaries, and symmetric PIR (SPIR).

We first discuss related works on colluding servers. In the modeling of the initial work [1], the servers do not share information about their communication with the user, *i.e.*, they do not collude to infer the identity of the requested message. However, it is possible in practical systems that some sets of servers can communicate and then potentially collude. In [2], the problem where any set of T servers may collude is studied (TPIR) and the capacity is established to be $C_{\text{TPIR}} = \left(1 + \frac{T}{N} + \cdots + \left(\frac{T}{N}\right)^{K-1}\right)^{-1}$. Beyond TPIR, PIR with arbitrary collusion patterns (where the cardinality of each set of colluding servers may vary) is first studied in [3] and the capacity of one particular case is found in [4], where the colluding sets are disjoint.

The second element is adversaries, referring to scenarios where the communication system is vulnerable to passive and/or active attacks. This is also a practical factor to consider in PIR protocol design, because applications of cloud storage are usually built on open systems and networks, where preserving data security and privacy is crucial and more challenging. The problem of TPIR with an active adversary who may erase the responses from any A servers is considered in [2] and interestingly, the capacity in this setting equals the capacity of TPIR with $N - A$ servers (the number of surviving servers). The problem of TPIR with a Byzantine (active) adversary who may introduce arbitrary errors to the responses from any B servers is considered in [5] and interestingly, the capacity in this setting equals the capacity of TPIR with $N - 2B$ servers times $(N - 2B)/N$ (where $2B$ out of all N servers are used to identify and correct the errors and are thus wasted, akin to the Hamming distance requirement in algebraic coding theory). In this work, we consider the problem of TPIR with a passive eavesdropper who may hear the responses from any E
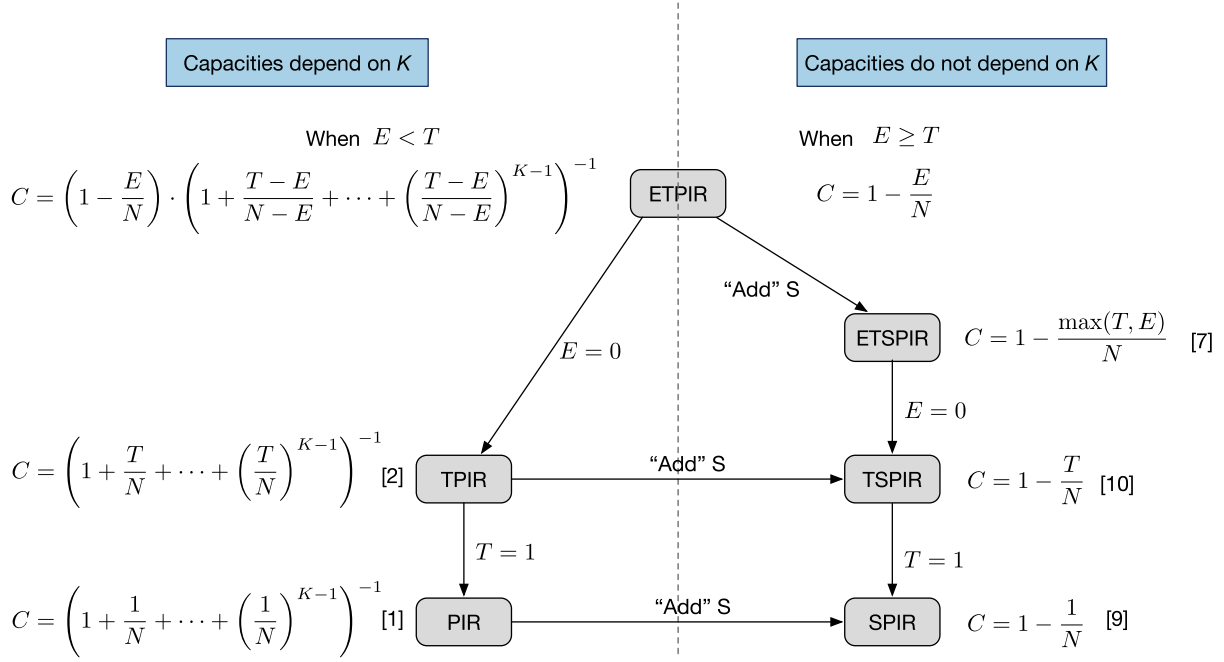
Fig. 1. Connections to prior works and results.

servers while learning nothing about the messages (ETPIR). The ETPIR problem is first studied in [6] and [7], where an achievable scheme and a converse are provided, although the two do not match. We will close the gap and establish the exact capacity in this work. It turns out that neither the achievable scheme nor the converse in [6] and [7] is tight and we need to improve both. Recent work [8] considers a related PIR setting with eavesdroppers, but the eavesdropping model is different. In [8], the eavesdropper can wiretap all servers, but can only observe a fraction of each server's answer. While in our work, the eavesdropper can only wiretap E (instead of all N) servers, but if a server is wiretapped, then its answer is fully (instead of partially) observed by the eavesdropper.

We finally consider SPIR. Relative to PIR, SPIR requires one more constraint on data-privacy at the user side. Namely, the user should not learn any information about the other messages besides the desired one. Although this work focuses exclusively on the PIR problem, our result turns out to relate intimately to SPIR so that a brief summary of known results along this line is presented next. For a PIR problem, if the letter 'S' is added to its description, we are referring to the same problem with SPIR constraint. For example, SPIR refers to the PIR problem with data-privacy requirement, and its capacity is shown to be $C_{\text{SPIR}} = 1 - \frac{1}{N}$ [9]. The capacity of TSPIR is characterized as $C_{\text{TSPIR}} = 1 - \frac{T}{N}$, by specializing a more general result from [10]. The capacity of TSPIR with Byzantine (active) adversaries is characterized as $C_{\text{BTSPIR}} = 1 - \frac{2B+T}{N}$ in [7] and [11]. The capacity of ETSPIR is established to be $C_{\text{ETSPIR}} = 1 - \frac{\max(T,E)}{N}$ in [7] and [11]. An interesting observation from these results is that the capacity of a PIR problem is always the inverse of a sum of a geometric series with K terms, and decreases

with the number of messages K while the capacity of an SPIR problem does not depend K. Furthermore, the SPIR capacity is the limit of the corresponding PIR capacity by taking $K \to \infty$.

The main contribution of this work is that we characterize the capacity of ETPIR (where two elements – eavesdropper and colluding servers are added to the conventional setting of PIR). Specifically, K messages are replicated at N servers. A user wants to retrieve one message by communicating with the servers, without revealing the identity of the desired message. Any T servers may collude, that is, they may share their communication with the user to infer the desired message index. The system is vulnerable to eavesdroppers, who are curious about the messages and can tap in on the communication between any E servers and the user. The ETPIR protocol should prevent the eavesdropper from learning any information about the messages. We show that the capacity of ETPIR is

$$C_{\text{ETPIR}}$$
$$= \begin{cases} \left(1 - \frac{E}{N}\right) \cdot \left(1 + \frac{T-E}{N-E} + \left(\frac{T-E}{N-E}\right)^2 + \ldots + \left(\frac{T-E}{N-E}\right)^{K-1}\right)^{-1}, \\ \qquad \text{when } E < T; \\ 1 - \frac{E}{N}, \quad \text{when } E \geq T. \end{cases}$$

To achieve the capacity, the servers need to share some common randomness that is independent of the messages, in the amount of at least an $\frac{E}{N}$ fraction of one over the capacity times the message size. Otherwise, the capacity is zero and the ETPIR problem is not feasible.

The ETPIR capacity result includes the capacity results of various previously studied settings as special cases, e.g., TPIR, PIR, ETSPIR, TSPIR and SPIR. A pictorial illustration is shown in Figure 1. It is interesting that in the regime of

$E \geq T$, ETPIR capacity is always equal to ETSPIR capacity (instead of approaching it in the large K limit), *i.e.,* there is no penalty in the retrieval rate by further requiring data-privacy at the user side.

A. Notations

Let $[m : n]$ denote the set $\{m, m + 1, \ldots, n\}$ for $m \leq n$, and let $(m : n)$ denote the vector $(m, m + 1, \ldots, n)$. To simplify the notation, denote the set of random variables $\{X_m, X_{m+1}, \ldots, X_n\}$ by $X_{[m:n]}$ and the vector $(X_m, X_{m+1}, \ldots, X_n)$ by $X_{(m:n)}$. For an index set $\mathcal{I} = \{i_1, i_2, \ldots, i_n\}$, denote the set of variables $\{X_i : i \in \mathcal{I}\}$ by $X_{\mathcal{I}}$. The cardinality of a set $\mathcal{A}$ is denoted by $|\mathcal{A}|$.

We denote the dimension of a matrix by the superscript (the dimension is omitted when there is no ambiguity). For example, a matrix $\mathbf{M}$ with m rows and n columns is denoted by $\mathbf{M}^{m \times n}$. When the matrix is square ($m = n$), the notation is simplified to $\mathbf{M}^m$. The identity matrix with dimension $m \times m$ is denoted by $\mathbf{I}^m$. The set of all full-rank $n \times n$ matrices over a finite field $\mathbb{F}_q$ is denoted by $GL_n(\mathbb{F}_q)$. For a matrix $\mathbf{M}$, $\mathbf{M}[\mathcal{I}, :]$ denotes the submatrix of $\mathbf{M}$ formed by the rows corresponding to the elements of the vector $\mathcal{I}$. If $\mathbf{G}$ is comprised of $A \times B$ square matrices with dimension $I \times I$, *i.e.,* $\mathbf{G}$ is a block matrix with dimensions $AI \times BI$,

$$\mathbf{G} = \begin{bmatrix} \mathbf{M}_{1,1} & \cdots & \mathbf{M}_{1,B} \\ \vdots & \ddots & \vdots \\ \mathbf{M}_{A,1} & \cdots & \mathbf{M}_{A,B} \end{bmatrix}, \tag{1}$$

where $\mathbf{M}_{i,j}$ are all $I \times I$ matrices, denote the i-th *block row* by $\mathbf{G}^I[i, :]$ (*i.e.,* the submatrix of $\mathbf{G}$ comprised of I rows from the $[(i-1)I + 1]$-th row to the iI-th row),

$$\mathbf{G}^I[i, :] = \begin{bmatrix} \mathbf{M}_{i,1} & \cdots & \mathbf{M}_{i,B} \end{bmatrix}. \tag{2}$$

Block column is defined in a similar way. The notation $\otimes$ denotes the Kronecker product. We denote the $m \times n$ Vandermonde matrix generated from n distinct symbols $\lambda_1, \lambda_2, \ldots, \lambda_n$ from a finite field by $\mathbf{V}^m(\lambda_1, \ldots, \lambda_n)$, *i.e.,*

$$\mathbf{V}^m(\lambda_1, \ldots, \lambda_n) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{m-1} & \lambda_2^{m-1} & \cdots & \lambda_n^{m-1} \end{bmatrix}. \tag{3}$$

II. PROBLEM SETUP

A database comprised of K messages is replicated at N servers. The messages $\{W_k\}$ are independent and each message consists of L symbols from $\mathbb{F}_q$.

$$H(W_k) = L, \quad \text{for } k = 1, \ldots, K; \quad H(W_1, \ldots, W_K) = KL. \tag{4}$$

Here and throughout the paper we measure entropy to base q.

A user wants to retrieve a message W_κ with index κ from the servers, where the desired message index κ follows some prior distribution over $[1 : K]$. Denote the realization of κ by k. Based on k, the user generates random queries to send to the servers. The query received by Server n is denoted by $Q_n^{[k]}$.

Let $\mathcal{Q} = \left[Q_n^{[k]} \right]_{n \in [1:N], k \in [1:K]}$ denote the complete query scheme, namely, the collection of all queries under all choices of the desired message index. The queries are independent of the messages, *i.e.,*

$$I(\mathcal{Q}; W_{[1:K]}) = 0. \tag{5}$$

In the communication system, an eavesdropper is interested in the messages and can tap in on the communication between any E servers and the user. To prevent the eavesdropper from learning the messages, the servers share a common random variable, denoted by S, which is independent of the messages and the queries, *i.e.,*

$$I(S; W_{[1:K]}, \mathcal{Q}) = 0. \tag{6}$$

Let ρ denote the ratio of the amount of common randomness relative to the message size, *i.e.,*

$$\rho \triangleq \frac{H(S)}{H(W_k)} = \frac{H(S)}{L}. \tag{7}$$

The servers follow the protocol agreed with the user *a priori*, and generate answers based on the received query $Q_n^{[k]}$, the stored messages $W_{[1:K]}$, and the common random variable S. The answer sent to the user from Server n is denoted by $A_n^{[k]}$. We have

$$H(A_n^{[k]} | Q_n^{[k]}, W_{[1:K]}, S) = 0. \tag{8}$$

The eavesdropper must learn no information about the messages from the queries and answers of any E servers. That is, the following E-security constraint must be satisfied,

$$[E\text{-security}] \quad I(W_{[1:K]}; A_{\mathcal{E}}^{[k]}, Q_{\mathcal{E}}^{[k]}) = 0,$$
$$\forall \mathcal{E} \subset [1 : N], \ |\mathcal{E}| = E. \tag{9}$$

Any T servers may collude. To guarantee user-privacy, from the queries and answers of any T servers, together with the message contents and the common random variable, the servers should not be able to infer any information about the desired message index. Thus, the following T-privacy constraint must be satisfied,

$$[T\text{-privacy}] \quad I(A_{\mathcal{T}}^{[\kappa]}, Q_{\mathcal{T}}^{[\kappa]}, W_{[1:K]}, S; \kappa) = 0,$$
$$\forall \mathcal{T} \subset [1 : N], \ |\mathcal{T}| = T. \tag{10}$$

From all the answers downloaded from the servers and the available information at the user, the user should be able to decode the desired message with diminishing probability of error as L tends to infinity. By Fano's inequality, this corresponds to the following correctness constraint,

$$[\text{Correctness}] \quad H(W_k | A_{[1:N]}^{[k]}, Q_{[1:N]}^{[k]}) = o(L). \tag{11}$$

The ETPIR rate, R of a scheme characterizes the number of desired information symbols retrieved per downloaded information symbol,

$$R = \frac{H(W_k)}{\sum_{n=1}^{N} H(A_n^{[k]})} = \frac{L}{\sum_{n=1}^{N} H(A_n^{[k]})}. \tag{12}$$

A rate R is said to be ϵ-error achievable if there exists a sequence of ETPIR schemes with rate at least R, and probability of error $P_e \to 0$ as $L \to \infty$. The supremum of all ϵ-error achievable rates is called the ϵ-error capacity C.

III. MAIN RESULT

Theorem 1 below states the main result of this work.

Theorem 1: The capacity of the ETPIR problem is

$$
C_{\text{ETPIR}}
= \begin{cases}
\left(1 - \frac{E}{N}\right) \cdot \left(1 + \frac{T-E}{N-E} + \left(\frac{T-E}{N-E}\right)^2 + \cdots + \left(\frac{T-E}{N-E}\right)^{K-1}\right)^{-1}, \\
\qquad \text{when } E < T; \\
1 - \frac{E}{N}, \quad \text{when } E \geq T.
\end{cases}
$$

To achieve the capacity, $\rho_{\text{ETPIR}} \geq \frac{E}{N} \cdot C_{\text{ETPIR}}^{-1}$. If $\rho_{\text{ETPIR}} < \frac{E}{N} \cdot C_{\text{ETPIR}}^{-1}$, the ETPIR problem is not feasible.

For our theorem, we have the following interesting observations.

Observation 1: When $E = 0$, $C_{\text{ETPIR}} = C_{\text{TPIR}} = \left(1 + \frac{T}{N} + \cdots + (\frac{T}{N})^{K-1}\right)^{-1}$. Therefore, Theorem 1 includes the capacity of TPIR [2] as a special case.

Observation 2: When $T = 1$, $C_{\text{ETPIR}} = C_{\text{EPIR}} = 1 - \frac{E}{N}$. Note that the EPIR problem (PIR with eavesdroppers) has not been studied in the literature. Therefore, the capacity of EPIR is obtained from Theorem 1 as a rewarding by-product.

Observation 3: When $E = 0$, $T = 1$, $C_{\text{ETPIR}} = C_{\text{PIR}} = \left(1 + \frac{1}{N} + \cdots + (\frac{1}{N})^{K-1}\right)^{-1}$. Therefore, Theorem 1 includes the capacity of PIR [1] as a special case.

Observation 4: When $T \leq N - E$, our achievable scheme attains the capacity with zero-error. When $T > N - E$, our achievable scheme has ϵ-error. It is interesting to see if the scheme can be strengthened to have zero-error.

Observation 5: When $E < T$, the capacity is the product of two terms. The first term $1 - \frac{E}{N} = \frac{N-E}{N}$ has an intuitive illustration – from the security constraint (9), the answers from any E servers provide no useful information to decode the messages (out of the N servers, only $N - E$ are useful. Thus we have the ratio $\frac{N-E}{N}$). The second term $\left(1 + \frac{T-E}{N-E} + \left(\frac{T-E}{N-E}\right)^2 + \cdots + \left(\frac{T-E}{N-E}\right)^{K-1}\right)^{-1}$ is more intriguing. Note that it is equal to the capacity of TPIR with $N - E$ servers where any $T - E$ servers might collude. If we view the first E servers as noise (common randomness) providers, then we are left with $N - E$ servers that return noiseless answers. For the remaining $N - E$ servers, a first thought is that any T (instead of $T - E$) servers might collude. Interestingly, it turns out that the first E noise providing servers also contribute to reduce the T-privacy constraint in the original N servers setting to the $(T - E)$-privacy constraint in the $N - E$ servers setting after noise is cancelled.

Observation 6: When $E \geq T$, the second term disappears (the T-privacy constraint is dominated by the E-security constraint) and the capacity depends neither on K nor on T. As the capacity does not depend on T, there is no penalty in increasing T to E, *i.e.,* we automatically have E-privacy (see the achievable scheme in Section V).

Observation 7: When $E \geq T$, the capacity of ETPIR turns out to be equal to the capacity of ETSPIR [11]. As ETSPIR requires one more data-privacy constraint at the user, it is evident that the converse of ETPIR also holds for ETSPIR. Our achievable scheme for ETPIR directly guarantees data-privacy such that data-privacy is obtained for free.

Observation 8: When $E < T$, $\lim_{K \to \infty} C_{\text{ETPIR}} = C_{\text{ETSPIR}}$ (consistent with previous relations between PIR capacity and SPIR capacity); when $E \geq T$, $C_{\text{ETPIR}} \equiv C_{\text{ETSPIR}}$ (distinct from previous relations).

IV. ACHIEVABILITY WHEN $E < T$

In this section, we present the proof of the achievability part of Theorem 1 for the case of $E < T$. In the setting of TPIR [2], the queries are constructed using an MDS code. However, for our setting of ETPIR (with an additional element of eavesdroppers), MDS coded queries no longer suffice and we need a more sophisticated design. A high level description of the protocol design is as follows. As the eavesdropper can tap in on any E answers (meaning that the combination of any E answers can not contain any useful information about the messages), the shared common randomness is coded with an (N, E)-MDS matrix and added to the answers as the noise to guarantee E-security and the property that the noise can be cancelled by projecting the answers to the dual space of the (N, E)-MDS code. After the noise is cancelled, we wish to ensure that the undesired symbols retain certain linear dependency (for example, MDS property suffices) such that they (named *interference*) can be cancelled and the desired symbols can be decoded. Beyond this correctness constraint, we further need to satisfy T-privacy by ensuring that every T servers observe linearly independent queries. For such a purpose, an (N, T)-MDS structure on the queries is desired. To summarize, the essence of our scheme is that the queries are coded by a (first) MDS code, and after projecting queries to the dual space of the (second) MDS code of the common randomness, the projected queries still form a (third) MDS code. Next we illustrate this design progressively by a series of examples with increasing complexity in Sections IV-A - IV-E, leading to the general scheme in Section IV-F.

A. Elemental Case $K = 2, N = 3, T = 2, E = 1$

Let's start with the smallest setting with $K = 2$ messages, $N = 3$ servers where any $T = 2$ may collude, and an eavesdropper who may hear the communication associated with any $E = 1$ server.

Let $\{a_i\}$ and $\{b_i\}$ denote the *mixtures* (linear combinations) of symbols from W_1 and W_2 respectively, and let r, s, t denote three uniform independent symbols shared by the servers (unknown to the eavesdropper and the user). It is natural to start with a scheme in the following form where the answer from 1 server is pure noise. Without loss of generality, suppose the user wants to retrieve W_1.

Server 1	Server 2	Server 3
$a_1 + r$	$a_3 + r$	r
$b_1 + s$	$b_2 + s$	s
$a_2 + b_2 + t$	$a_4 + b_1 + t$	t

Although the above scheme is correct as noise variables and non-desired symbols can be cancelled, it is not 2-private as Server 1 and Server 2 can collude and figure out that the b symbols have the same indices while the a symbols have distinct indices. To make it 2-private, we add message symbols

with carefully chosen linear coefficients to Server 3 (so that it will enhance privacy and simultaneously act as the noise provider), as shown below.

Server 1	Server 2	Server 3
$a_1 + r$	$a_3 + r$	$a_5 + r$
$b_1 + s$	$b_3 + s$	$b_5 + s$
$a_2 + b_2 + t$	$a_4 + b_4 + t$	$a_6 + b_6 + t$

The specific construction of the mixtures $\{a_i\}$ and $\{b_i\}$ is as follows. Assume each message consists of 4 symbols from a field $\mathbb{F}_q$ with $q \geq 3$. Let W_1 and W_2 denote the column vectors comprised of the $L = 4$ symbols of each message. The user privately chooses two 4×4 matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ independently and uniformly from $GL_4(\mathbb{F}_q)$. Recall that $a_{(1:4)}$ and $b_{(1:4)}$ denote the column vectors comprised of $\{a_1, \ldots, a_4\}$ and $\{b_1, \ldots, b_4\}$, respectively. They are generated by

$$a_{(1:4)} = \mathbf{U}_1 W_1, \tag{13}$$
$$b_{(1:4)} = \mathbf{U}_2 W_2. \tag{14}$$

Further, a_5, a_6, b_5, b_6 are assigned as follows.

Server 1	Server 2	Server 3
$a_1 + r$	$a_3 + r$	$a_1 + a_3 + r$
$b_1 + s$	$b_3 + s$	$b_3 + (b_4 - b_2) + s$
$a_2 + b_2 + t$	$a_4 + b_4 + t$	$(a_2 + a_4) + (b_3 - b_1) + (2b_4 - b_2) + t$

The query sent to Server 1 is $(\mathbf{U}_1[1, :], \mathbf{U}_2[1, :], \mathbf{U}_1[2, :], \mathbf{U}_2[2, ;])$, i.e., the coefficients for the message symbols. Using the coefficients received, the servers then produce the answers according to the table above.

Next, we remove the noise variables by subtracting the answers of Server 1 from the answers of Server 2 and Server 3.

Server 2 − Server 1	Server 3 − Server 1
$(a_3 - a_1)$	a_3
$(b_3 - b_1)$	$(b_3 - b_1) + (b_4 - b_2)$
$(a_4 - a_2) + (b_4 - b_2)$	$a_4 + (b_3 - b_1) + 2(b_4 - b_2)$

We are now ready to prove that the scheme satisfies the required properties.

1) Correctness: We call one row of the retrieval table a *query row*. That is, a query row is comprised of one symbol (in the same row) downloaded from each server. After removing the noise symbols r, s, t, from the first query row (i.e., $a_1 + r, a_3 + r, a_1 + a_3 + r$), the user decodes a_1 and a_3. From the second query row, the user obtains $b_3 - b_1$ and $b_4 - b_2$, which are used to cancel the interference caused by W_2 in the third query row. After canceling the interference, the user decodes a_2 and a_4 from the third query row. Because $\mathbf{U}_1$ is full-rank and known by the user, the user can decode the 4 symbols of W_1 from $a_{(1:4)}$.

2) 1-Security: ecause r, s, t are uniformly distributed random variables independent of the message symbols, we have $I(W_1, W_2; A_n^{[k]}, Q_n^{[k]}) = I(W_1, W_2; A_n^{[k]} | Q_n^{[k]}) = H(A_n^{[k]} | Q_n^{[k]}) - H(A_n^{[k]} | W_1, W_2, Q_n^{[k]}) = H(A_n^{[k]} | Q_n^{[k]}) - H(r, s, t) = 0$. Therefore, from the answer of any server, the eavesdropper can learn nothing about the messages W_1, W_2.

3) 2-Privacy: Privacy is guaranteed by the observation that the answers of any 2 servers contain 4 linearly independent random combinations of symbols from each message. For example, Server 1 and Server 2 observe $a_{(1:4)}$ and $b_{(1:4)}$, which are linearly independent random combinations of the 2 messages. Server 1 and Server 3 observe

$$\begin{bmatrix} a_1, a_2, b_1, b_2 \\ a_1 + a_3, a_2 + a_4, -b_2 + b_3 + b_4, -b_1 - b_2 + b_3 + 2b_4 \end{bmatrix} \tag{15}$$

$$\overset{\text{invertible}}{\Longleftrightarrow} \begin{bmatrix} a_1, a_2, b_1, b_2 \\ a_3, a_4, b_3, b_4 \end{bmatrix} \tag{16}$$

Therefore, Server 1 and Server 3 also observe linearly independent random combinations of W_1 and W_2. The case where Server 2 and Server 3 collude follows similarly.

Because $\mathbf{U}_1$ and $\mathbf{U}_2$ are independently and uniformly chosen from all 4×4 full-rank matrices, the mapping from W_1 to $a_{(1:4)}$ is i.i.d. as the mapping from W_2 to $b_{(1:4)}$. The invertible relation in (16) indicates that the random variables on both sides of (16) are identically distributed (a more detailed proof of this argument appears in [2, Lemma 1]), so that any 2 servers cannot distinguish W_1 and W_2 and the scheme is private.

4) Achievable Rate: The scheme downloads 9 symbols in total, out of which the user can decode 4 desired symbols. Therefore, the rate achieved by this scheme is 4/9, which matches the capacity promised in Theorem 1. The achieved common randomness size ρ (1/3 of the total download) is also as promised.

In this section, we present a capacity achieving scheme for the smallest non-trivial setting. When we proceed to larger parameters, the main challenge is to design the linear combination coefficients of the desired and undesired symbols (named *precoding* or the *precoding matrices*), such that

1) Any T servers observe linearly independent combinations of symbols from each message (to guarantee T-privacy);
2) After the noise is cancelled, the interference of undesired symbols can be canceled as well;
3) After the noise and the interference of undesired symbols are cancelled, the desired symbols have full rank (to guarantee that the desired message can be recovered).

B. More Layers When the Number of Messages K Increases, $K = 3, N = 3, T = 2, E = 1$

We consider an example where the number of messages increases. In this case, the scheme has more layers and interestingly, the layers can be designed independently.

Suppose each message consists of 8 symbols from a finite field $\mathbb{F}_q$ with $q \geq 3$. The user privately chooses 3 matrices $\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3$ uniformly and independently from $GL_8(\mathbb{F}_q)$. Let $a_{(1:8)} = \mathbf{U}_1 W_1$, $b_{(1:8)} = \mathbf{U}_2 W_2$, and $c_{(1:8)} = \mathbf{U}_3 W_3$, and let $s_{(1:7)}$ contain the noise variables. The retrieval scheme for W_1 is shown below, which includes 3 layers separated by dashed lines.

Note that the scheme uses the same precoding in Section IV-A 3 times, *i.e.,* for the b symbols in layer 1 and

Server 1	Server 2	Server 3
$a_1 + s_1$	$a_5 + s_1$	$a_1 + a_5 + s_1$
$b_1 + s_2$	$b_5 + s_2$	$b_5 + (b_6 - b_2) + s_2$
$c_1 + s_3$	$c_5 + s_3$	$c_5 + (c_6 - c_2) + s_3$
$a_2 + b_2 + s_4$	$a_6 + b_6 + s_4$	$a_2 + a_6 + (b_5 - b_1) + (2b_6 - b_2) + s_4$
$a_3 + c_2 + s_5$	$a_7 + c_6 + s_5$	$a_3 + a_7 + (c_5 - c_1) + (2c_6 - c_2) + s_5$
$b_3 + c_3 + s_6$	$b_7 + c_7 + s_6$	$b_7 + c_7 + (b_8 + c_8 - b_4 - c_4) + s_6$
$a_4 + b_4 + c_4 + s_7$	$a_8 + b_8 + c_8 + s_7$	$a_4 + a_8 + (b_7 + c_7 - b_3 - c_3) + (2b_8 + 2c_8 - b_4 - c_4) + s_7$

layer 2, the c symbols in layer 1 and layer 2, and the $b + c$ symbols in layer 2 and layer 3. The design cross different layers may not be the same in general (for more details, refer to Section IV-F).

Security is obvious. Correctness and privacy can be checked similar to the proof in Section IV-A. The rate achieved is $8/21$, matching the capacity in Theorem 1.

C. Precoding Design When the Number of Servers N Increases, $K = 2, N = 4, T = 2, E = 1$

Assume each message consists of 9 symbols from a field $\mathbb{F}_q$ with $q \geq 9$ and the user wants to retrieve W_1. In the following retrieval table, the 4 symbols r, s, t, z are independent and uniform random symbols privately shared by the servers.

Server 1	Server 2	Server 3	Server 4
$a_1 + r$	$a_4 + r$	$a_7 + r$	$a_{10} + r$
$b_1 + s$	$b_4 + s$	$b_7 + s$	$b_{10} + s$
$a_2 + b_2 + t$	$a_5 + b_5 + t$	$a_8 + b_8 + t$	$a_{11} + b_{11} + t$
$a_3 + b_3 + z$	$a_6 + b_6 + z$	$a_9 + b_9 + z$	$a_{12} + b_{12} + z$

The user privately chooses 2 9×9 matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ from $GL_9(\mathbb{F}_q)$. Let $\mathbf{G}_1 = \mathbf{V}^4(\phi_1, \phi_2, \phi_3)$, *i.e.*, a Vandermonde matrix with distinct nonzero ϕ_i's, and $\phi_i \neq 1, i \in [1:3]$,

$$\mathbf{G}_1 = \begin{bmatrix} 1 & 1 & 1 \\ \phi_1 & \phi_2 & \phi_3 \\ \phi_1^2 & \phi_2^2 & \phi_3^2 \\ \phi_1^3 & \phi_2^3 & \phi_3^3 \end{bmatrix}. \tag{17}$$

Let $\mathbf{G}_2$ denote a 12×9 matrix,

$$\mathbf{G}_2 = \begin{bmatrix} \mathbf{I}^3 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}^3 \\ \mathbf{I}^3 - \mathbf{M}_1 & \mathbf{M}_1 \\ \mathbf{I}^3 - \mathbf{M}_2 & \mathbf{M}_2 \end{bmatrix}, \tag{18}$$

where $\mathbf{I}^3$ is the 3×3 identity matrix, and $\mathbf{M}_1$ and $\mathbf{M}_2$ are 3×3 matrices specified as follows. For 3 distinct nonzero elements ψ_1, ψ_2, ψ_3 that are not the 6-th roots of unity in $\mathbb{F}_q$, *i.e.*, $\psi_i^6 \neq 1, i \in [1:3]$, we set

$$\mathbf{M}_1 = \mathbf{V}^3(\psi_1, \psi_2, \psi_3) \cdot \text{diag}(\psi_1^3, \psi_2^3, \psi_3^3) \cdot [\mathbf{V}^3(\psi_1, \psi_2, \psi_3)]^{-1}, \tag{19}$$

$$\mathbf{M}_2 = \mathbf{V}^3(\psi_1, \psi_2, \psi_3) \cdot \text{diag}(\psi_1^6, \psi_2^6, \psi_3^6) \cdot [\mathbf{V}^3(\psi_1, \psi_2, \psi_3)]^{-1}. \tag{20}$$

$a_{(1:12)}$ are designed as follows, where each query row is precoded by $\mathbf{G}_1$.

$$[a_1, a_4, a_7, a_{10}]^T = \mathbf{G}_1 \mathbf{U}_1[(1:3), :]W_1, \tag{21}$$
$$[a_2, a_5, a_8, a_{11}]^T = \mathbf{G}_1 \mathbf{U}_1[(4:6), :]W_1, \tag{22}$$
$$[a_3, a_6, a_9, a_{11}]^T = \mathbf{G}_1 \mathbf{U}_1[(7:9), :]W_1, \tag{23}$$

$b_{(1:12)}$ are precoded by $\mathbf{G}_2$,

$$b_{(1:12)} = \mathbf{G}_2 \mathbf{U}_2[(1:6), :]W_2. \tag{24}$$

The description of the scheme is complete and we prove that the scheme is correct, private, secure and capacity achieving.

1) Correctness: We first cancel the noise.

Server 2 − Server 1	Server 3 − Server 1	Server 4 − Server 1
$a_4 - a_1$	$a_7 - a_1$	$a_{10} - a_1$
$b_4 - b_1$	$b_7 - b_1$	$b_{10} - b_1$
$(a_5 - a_2) + (b_5 - b_2)$	$(a_8 - a_2) + (b_8 - b_2)$	$(a_{11} - a_2) + (b_{11} - b_2)$
$(a_6 - a_3) + (b_6 - b_3)$	$(a_9 - a_3) + (b_9 - b_3)$	$(a_{12} - a_3) + (b_{12} - b_3)$

From the precoding of $b_{(1:12)}$ (24) and the structure of $\mathbf{G}_2$ (18), for the undesired symbols we have

$$\begin{bmatrix} b_7 - b_1 \\ b_8 - b_2 \\ b_9 - b_3 \\ b_{10} - b_1 \\ b_{11} - b_2 \\ b_{12} - b_3 \end{bmatrix} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \end{bmatrix} \cdot \begin{bmatrix} b_4 - b_1 \\ b_5 - b_2 \\ b_6 - b_3 \end{bmatrix}. \tag{25}$$

From the definition of $\mathbf{M}_1$ (19) and $\mathbf{M}_2$ (20), we know that $[\mathbf{I}^3; \mathbf{M}_1; \mathbf{M}_2] = \mathbf{V}^9(\psi_1, \psi_2, \psi_3) \cdot [\mathbf{V}^3(\psi_1, \psi_2, \psi_3)]^{-1}$ is the systematic form of a Vandermonde matrix, and thus is a 9×3 MDS matrix. Therefore, the 3 exposed symbols $b_4 - b_1, b_7 - b_1, b_{10} - b_1$ in the second query row are sufficient to reconstruct the remaining 6 symbols $b_5 - b_2, b_6 - b_3, b_8 - b_2, b_9 - b_3, b_{11} - b_2, b_{12} - b_3$ in the third and fourth query rows. Hence, the interference caused by $b_{(1:12)}$ can be cancelled and the user obtains

Server 2 − Server 1	Server 3 − Server 1	Server 4 − Server 1
$a_4 - a_1$	$a_7 - a_1$	$a_{10} - a_1$
$a_5 - a_2$	$a_8 - a_2$	$a_{11} - a_2$
$a_6 - a_3$	$a_9 - a_3$	$a_{12} - a_3$

From the definition of $\mathbf{G}_1$ (17), and the assumption that the ϕ_i's are nonzero, distinct, and $\phi_i \neq 1$, it is easy to check that each query row above contains 3 linear independent combinations of the desired symbols. Further, $\mathbf{U}_1$ has full rank and the user can invert the 9 linear independent combinations of the desired symbols to decode W_1.

TABLE I
EXAMPLE IN SECTION IV-D WITH $K = 2, N = 5, T = 3, E = 2$

Server 1	Server 2	Server 3	Server 4	Server 5
$a_1 + r_1$	$a_4 + r_2$	$a_7 + r_1 + r_2$	$a_{10} + r_1 + 2r_2$	$a_{13} + 2r_1 + 3r_2$
$b_1 + s_1$	$b_4 + s_2$	$b_7 + s_1 + s_2$	$b_{10} + s_1 + 2s_2$	$b_{13} + 2s_1 + 3s_2$
$a_2 + b_2 + t_1$	$a_5 + b_5 + t_2$	$a_8 + b_8 + t_1 + t_2$	$a_{11} + b_{11} + t_1 + 2t_2$	$a_{14} + b_{14} + 2t_1 + 3t_2$
$a_3 + b_3 + z_1$	$a_6 + b_6 + z_2$	$a_9 + b_9 + z_1 + z_2$	$a_{12} + b_{12} + z_1 + 2z_2$	$a_{15} + b_{15} + 2z_1 + 3z_2$

2) 2-Privacy: From the precoding of the desired symbols (21)-(23), the linear combinations of W_1 across query rows are independent. In addition, from the structure of $\mathbf{G}_1$ (17), within each query row, every 3 a_i symbols are linearly independent. Therefore, every 2 servers observe 6 linear independent combinations of W_1.

We next show that every 2 servers observe linear independent combinations of W_2. From (24), $\mathbf{G}_2$ expands 6 linear independent combinations of W_2 to 12 symbols, which are distributed to the servers. The precoding coefficients of the 6 b_i symbols observed by any 2 servers are given by the following 6 matrices (each of which is a submatrix of $\mathbf{G}_2$),

$$\begin{bmatrix} \mathbf{I}^3 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}^3 \end{bmatrix}, \begin{bmatrix} \mathbf{I}^3 & \mathbf{0} \\ \mathbf{I}^3 - \mathbf{M}_1 & \mathbf{M}_1 \end{bmatrix}, \begin{bmatrix} \mathbf{I}^3 & \mathbf{0} \\ \mathbf{I}^3 - \mathbf{M}_2 & \mathbf{M}_2 \end{bmatrix},$$
$$\begin{bmatrix} \mathbf{0} & \mathbf{I}^3 \\ \mathbf{I}^3 - \mathbf{M}_1 & \mathbf{M}_1 \end{bmatrix}, \begin{bmatrix} \mathbf{0} & \mathbf{I}^3 \\ \mathbf{I}^3 - \mathbf{M}_2 & \mathbf{M}_2 \end{bmatrix}, \begin{bmatrix} \mathbf{I}^3 - \mathbf{M}_1 & \mathbf{M}_1 \\ \mathbf{I}^3 - \mathbf{M}_2 & \mathbf{M}_2 \end{bmatrix}. \quad (26)$$

We are left to prove that these 6 matrices are invertible, which is is equivalent to prove that $\mathbf{M}_1, \mathbf{M}_2, \mathbf{I}^3 - \mathbf{M}_1, \mathbf{I}^3 - \mathbf{M}_2$ and $\mathbf{M}_1 - \mathbf{M}_2$ are invertible. This follows from (19), (20) and the assumption that $\psi_i \neq 0, \psi_i^6 \neq 1$.

Therefore, any 2 servers observe linear independent random combinations of W_1 and W_2 so that the mappings to W_1 and W_2 are i.i.d. and 2-privacy is guaranteed.

Security is easy to verify. The rate achieved is equal to the capacity, 9/16 and the size of common randomness used is 1/4 of the total download, as desired.

D. Precoding of Noise Symbols When $E > 1$, $K = 2, N = 5, T = 3, E = 2$

In the 3 examples above, $E = 1$ and all servers use the same noise (common randomness) symbols to add on the message symbols (*i.e.,* repetition code is used). When $E > 1$, the noise symbols are precoded by an (N, E)-MDS code. To illustrate how this is done, we present an example with $E = 2$.

Assume each message is comprised of 9 symbols from a sufficiently large field $\mathbb{F}_q$, and the user wants W_1. In the scheme described in Table I, the user downloads 20 symbols in total, achieving the capacity of 9/20.

In the retrieval table above, the 8 shared symbols $r_1, r_2, s_1, s_2, t_1, t_2, z_1, z_2$ are uniform and independent random variables. The precoding matrix and parity check matrix of the noise symbols are

$$\mathbf{C}_S = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 2 \\ 2 & 3 \end{bmatrix}, \quad \mathbf{H}_S = \begin{bmatrix} -1 & -1 & 1 & 0 & 0 \\ -1 & -2 & 0 & 1 & 0 \\ -2 & -3 & 0 & 0 & 1 \end{bmatrix}. \quad (27)$$

where the same precoding matrix is applied to the noise symbols in each query row. For example, in the first query row, r_1, r_2 is precoded by $\mathbf{C}_S$ to produce $r_1, r_2, r_1 + r_2$, $r_1 + 2r_2, 2r_1 + 3r_2$ (one for each server). 2-security follows by noting that $\mathbf{C}_S$ is an MDS matrix where any 2 rows are linearly independent.

In what follows, we present the precoding design of $a_{(1:15)}$ and $b_{(1:15)}$. The user privately chooses 2 matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ uniformly and independently from $GL_9(\mathbb{F}_q)$. The precoding of the desired message symbols occurs in each query row, by the same 5×3 matrix $\mathbf{G}_1$,

$$\mathbf{G}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 3 & 2 \\ 2 & 3 & 1 \end{bmatrix}. \quad (28)$$

The precoding of the undesired message symbols is over all 15 b_i symbols, the coefficients of which are described by a 15×9 matrix $\mathbf{G}_2$. Thus, $a_{(1:15)}$ and $b_{(1:15)}$ are generated by

$$[a_1, a_4, a_7, a_{10}, a_{13}]^T = \mathbf{G}_1 \mathbf{U}_1[(1 : 3), :]W_1, \quad (29)$$
$$[a_2, a_5, a_8, a_{11}, a_{14}]^T = \mathbf{G}_1 \mathbf{U}_1[(4 : 6), :]W_1, \quad (30)$$
$$[a_3, a_6, a_9, a_{12}, a_{15}]^T = \mathbf{G}_1 \mathbf{U}_1[(7 : 9), :]W_1, \quad (31)$$
$$b_{(1:15)} = \mathbf{G}_2 \mathbf{U}_2 W_2. \quad (32)$$

The matrix $\mathbf{G}_2$ will be specified later.

Upon receiving the answers, the user first cancels the noise by multiplying $\mathbf{H}_S$ with the answers in each query row, obtaining Table II, where after the noise symbols are cancelled, the 9×9 precoding matrices associated with $\mathbf{U}_1 W_1$ and $\mathbf{U}_2 W_2$ respectively are[1]

$$\mathbf{F}_1 = (\mathbf{H}_S \cdot \mathbf{G}_1) \otimes \mathbf{I}^3, \quad (33)$$
$$\mathbf{F}_2 = (\mathbf{H}_S \otimes \mathbf{I}^3) \cdot \mathbf{G}_2. \quad (34)$$

To successfully decode the 9 symbols of W_1, the b_i symbols in the third and fourth query rows should be cancelled by the symbols in the second query row; the 9 a_i symbols should be linearly independent. Therefore, the correctness criterion on $\mathbf{F}_1$ and $\mathbf{F}_2$ is

Correctness condition 1 (for desired symbols): $\mathbf{F}_1$ is non-singular.

---

[1] From (29)-(32), $\mathbf{F}_1$ is associated with $\mathbf{U}_1 W_1$ in an order specified by the query rows (e.g., the coefficients of the first two queries $a_7 - a_1 - a_4$ and $a_{10} - a_1 - 2a_4$ in the first query row are specified by the first two rows of $\mathbf{F}_1$, respectively); $\mathbf{F}_2$ is associated with $\mathbf{U}_2 W_2$ in lexicographic order (e.g., the coefficients of the two queries with smallest symbol indices in b, $b_7 - b_1 - b_4$ and $b_8 - b_2 - b_5$ are specified by the first two rows of $\mathbf{F}_2$, respectively). Note that these orders are not essential in the later proof.

TABLE II
EXAMPLE IN SECTION IV-D: AFTER CANCELLING THE NOISE

Server 3 − Server 1 − Server 2	Server 4 − Server 1 − 2Server 2	Server 5 − 2Server 1− 3Server 2
$a_7 - a_1 - a_4$	$a_{10} - a_1 - 2a_4$	$a_{13} - 2a_1 - 3a_4$
$b_7 - b_1 - b_4$	$b_{10} - b_1 - 2b_4$	$b_{13} - 2b_1 - 3b_4$
$a_8 - a_2 - a_5 + b_8 - b_2 - b_5$	$a_{11} - a_2 - 2a_5 + b_{11} - b_2 - 2b_5$	$a_{14} - 2a_2 - 3a_5 + b_{14} - 2b_2 - 3b_5$
$a_9 - a_3 - a_6 + b_9 - b_3 - b_6$	$a_{12} - a_3 - 2a_6 + b_{12} - b_3 - 2b_6$	$a_{15} - 2a_3 - 3a_6 + b_{15} - 2b_3 - 3b_6$

Correctness condition 2 (for undesired symbols): there exists a systematic MDS matrix $\mathbf{M}_{\text{interference}} = \begin{bmatrix} \mathbf{I}^3 \\ \mathbf{M}_1 \\ \mathbf{M}_2 \end{bmatrix}$ where $\mathbf{M}_1$ and $\mathbf{M}_2$ are 2 3×3 matrices, such that

$$\mathbf{F}_2 = \mathbf{M}_{\text{interference}} \cdot \mathbf{F}_2[(1:3), :]. \qquad (35)$$

Note that this condition is sufficient but not necessary to cancel the interference.

From (35), $\mathbf{G}_2$ can be represented in terms of $\mathbf{M}_1$ and $\mathbf{M}_2$,

$$\mathbf{G}_2 = \begin{bmatrix} \mathbf{I}^9 \\ \mathbf{I}^3 - \mathbf{M}_1, 2\mathbf{I}^3 - \mathbf{M}_1, \mathbf{M}_1 \\ 2\mathbf{I}^3 - \mathbf{M}_2, 3\mathbf{I}^3 - \mathbf{M}_2, \mathbf{M}_2 \end{bmatrix} \cdot \mathbf{G}_2[(1:9), :], \quad (36)$$

where we set $\mathbf{G}_2[(1:9), :] = \mathbf{I}^9$.

To guarantee 3-privacy, we require every 3 servers observe independent random linear combinations of both messages, which is translated into the following two conditions.

Privacy condition 1 (for desired symbols): 10 submatrices of $\mathbf{G}_1$, each of which is comprised of 3 distinct rows, are non-singular.

Privacy condition 2 (for undesired symbols): 10 submatrices of $\mathbf{G}_2$, each of which is comprised of 9 rows that specify the precoding coefficients of the b_i symbols observed by 3 distinct servers, are non-singular.

It is easy to verify that the 10 submatrices of $\mathbf{G}_1$ (refer to (28)) are invertible, and privacy condition 1 is satisfied. Further,

$$\mathbf{H}_S \cdot \mathbf{G}_1 = \begin{bmatrix} -1 & -1 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix}, \qquad (37)$$

which is invertible. Hence, $\mathbf{F}_1 = (\mathbf{H}_S \cdot \mathbf{G}_1) \otimes \mathbf{I}^3$ is invertible, and correctness condition 1 is satisfied.

Instead of finding an explicit choice of $\mathbf{G}_2$ as in previous sections (which is complicated in general because $\mathbf{G}_2$ contains certain non-trivial structure, e.g., see (36)), we show the existence of such $\mathbf{G}_2$ over a large field via the Schwartz Zippel lemma [12], [13]. In fact, by choosing the entries of $\mathbf{M}_1, \mathbf{M}_2$ uniformly and independently from a sufficiently large field, the conditions are satisfied with high probability so that almost all choices will work.

Note that $\mathbf{G}_2$ is determined by $\mathbf{M}_1$ and $\mathbf{M}_2$ and we will choose the 18 entries in $\mathbf{M}_1$ and $\mathbf{M}_2$ independently and uniformly over a sufficiently large field. Correctness condition 2 requires $\mathbf{M}_{\text{interference}} = [\mathbf{I}^3; \mathbf{M}_1; \mathbf{M}_2]$ to be MDS,

meaning that its $\binom{9}{3}$ submatrices (each of which contains 3 distinct rows from $\mathbf{M}_{\text{interference}}$) are invertible. It is easy to see that the determinant polynomial for each of them is not the zero-polynomial, then they are invertible with high probability by the Schwartz Zippel lemma [12], [13]. Privacy condition 2 requires 10 specific submatrices of $\mathbf{G}_2$ to be invertible. Similarly, we need to show that their determinant polynomials are non-zero, for which we defer the detailed proof to Section IV-F for the general case to avoid repetition.

Therefore, we have proved the existence of certain $\mathbf{G}_1, \mathbf{G}_2$ that produces a private and correct scheme.

E. Servers Observe Linearly Dependent Symbols When $T > N - E$, $K = 2, N = 4, T = 3, E = 2$

We consider the smallest case where $T > N - E$. In this regime, it turns out that the servers observe linearly dependent symbols. This phenomenon has not been observed in the literature and existing privacy proofs will fail. We tackle this challenge by increasing the message size (*i.e.,* append dummy variables that are globally known) so that the symbols (after lengthening) become linearly independent again.

Assume each message consists of 4 symbols from a field $\mathbb{F}_q$, and the user wants W_1. The scheme is described below.

Server 1	Server 2	Server 3	Server 4
$a_1 + r_1$	$a_3 + r_2$	$a_5 + r_1 + r_2$	$a_7 + r_1 + 2r_2$
$b_1 + s_1$	$b_3 + s_2$	$b_5 + s_1 + s_2$	$b_7 + s_1 + 2s_2$
$a_2+b_2+t_1$	$a_4+b_4+t_2$	$a_6+b_6+t_1+t_2$	$a_8+b_8+t_1+2t_2$

where the precoding matrix of the noise symbols r_1, r_2, s_1, s_2, t_1, t_2 and the parity check matrix are

$$\mathbf{C}_S = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{H}_S = \begin{bmatrix} -1 & -1 & 1 & 0 \\ -1 & -2 & 0 & 1 \end{bmatrix}. \qquad (38)$$

Note that every 3 servers observe 6 (mixed) symbols from each message and each message has 4 symbols, so the colluding servers do observe linear dependency in the symbols. To deal with this, we add 2 dummy variables (known to the user) to each message. Denote the extended messages by $\tilde{W}_1$ and $\tilde{W}_2$, each of length 6.

The user privately chooses 2 matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ uniformly and independently from $GL_6(\mathbb{F}_q)$, $a_{(1:8)}$ and $b_{(1:8)}$ are generated by

$$a_{(1:8)} = \overline{\mathbf{G}}_1 \mathbf{U}_1 \tilde{W}_1, \qquad (39)$$

$$b_{(1:8)} = \mathbf{G}_2 \mathbf{U}_2 \tilde{W}_2, \qquad (40)$$

where $\overline{\mathbf{G}}_1$ is an 8×6 matrix where each entry is chosen independently and uniformly from a sufficiently large field, and

$$\mathbf{G}_2^{8 \times 6} = \begin{bmatrix} \mathbf{I}^6 \\ \mathbf{I}^2 - \mathbf{M}, 2\mathbf{I}^2 - \mathbf{M}, \mathbf{M} \end{bmatrix}, \qquad (41)$$

where

$$\mathbf{M} = \begin{bmatrix} 1 & 1 \\ \psi_1 & \psi_2 \end{bmatrix} \cdot \mathrm{diag}(\psi_1^2, \psi_2^2) \cdot \begin{bmatrix} 1 & 1 \\ \psi_1 & \psi_2 \end{bmatrix}^{-1},$$
$$\psi_1 \neq \psi_2, \quad \psi_i^2 - 1 \neq 0, \quad \psi_i^2 - 2 \neq 0. \quad (42)$$

1) 3-Privacy: Note that the 4 submatrices of $\overline{\mathbf{G}}_1$ corresponding to 4 possible 3-colluding servers have non-zero determinant polynomials, then by the Schwartz Zippel lemma, the probability that $\overline{\mathbf{G}}_1$ works approaches 1 as q increases. With any such realization of $\overline{\mathbf{G}}_1$, every 3 servers observe linearly independent combinations of $\tilde{W}_1$. We may similarly check that every 3 servers observe linearly independent combinations of $\tilde{W}_2$. Hence, 3-privacy is guaranteed.

2) Correctness: After the noise symbols are cancelled, we have

Server 3−Server 1−Server 2	Server 4−Server 1−2Server 2
$a_5 - a_1 - a_3$	$a_7 - a_1 - 2a_3$
$b_5 - b_1 - b_3$	$b_7 - b_1 - 2b_3$
$a_6 - a_2 - a_4 + b_6 - b_2 - b_4$	$a_8 - a_2 - 2a_4 + b_8 - b_2 - 2b_4$

where the coefficients of the b symbols are

$$\mathbf{F}_2 = (\mathbf{H}_S \otimes \mathbf{I}^2) \cdot \mathbf{G}_2 = \begin{bmatrix} -\mathbf{I} & -\mathbf{I} & \mathbf{I} \\ -\mathbf{M} & -\mathbf{M} & \mathbf{M} \end{bmatrix} \quad (43)$$

$$= \begin{bmatrix} \mathbf{I} \\ \mathbf{M} \end{bmatrix} \cdot \mathbf{F}_2[(1:2), :], \quad (44)$$

where $[\mathbf{I}; \mathbf{M}]$ is MDS from (42). Then $b_6 - b_2 - b_4$ and $b_8 - b_2 - 2b_4$ can be constructed from $b_5 - b_1 - b_3$ and $b_7 - b_1 - 2b_3$. The coefficients of the 4 desired symbols of W_1 are

$$\overline{\mathbf{F}}_1 \cdot \mathbf{U}_1[:, (1:4)] = (\mathbf{H}_S \otimes \mathbf{I}^2) \cdot \overline{\mathbf{G}}_1 \cdot \mathbf{U}_1[:, (1:4)]. \quad (45)$$

To correctly decode the desired symbols, $\overline{\mathbf{F}}_1 \cdot \mathbf{U}_1[:, (1:4)]$ must be non-singular. Consider an arbitrary realization of $\mathbf{U}_1$. Note that $\mathbf{U}_1[:, (1:4)]$ has full-column rank, i.e., there exists 4 rows of $\mathbf{U}_1[:, (1:4)]$ that are linearly independent (assume to be the first 4 rows without loss of generality). Then, we can set the last 2 columns of $\overline{\mathbf{G}}_1$ to be zeros, and $\overline{\mathbf{G}}_1[:, (1:4)]$ to

$\mathbf{G}_1 \otimes \mathbf{I}^2$, where $\mathbf{G}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 2 \\ 2 & 4 \end{bmatrix}$. As a result, $(\mathbf{H}_S \otimes \mathbf{I}^2) \cdot \overline{\mathbf{G}}_1[:, (1:$

$4)] = (\mathbf{H}_S \cdot \mathbf{G}_1) \otimes \mathbf{I}^2$, where $\mathbf{H}_S \cdot \mathbf{G}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ is invertible.

Therefore, $(\mathbf{H}_S \otimes \mathbf{I}^2) \cdot \overline{\mathbf{G}}_1[:, (1:4)]$ and $\overline{\mathbf{F}}_1 \cdot \mathbf{U}_1[:, (1:4)]$ are invertible, and the determinant polynomial of $\overline{\mathbf{F}}_1 \cdot \mathbf{U}_1[:, (1:4)]$ is not the zero polynomial. By the Schwartz Zippel lemma, the determinant polynomial evaluates to a nonzero value with probability approaching 1 as q increases. Averaging over all possible choices of $\mathbf{U}_1$, the probability of $\overline{\mathbf{F}}_1 \cdot \mathbf{U}_1[:, (1:4)]$ being non-singular (hence we can decode correctly) tends to 1 as q increases.

TABLE III
THE STRUCTURE OF THE DOWNLOADED SYMBOL FROM SERVER n

	Server n
Layer 1:	$\{X_i^{[1]} + S_i^{[1]}\}_{i \in \mathcal{I}(n,1,1)}$ $\vdots$ $\{X_i^{[K]} + S_i^{[K]}\}_{i \in \mathcal{I}(n,1,K)}$
Layer 2:	$\{X_i^{[1]} + X_i^{[2]} + S_i^{[1,2]}\}_{i \in \mathcal{I}(n,2,1)}$ $\{X_i^{[1]} + X_i^{[3]} + S_i^{[1,3]}\}_{i \in \mathcal{I}(n,2,2)}$ $\vdots$ $\{X_i^{[1]} + X_i^{[K]} + S_i^{[1,K]}\}_{i \in \mathcal{I}(n,2,K-1)}$ $\{X_i^{[2]} + X_i^{[3]} + S_i^{[2,3]}\}_{i \in \mathcal{I}(n,2,K)}$ $\vdots$ $\{X_i^{[K-1]} + X_i^{[K]} + S_i^{[K-1,K]}\}_{i \in \mathcal{I}(n,2,\binom{K}{2})}$
Layers $3:K-1$	$\vdots$
Layer K	$\{X_i^{[1]} + X_i^{[2]} + \cdots + X_i^{[K]} + S_i^{[1:K]}\}_{i \in \mathcal{I}(n,K,1)}$

Security and capacity-achieving are easy to verify. The proof is thus complete.

F. General Parameters K, N, T, E

In this section, we present a capacity-achieving scheme for general parameters of K, N, T and E when $E < T$. We first introduce the structure of the retrieval scheme and then discuss the detailed precoding of the downloaded symbols.

1) Structure of the Retrieval Scheme: In our scheme, we force symmetry across the servers, *i.e.,* the structure of the downloaded symbols from all servers is the same. Table III shows the detailed structure for Server n and the structure is explained in detail as follows.

The downloaded symbols from Server n may be divided into K layers. For layer $\iota \in [1:K]$, each symbol is a mixture of symbols from ι distinct messages (hence called an ι-**sum**) and a distinct noise symbol. In particular, $X_i^{[k]}$ denotes a mixture of symbols from message W_k (corresponding to a_i for a mixture of symbols from W_1 in the examples), and $S_i^{[\mathcal{K}]}$ denotes a noise symbol added to an ι-sum with type $\mathcal{K} \subset [1:K]$ (where the type $\mathcal{K}$ is the set of message indices in the ι-sum). Each layer ι includes all $\binom{K}{\iota}$ different types, and for each type, a same number of ι-sums are downloaded. The types are ordered lexicographically and assigned a type index t (see Table IV). For each type $\{t_1, t_2, \ldots, t_\iota\} \subset [1:K]$ with type index t, the user downloads $|\mathcal{I}(n, \iota, t)|$ ι-sums $\{X_i^{[t_1]} + X_i^{[t_2]} + \cdots + X_i^{[t_\iota]} + S_i^{[t_1, t_2, \ldots, t_\iota]}\}$, where $i \in \mathcal{I}(n, \iota, t)$,

$$|\mathcal{I}(n, \iota, t)| \triangleq I_\iota = (N-T)^{\iota-1}(T-E)^{K-\iota}, \quad (46)$$

and $\mathcal{I}(n, \iota, t)$ denotes the index set of the ι-sums. All the indices are formulated by

$$\{\mathcal{I}(n, \iota, t)\}_{n=[1:N], \iota=[1:K], t=[1:\binom{K}{\iota})]} = [1:NL_n], \quad (47)$$

where L_n is calculated below in (49) and NL_n is the total number of mixtures of any message appeared in all N servers. The indices are enumerated from 1 to NL_n, and are distributed

Type index	Type of ι-sum	Index set	Downloaded symbols
1	$\{1, 2, \ldots, \iota - 2, \iota - 1, \iota\}$	$\mathcal{I}(n, \iota, 1)$	$\{X_i^{[1]} + X_i^{[2]} + \cdots + X_i^{[\iota]} + S_i^{[1:\iota]}\}_{i \in \mathcal{I}(n,\iota,1)}$
2	$\{1, 2, \ldots, \iota - 2, \iota - 1, \iota + 1\}$	$\mathcal{I}(n, \iota, 2)$	$\{X_i^{[1]} + \cdots + X_i^{[\iota-1]} + X_i^{[\iota+1]} + S_i^{[1,2,\ldots,\iota-1,\iota+1]}\}_{i \in \mathcal{I}(n,\iota,2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
t	$\{t_1, t_2, \ldots, t_\iota\}$	$\mathcal{I}(n, \iota, t)$	$\{X_i^{[t_1]} + X_i^{[t_2]} + \cdots + X_i^{[t_\iota]} + S_i^{[t_1, t_2, \ldots, t_\iota]}\}_{i \in \mathcal{I}(n,\iota,t)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\binom{K}{\iota}$	$\{K - \iota + 1, K - \iota + 2, \ldots, K\}$	$\mathcal{I}(n, \iota, \binom{K}{\iota})$	$\{X_i^{[K-\iota+1]} + X_i^{[K-\iota+2]} + \cdots + X_i^{[K]} + S_i^{[K-\iota+1:K]}\}_{i \in \mathcal{I}(n,\iota,\binom{K}{\iota})}$

from Server 1 to Server N, and at each server from layer 1 to layer K. (The same indices assignment rule is used in the examples.)

We calculate the rate achieved. The number of symbols downloaded from each server is

$$D_n = \sum_{\iota=1}^{K} \binom{K}{\iota} \cdot I_\iota$$

$$= \sum_{\iota=1}^{K} \binom{K}{\iota} \cdot (N - T)^{\iota-1}(T - E)^{K-\iota}$$

$$= \frac{(N - E)^K - (T - E)^K}{N - T}. \tag{48}$$

For any message W_k, the number of mixtures $X_i^{[k]}$ of symbols from W_k involved in the downloads from each server is

$$L_n = \sum_{\iota=1}^{K} \binom{K-1}{\iota-1} \cdot I_\iota$$

$$= \sum_{\iota=1}^{K} \binom{K-1}{\iota-1} \cdot (N - T)^{\iota-1}(T - E)^{K-\iota}$$

$$= (N - E)^{K-1}. \tag{49}$$

In Section IV-F2 below, we will show that with proper precoding design of the mixtures $\{X_i^{[k]}\}$ and the noise symbols $\{S_i^{[\mathcal{K}]}\}$, the scheme can decode $(N-E) \cdot L_n = (N-E)^K$ symbols from the desired message. Hence, the scheme achieves the rate

$$R = \frac{(N - E) \cdot L_n}{N \cdot D_n}$$

$$= \left(1 - \frac{E}{N}\right) \cdot \frac{(N - E)^{K-1}}{\frac{(N-E)^K - (T-E)^K}{N-T}}$$

$$= \left(1 - \frac{E}{N}\right) \cdot \frac{1 - \frac{T-E}{N-E}}{1 - \left(\frac{T-E}{N-E}\right)^K}. \tag{50}$$

In the following, we present the details of the precoding design and the justification of security, privacy, and correctness.

2) Precoding Design: Suppose each message consists of $L = (N - E)^K$ symbols from a sufficiently large field $\mathbb{F}_q$. If $T > N - E$, for ease of presentation, we add $(T + E - N)(N - E)^{K-1}$ dummy variables to each message to extend

the message length to $T(N - E)^{K-1}$. We abuse the notation by letting W_k denote the extended message. Define $N_{\text{eff}} = \max(N - E, T)$ and $L' = L_n \cdot N_{\text{eff}}$.

a) Precoding of the noise symbols: From Table III in Section IV-F1, the scheme downloads D_n symbols from each server and each symbol is added with a noise symbol. The noise symbols from each query row are precoded by an (N, E)-MDS matrix $\mathbf{C}_S$. Then the corresponding systematic parity check matrix is denoted by $\mathbf{H}_S$. From the dimension of the MDS code, we know that in total, $E \cdot D_n$ independent noise symbols are used, matching the minimum required size for common randomness. The MDS property of the noise precoding matrix translates to E-security.

b) Precoding of the desired symbols: Suppose the user wants to retrieve W_l. The user privately chooses K matrices $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_K$ independently and uniformly from $GL_{L'}(\mathbb{F}_q)$. Let $\overline{\mathbf{G}}_l^{NL_n \times N_{\text{eff}} L_n}$ be the precoding matrix of all $X_i^{[l]}$ symbols from W_l. Thus, all $N \cdot L_n$ mixtures of symbols from W_l in the retrieval scheme are generated by

$$\{X_i^{[l]}\}_{i \in \mathcal{I}([1:N],:,:)} = \overline{\mathbf{G}}_l^{NL_n \times N_{\text{eff}} L_n} \cdot \mathbf{U}_l W_l, \tag{51}$$

and then $\{X_i^{[l]}\}_{i \in \mathcal{I}([1:N],:,:)}$ are distributed evenly to each query row. To guarantee T-privacy, we require every T servers observe linearly independent random mixtures of symbols from W_l. That is

Privacy condition 1 (for desired symbols): The $\binom{N}{T}$ submatrices of $\overline{\mathbf{G}}_l^{NL_n \times N_{\text{eff}} L_n}$, each of which is comprised of $T L_n$ rows corresponding to a set of T servers, are non-singular.

From all the received symbols, the user cancels the noise by projecting the N symbols in each query row to the dual space of $\mathbf{C}_S$ (*i.e.*, multiplying with $\mathbf{H}_S$). The user further cancels the interference of undesired symbols (to be justified in the following). The coefficients of $\mathbf{U}_l W_l$ are given by $\overline{\mathbf{F}}_l = \left(\mathbf{H}_S^{(N-E) \times N} \otimes \mathbf{I}^{L_n}\right) \cdot \overline{\mathbf{G}}_l^{NL_n \times N_{\text{eff}} L_n}$. Therefore, we have the following condition for the correct decoding of W_l,

Correctness condition 1 (for desired symbols): The $L \times L$ square matrix $\overline{\mathbf{F}}_l \cdot \mathbf{U}_l[:, (1 : L)] = \left(\mathbf{H}_S^{(N-E) \times N} \otimes \mathbf{I}^{L_n}\right) \cdot \overline{\mathbf{G}}_l^{NL_n \times N_{\text{eff}} L_n} \cdot \mathbf{U}_l[:, (1 : L)]$ is non-singular.

In the following, we show explicit precoding design for the case $T \leq N - E$, and a random choice will work almost surely $T > N - E$.

Case 1 (Explicit Precoding Design When $T \leq N - E$): In this case, $N_{\text{eff}} = N - E$. Let the precoding matrix $\mathbf{C}_S$ of the noise symbols be the Vandermonde matrix $\mathbf{V}^N(\lambda_1, \ldots, \lambda_E)$. Further, let each query row be precoded by the same matrix $\mathbf{G}_l^{N \times (N-E)}$, where $\mathbf{G}_l$ is the Vandermonde matrix $\mathbf{V}^N(\phi_1, \ldots, \phi_{N-E})$. The precoding matrix in (51) is $\overline{\mathbf{G}}_l^{N L_n \times N_{\text{eff}} L_n} = \mathbf{G}_l^{N \times (N-E)} \otimes \mathbf{I}^{L_n}$. Then $\{X_i^{[l]}\}$ are independent across different query rows. Further, every T rows of $\mathbf{G}_l$ are linearly independent. Therefore, every T servers observe linearly independent random mixtures of symbols from W_l, and privacy condition 1 is satisfied.

In this case, $L' = L$, and $\mathbf{U}_l[:, (1:L)] = \mathbf{U}_l$ is non-singular. Note that $\overline{\mathbf{F}}_l = \left(\mathbf{H}_S^{(N-E) \times N} \otimes \mathbf{I}^{L_n} \right) \cdot \left(\mathbf{G}_l^{N \times (N-E)} \otimes \mathbf{I}^{L_n} \right) = (\mathbf{H}_S \cdot \mathbf{G}_l) \otimes \mathbf{I}^{L_n}$, so correctness condition 1 translates to the requirement of $\mathbf{H}_S \cdot \mathbf{G}_l$ being non-singular. This ensures that for any realization of $\mathbf{U}_l$, the user obtains $N - E$ linearly independent combinations in desired symbols from each query row. We require that $\{\lambda_1, \ldots, \lambda_E, \phi_1, \ldots, \phi_{N-E}\}$ are distinct nonzero elements from $\mathbb{F}_q$ such that the columns of $\mathbf{C}_S$ and $\mathbf{G}_l$ are linearly independent. Note that the parity check matrix $\mathbf{H}_S$ projects any vector that is in the span of the columns of $\mathbf{C}_S$ to zero. As the columns of $\mathbf{C}_S$ and $\mathbf{G}_l$ are linearly independent, we know that $\mathbf{H}_S \cdot \mathbf{G}_l$ is non-singular, and correctness condition 1 is satisfied.

Case 2 (Random Choices Work Almost Surely When $T > N - E$): In this case, $N_{\text{eff}} = T$. Let each entry of $\overline{\mathbf{G}}_l^{N L_n \times T L_n}$ be chosen independently and uniformly over $\mathbb{F}_q$. We show that as $q \to \infty$, the probability that $\overline{\mathbf{G}}_l$ satisfies correctness condition 1 approaches 1.

The proof relies on the Schwartz Zippel lemma [12], [13] about the roots of polynomials, where the variables of the polynomials to be considered are the entries of $\overline{\mathbf{G}}_l$.

Privacy condition 1 requires that $\binom{N}{T}$ submatrices of $\overline{\mathbf{G}}_l$ corresponding to $\binom{N}{T}$ possible sets of T-colluding servers are non-singular. It is evident that the determinant polynomial of each submatrix is not the zero polynomial. Therefore, as the field size q increases, the probability that each determinant polynomial evaluates to a nonzero value (the submatrix being non-singular) approaches 1. It follows that all $\binom{N}{T}$ submatrices are non-singular with probability approaching 1 as q increases. As a uniform choice of $\overline{\mathbf{G}}_l$ works almost surely, there are infinite many choices of feasible $\overline{\mathbf{G}}_l$. The final choice of $\overline{\mathbf{G}}_l$ will be uniformly chosen from this feasible set.

To check correctness condition 1, consider an arbitrary realization of $\mathbf{U}_l$. As $\mathbf{U}_l$ is a full-rank matrix, there exist L linearly independent rows in $\mathbf{U}_l[:, (1:L)]$, the indices of which are denoted by $\mathcal{L}_{\mathbf{U}_l}$. We argue that the determinant polynomial of $\overline{\mathbf{F}}_l \cdot \mathbf{U}_l[:, (1:L)] = (\mathbf{H}_S \otimes \mathbf{I}^{L_n}) \cdot \overline{\mathbf{G}}_l \cdot \mathbf{U}_l[:, (1:L)]$ is not the zero polynomial, because one can set the L columns of $\overline{\mathbf{G}}_l$ with indices $\mathcal{L}_{\mathbf{U}_l}$ to be independent of the null space of $\mathbf{H}_S$ (similar choice as in Case 1 suffices), and other columns to be zeros, such that $\overline{\mathbf{F}}_l \cdot \mathbf{U}_l[:, (1:L)]$ is non-singular. Therefore, by Schwartz Zippel lemma, the determinant polynomial of $\overline{\mathbf{F}}_l \cdot \mathbf{U}_l[:, (1:L)]$ evaluates to a nonzero value with probability

approaching 1 as the field size q increases. Averaging over all possible choice of $\mathbf{U}_l$, the probability of $\overline{\mathbf{F}}_l \cdot \mathbf{U}_l[:, (1:L)]$ being non-singular (hence we can decode correctly) tends to 1 as q increases.

c) Precoding of undesired symbols: We now consider the precoding for the undesired messages, $W_k, k \neq l$. We wish to ensure that the undesired symbols that appear in an ι-sum (thus from layer $\iota \in [1 : K - 1]$) with a type that does not contain the desired message W_l are sufficient to reconstruct the undesired symbols that appear in an $\iota + 1$-sum (thus from layer $\iota + 1$) with a type that contains the desired message W_l. If this is satisfied, then all interferences can be cancelled. Next, we proceed to consider the precoding of these undesired symbols in the pure undesired ι-sums and the $\iota + 1$-sums that contain W_l. The design for different ι is done separately and from $\iota = 1$ to $\iota = K - 1$. Consider an arbitrary fixed ι. Further, the design for different types of the undesired ι-sum is done separately and in increasing order of the type index. Consider an arbitrary undesired type $\mathcal{K} = \{k_1, \ldots, k_\iota\}$ where $l \notin \mathcal{K}$. Assume the lexicographic index of the type $\mathcal{K}$ among all types with cardinality ι (*i.e.,* in layer ι) is t, and the lexicographic index of type $\mathcal{K} \cup l$ in layer $\iota + 1$ is t'. To simplify the notation, $X_i^{[k_1]} + \cdots + X_i^{[k_\iota]}$ is denoted as $X_i^{[\mathcal{K}]}$. Therefore, we are considering the design of the symbols in Table V.

Recall that $|\mathcal{I}(n, \iota, t)| = I_\iota = (N - T)^{\iota-1}(T - E)^{K-\iota}$, $|\mathcal{I}(n, \iota, t')| = I_{\iota+1} = (N - T)^\iota (T - E)^{K-\iota-1}$, and $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_K$ are uniform $L' \times L'$ full-rank matrices, where $L' = N_{\text{eff}} \cdot (N - E)^{K-1}$. Let $\mathbf{G}_{\mathcal{K},\iota}$ denote an $N(I_\iota + I_{\iota+1}) \times T(I_\iota + I_{\iota+1})$ precoding matrix. The symbols $\{X_i^{[\mathcal{K}]}\}$ in the above table are generated by,

$$\{X_i^{[\mathcal{K}]}\}_{i \in \mathcal{I}([1:N],[\iota:\iota+1],\{t,t'\})} = \mathbf{G}_{\mathcal{K},\iota} \cdot \sum_{k \in \mathcal{K}} \mathbf{U}_k[\mathcal{J}_k, :]W_k, \quad (52)$$

where $\mathcal{J}_k$ is a vector with $T(I_\iota + I_{\iota+1})$ distinct elements in $[1 : L']$ that have not been used (by default, the rows of $\mathbf{U}_k$ are used from the first to the last). Then $\{X_i^{[\mathcal{K}]}\}$ are distributed evenly to the servers. Note that the precoding matrix $\mathbf{G}_{\mathcal{K},\iota}$ does not depend on the message index or the type index. In other words, the same precoding matrix is used for all undesired messages and all types. This is important to align the interference so that the interfering sum could be cancelled without knowing the individual terms in the sum.

The dimension of $\mathbf{G}_{\mathcal{K},\iota}$ is chosen to ensure T-privacy, which requires that each T servers observe linearly independent random combinations of symbols from W_k. This translates to the following condition,

> **Privacy condition 2 (for undesired symbols):** The $\binom{N}{T}$ submatrices of $\mathbf{G}_{\mathcal{K},\iota}$, each of which is comprised of T distinct block rows, are non-singular.

Note that $\mathbf{G}_{\mathcal{K},\iota}$ can be viewed as a block matrix consisting of NT square submatrices of dimension $(I_\iota + I_{\iota+1})$ and $\mathbf{G}_{\mathcal{K},\iota}^{I_\iota + I_{\iota+1}}[i, :]$ is a *block row* that is comprised of a row of square submatrices.

After multiplying the symbols in the above table with $\mathbf{H}_S = \left[(-\mathbf{P})^{(N-E) \times E} | \mathbf{I}^{N-E} \right]$ to cancel the noise, we have the

TABLE V
INTERFERENCE OF UNDESIRED SYMBOLS FROM LAYER ι TO LAYER $\iota + 1$

	Server 1	...	Server N
Layer ι:	$\left\{X_i^{[\mathcal{K}]} + S_i^{[\mathcal{K}]}\right\}_{i \in \mathcal{I}(1,\iota,t)}$	...	$\left\{X_i^{[\mathcal{K}]} + S_i^{[\mathcal{K}]}\right\}_{i \in \mathcal{I}(N,\iota,t)}$
Layer $\iota + 1$:	$\left\{X_i^{[l]} + X_i^{[\mathcal{K}]} + S_i^{[\mathcal{K},l]}\right\}_{i \in \mathcal{I}(1,\iota+1,t')}$	...	$\left\{X_i^{[l]} + X_i^{[\mathcal{K}]} + S_i^{[\mathcal{K},l]}\right\}_{i \in \mathcal{I}(N,\iota+1,t')}$

precoding matrix of $\mathbf{U}_k[\mathcal{J}_k,:]W_k$ as

$$\mathbf{F}_{\mathcal{K},\iota} = \left(\mathbf{H}_S \otimes \mathbf{I}^{I_\iota + I_{\iota+1}}\right) \cdot \mathbf{G}_{\mathcal{K},\iota} \tag{53}$$

$$= \begin{bmatrix} \mathbf{G}_{\mathcal{K},\iota}^{I_\iota + I_{\iota+1}}[E+1,:] - \mathbf{P}[1,:] \cdot \mathbf{G}_{\mathcal{K},\iota}^{I_\iota + I_{\iota+1}}[(1:E),:] \\ \vdots \\ \mathbf{G}_{\mathcal{K},\iota}^{I_\iota + I_{\iota+1}}[N,:] - \mathbf{P}[N-E,:] \cdot \mathbf{G}_{\mathcal{K},\iota}^{I_\iota + I_{\iota+1}}[(1:E),:] \end{bmatrix}. \tag{54}$$

To guarantee correct decoding of the desired message symbols, we require that $\mathbf{F}_{\mathcal{K},\iota}$ has dependent rows in a way such that the interference of W_k can be cancelled. Specifically, after noise cancellation, there are $(N-E)I_\iota \{X_i^{[\mathcal{K}]}\}$ symbols in layer ι and we wish to ensure that they can reconstruct the $(N-E)I_{\iota+1} \{X_i^{[\mathcal{K}]}\}$ symbols (mixed with $\{X_i^{[l]}\}$) in layer $\iota + 1$. To this end, we require that the $\{X_i^{[\mathcal{K}]}\}$ symbols in layer ι and layer $\iota + 1$ form an MDS code. Therefore, we have the following (sufficient but not necessary) condition on correctness decoding,

Correctness condition 2 (for undesired symbols): There exists an MDS matrix $\mathbf{M}_{\text{interference}}$ of dimension $(N-E)(I_\iota + I_{\iota+1}) \times (T-E)(I_\iota + I_{\iota+1})$, such that

$$\mathbf{F}_{\mathcal{K},\iota} = \mathbf{M}_{\text{interference}}\mathbf{F}_{\mathcal{K},\iota}^{I_\iota + I_{\iota+1}}[(1:T-E),:]. \tag{55}$$

Note that $(N-E) \cdot I_\iota = (T-E) \cdot (I_\iota + I_{\iota+1})$.

In the following, we expand $\mathbf{M}_{\text{interference}}$ and (55) to represent $\mathbf{G}_{\mathcal{K},\iota}$ by $\mathbf{M}_{\text{interference}}$, such that the privacy condition 2 can also be translated into a condition on $\mathbf{M}_{\text{interference}}$.

Let $\{\mathbf{M}_{i,j}\}_{i \in [1:N-T], j \in [1:T-E]}$ be $(N-T)(T-E)$ square matrices of dimension $I_\iota + I_{\iota+1}$, and $\mathbf{M}_{\text{interference}}$ be a systematic MDS matrix such that

$$\mathbf{M}_{\text{interference}} = \begin{bmatrix} & \mathbf{I}^{(T-E)(I_\iota + I_{\iota+1})} & \\ \mathbf{M}_{1,1} & \cdots & \mathbf{M}_{1,T-E} \\ \vdots & \vdots & \vdots \\ \mathbf{M}_{N-T,1} & \cdots & \mathbf{M}_{N-T,T-E} \end{bmatrix}. \tag{56}$$

Then (55) is equivalent to

$$\begin{bmatrix} \mathbf{G}_{\mathcal{K},\iota}^{I_\iota + I_{\iota+1}}[T+1,:] - \mathbf{P}[T-E+1,:] \cdot \mathbf{G}_{\mathcal{K},\iota}^{I_\iota + I_{\iota+1}}[(1:E),:] \\ \vdots \\ \mathbf{G}_{\mathcal{K},\iota}^{I_\iota + I_{\iota+1}}[N,:] - \mathbf{P}[N-E,:] \cdot \mathbf{G}_{\mathcal{K},\iota}^{I_\iota + I_{\iota+1}}[(1:E),:] \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{M}_{1,1} & \cdots & \mathbf{M}_{1,T-E} \\ \vdots & \vdots & \vdots \\ \mathbf{M}_{N-T,1} & \cdots & \mathbf{M}_{N-T,T-E} \end{bmatrix}$$

$$\cdot \begin{bmatrix} \mathbf{G}_{\mathcal{K},\iota}^{I_\iota + I_{\iota+1}}[E+1,:] - \mathbf{P}[1,:] \cdot \mathbf{G}_{\mathcal{K},\iota}^{I_\iota + I_{\iota+1}}[(1:E),:] \\ \vdots \\ \mathbf{G}_{\mathcal{K},\iota}^{I_\iota + I_{\iota+1}}[T,:] - \mathbf{P}[T-E,:] \cdot \mathbf{G}_{\mathcal{K},\iota}^{I_\iota + I_{\iota+1}}[(1:E),:]. \end{bmatrix} . \tag{57}$$

This imposes the structure of $\mathbf{G}_{\mathcal{K},\iota}$ to be

$$\mathbf{G}_{\mathcal{K},\iota} = \underbrace{\begin{bmatrix} & & \mathbf{I}^{T(I_\iota + I_{\iota+1})} & & & \\ \hline \mathbf{N}_{1,1} & \cdots & \mathbf{N}_{1,E} & \mathbf{M}_{1,1} & \cdots & \mathbf{M}_{1,T-E} \\ \vdots & & \vdots & \vdots & & \vdots \\ \mathbf{N}_{N-T,1} & \cdots & \mathbf{N}_{N-T,E} & \mathbf{M}_{N-T,1} & \cdots & \mathbf{M}_{N-T,T-E} \end{bmatrix}}_{\triangleq \tilde{\mathbf{G}}_{\mathcal{K},\iota}}$$

$$\cdot \mathbf{G}_{\mathcal{K},\iota}^{I_\iota + I_{\iota+1}}[(1:T),:], \tag{58}$$

where $\mathbf{N}_{i,j} = \mathbf{P}[T-E+i,j] \cdot \mathbf{I}^{I_\iota + I_{\iota+1}} - \sum_{t=1}^{T-E} \mathbf{P}[t,j] \cdot \mathbf{M}_{i,t}$. We set $\mathbf{G}_{\mathcal{K},\iota}^{I_\iota + I_{\iota+1}}[(1:T),:]$ as the identity matrix such that $\mathbf{G}_{\mathcal{K},\iota} = \tilde{\mathbf{G}}_{\mathcal{K},\iota}$.

Therefore, the privacy condition 2 is equivalent to

Privacy condition 2' (for undesired symbols): There exist $(N-T)(T-E)$ square matrices $\{\mathbf{M}_{i,j}\}_{i \in [1:N-T], j \in [1:T-E]}$ of dimension $I_\iota + I_{\iota+1}$, such that $\binom{N}{T}$ submatrices of $\tilde{\mathbf{G}}_{\mathcal{K},\iota}$, each of which is comprised of distinct T block rows, are non-singular.

Now correctness condition 2 and privacy condition 2' are both stated in terms of $\mathbf{M}_{i,j}$. We next proceed to show the existence of $\mathbf{M}_{i,j}$ such that both conditions are satisfied. To this end, we will pick each entry of $\mathbf{M}_{i,j}$ independently and uniformly from a sufficiently large field.

First, consider correctness condition 2, which requires $\mathbf{M}_{\text{interference}}$ to be MDS. From (56), it is easy to see that over a sufficiently large finite field, by randomly choosing $\mathbf{M}_{i,j}$, $\mathbf{M}_{\text{interference}}$ is MDS with high probability.

Second, consider privacy condition 2', which requires the submatrices of $\tilde{\mathbf{G}}_{\mathcal{K},\iota}$ (made up of T block rows) to be non-singular. For these submatrices, we have three cases.

- Case 1: The submatrix consists of the first T block rows, which is an identity matrix and therefore is non-singular directly.
- Case 2: The submatrix consists of t block rows from the first T block rows, and $T-t$ block rows from the remaining $N-T$ block rows. We show that by elementary row and column operations and a proper choice of a realization of $\mathbf{M}_{i,j}$, the matrix is non-singular.
  First consider the first t block rows. Each one of the first t block rows contains one identity matrix, and by exchanging columns, we have the identity matrix on the

upper left corner and then the upper right corner is a zero matrix (see (60)). It remains to show that the lower right corner (denoted by $\mathbf{M}^{T-t}$ in (59)) is non-singular.

The last $T-t$ block rows depend on the $\{\mathbf{M}_{i,j}\}$ matrices. Here we have 3 cases: 1) $\mathbf{M}^{T-t}$ contains only $\{\mathbf{M}_{i,j}\}$ matrices, 2) $\mathbf{M}^{T-t}$ contains both $\{\mathbf{M}_{i,j}\}$ and $\{\mathbf{N}_{i,j}\}$ matrices, and 3) $\mathbf{M}^{T-t}$ contains only $\{\mathbf{N}_{i,j}\}$ matrices. In the following, we prove the non-singularity of $\mathbf{M}^{T-t}$ for all 3 cases.

$$
\begin{bmatrix}
\mathbf{I} & 0 & 0 & \\
0 & \ddots & 0 & \mathbf{0}^{t \times (T-t)} \\
0 & 0 & \mathbf{I} & \\
\hdashline
\multicolumn{3}{c}{\cdots\cdots\cdots} & \mathbf{M}^{T-t}
\end{bmatrix}
\tag{59}
$$

Case 2.1: $\mathbf{M}^{T-t}$ consists of only $\{\mathbf{M}_{i,j}\}$ matrices (this may happen when $t \geq E$) with distinct (i,j) indices and therefore is non-singular by setting it to identity.

Case 2.2: $\mathbf{M}^{T-t}$ consists of $T-t-j$ block columns of $\{\mathbf{M}_{i,j}\}$ matrices and j block columns of $\{\mathbf{N}_{i,j}\}$ matrices, where $\max(E-t,0) \leq j \leq E$. We rearrange $\{\mathbf{M}_{i,j}\}$ matrices to the left so that $\mathbf{M}^{T-t}$ is in the form of (60). Note that this operation does not change the upper right zero matrix in (59). It is easy to see that the upper left part in (60) is non-singular. The lower left part (which contains $\mathbf{M}_{i,j}$ matrices) is set to the zero matrix. The lower right part is the Kronecker product of a square submatrix of $\mathbf{P}$ (of dimension $j \leq E$) and the identity matrix (note that the $\{\mathbf{M}_{i,j}\}$ matrices in these $\{\mathbf{N}_{i,j}\}$ matrices in the lower right part are set to $\mathbf{0}$ so that we are left with only $\mathbf{P}^j$). Remember that $\mathbf{P}$ is MDS, so every square submatrix of $\mathbf{P}$ is non-singular. Then, $\mathbf{P}^j \otimes \mathbf{I}$ is also non-singular and $\mathbf{M}^{T-t}$ has full rank.

$$
\mathbf{M}^{T-t} = \begin{bmatrix}
\mathbf{M}_{i_1,1} & \cdots & \mathbf{M}_{i_1,T-t-j} & \vdots \\
\vdots & \ddots & \vdots & \vdots \\
\mathbf{M}_{i_{T-t-j},1} & \cdots & \mathbf{M}_{i_{T-t-j},T-t-j} & \\
\hdashline
\multicolumn{3}{c}{\mathbf{0}^{j \times (T-t-j)}} & \mathbf{P}^j \otimes \mathbf{I}
\end{bmatrix}
\tag{60}
$$

Case 2.3: $\mathbf{M}^{T-t}$ consists of only $\{\mathbf{N}_{i,j}\}$ matrices (this may happen when $t \geq T-E$) with distinct (i,j) indices. In this case, we set all the $\{\mathbf{M}_{i,j}\}$ matrices which appear in $\{\mathbf{N}_{i,j}\}$ matrices to be zero matrices. Then

$$
\mathbf{M}^{T-t} = \mathbf{P}^{T-t} \otimes \mathbf{I},
\tag{61}
$$

where $T-t \leq E$. Similar to the arguments in Case 2.2, $\mathbf{P}^{T-t}$ is non-singular and so is $\mathbf{M}^{T-t}$.

- Case 3: The submatrix consists of T block rows from the last $N-T$ blocks rows of $\tilde{\mathbf{G}}_{\mathcal{K},t}$. This case corresponds to Case 2.2 when $t=0$, and the submatrix is the same as that in (60) with $t=0$ and $j=E$. In this case, we must have $T \leq N-T$ such that $N \geq 2T > 2E$, and $N-E > E$. Then the $E \times E$ submatrix $\mathbf{P}^E$ of $\mathbf{P}$ is non-singular and so is the submatrix.

The analysis above shows that the determinant polynomials of the required submatrices of $\tilde{\mathbf{G}}_{\mathcal{K},t}$ are nonzero. Then, by the Schwartz Zippel lemma [12], [13], a uniform random choice of

$\mathbf{M}_{i,j}$ from a sufficiently large field will ensure that with high probability, the determinants of the submatrices are nonzero. Therefore, the submatrices are non-singular and note that this holds for undesired symbols with all types and all layers. Further, for the symbols designed for each type and each two layers, disjoint row vectors from $\mathbf{U}_k$ are multiplied with $\mathbf{G}_{\mathcal{K},t}$ and in total $T L_n$ vectors from $\mathbf{U}_k$ are used. Then by Lemma 1 in [2], the precoding coefficients for $W_k, k \in [1:K]$ are identically distributed as $\mathbf{U}_k[(1:T L_n);:]$. As $\mathbf{U}_k$ are identically distributed for all k, the message indices are equally likely and can not be distinguished. In other words, any T servers see linearly independent uniformly random combinations of both the desired message and undesired messages, and T-privacy is guaranteed.

To sum up, we have completed the description of the precoding of noise, desired and undesired symbols, and have proved that the scheme is secure, correct, private and the rate achieved is equal to the capacity. The achievability proof is thus complete.

V. ACHIEVABILITY WHEN $E \geq T$

Proving achievability in this case is relatively straightforward and follows from that in [11]. We present the proof here for completeness.

Suppose each message is comprised of $N-E$ symbols from $\mathbb{F}_q$ with $q > E$. Let $W_{(1:K)} = (W_1^{[1]}, \ldots, W_{N-E}^{[1]}, \ldots, W_1^{[K]}, \ldots, W_{N-E}^{[K]})$ represent the symbols of all K messages. The user wants to retrieve $W_k = (W_1^{[k]}, \ldots, W_{N-E}^{[k]})$ privately from the servers.

The queries are generated in the following way. The user firstly generates E independent uniformly random vectors $U_1, \ldots, U_E$ of length $K(N-E)$, and then chooses an (N, E)-GRS (generalized Reed-Solomon) code with generating matrix $\mathbf{G}_{(N,E)}$. Let $\Lambda = (\lambda_1, \ldots, \lambda_N) \in \mathbb{F}_q^N$ where $\lambda_i \neq \lambda_j$. Let $\Phi = (\phi_1, \ldots, \phi_N) \in \mathbb{F}_q^N$ where $\phi_i \neq 0$. The generating matrix $\mathbf{G}_{(N,E)}$ is defined by

$$
\mathbf{G}_{(N,E)} = \mathbf{V}^E(\lambda_1, \ldots, \lambda_N) \cdot \text{diag}(\Phi).
\tag{62}
$$

Let $e_i^{[k]}$ denote the length-$(K(N-E))$ unit vector where only the $((k-1)(N-E)+i)$-th entry is 1 and all the other entries are equal to 0. The purpose of $e_i^{[k]}$ is to retrieve the i-th entry of W_k. The query vectors are generated by

$$
[Q_1^{[k]}, \ldots, Q_N^{[k]}]
$$
$$
= [U_1, \ldots, U_E] \cdot \mathbf{G}_{(N,E)} + [0, \ldots, 0, e_1^{[k]}, \ldots, e_{N-E}^{[k]}].
\tag{63}
$$

The nodes share E common randomness symbols $(S_1, \ldots, S_E) = S_{(1:E)}$ that are uniformly and independently chosen from $\mathbb{F}_q$. The servers generate their answers by taking the inner product of the query vector and the stored data vector, then adding a linear combination of the common randomness,

$$
A_n^{[k]} = \langle Q_n^{[k]}, W_{(1:K)} \rangle + \langle \mathbf{G}_{(N,E)}[:,n], S_{(1:E)} \rangle.
\tag{64}
$$

Let $X_j = \langle U_j, W_{(1:K)} \rangle + S_j, j \in [1:E]$, the answers received by the user can be presented by

$$
[A_1^{[k]}, \ldots, A_N^{[k]}]
$$
$$
= [X_1, \ldots, X_E, W_1^{[k]}, \ldots, W_{N-E}^{[k]}] \cdot \begin{bmatrix} \mathbf{G}_{(N,E)} \\ \mathbf{0} \quad \mathbf{I} \end{bmatrix},
\tag{65}
$$

where the dimensions of $\mathbf{0}$ and $\mathbf{I}$ are clear from the context.

A. Correctness

From (62), $\begin{bmatrix} \mathbf{G}_{(N,E)} \\ \mathbf{0} \quad \mathbf{I} \end{bmatrix}$ is invertible such that the user can obtain W_k.

B. T-Privacy

The T-privacy constraint is guaranteed, because from (63), every E query vectors are independently and uniformly distributed. Note that $E \geq T$, so every T nodes can not tell which message the user requests. Note also that E-privacy is obtained for free.

C. E-Security

Guaranteed by the MDS property of the GRS code, *i.e.*, any E columns of $\mathbf{G}_{(N,E)}$ are linearly independent.

Remark 1: Besides the desired message, the user could further obtain $[X_1, \ldots, X_E]$ from (65). Theses E symbols are protected by independent noise symbols $S_{(1:E)}$. Therefore, the user learns no extra information beyond the desired message, *i.e.*, symmetric PIR constraint is satisfied at no additional cost.

VI. CONVERSE WHEN $E < T$

Let us start with a few lemmas. The lemma below shows that the answers of a set of servers do not depend on the queries when other messages are desired or the queries to other servers, after conditioning on their own queries.

Lemma 1: For any set of nodes $\mathcal{N} \subset [1:N]$, and any set of message index $\mathcal{K} \subset [1:K]$,

$$H(A_{\mathcal{N}}^{[k]}|\mathcal{Q}, W_{\mathcal{K}}, Q_{\mathcal{N}}^{[k]}) = H(A_{\mathcal{N}}^{[k]}|W_{\mathcal{K}}, Q_{\mathcal{N}}^{[k]}); \quad (66)$$

$$H(A_{\mathcal{N}}^{[k]}|W_{\mathcal{K}}, Q_{[1:N]}^{[k]}) = H(A_{\mathcal{N}}^{[k]}|W_{\mathcal{K}}, Q_{\mathcal{N}}^{[k]}). \quad (67)$$

Proof: As $I(A_{\mathcal{N}}^{[k]}; \mathcal{Q}|W_{\mathcal{K}}, Q_{\mathcal{N}}^{[k]}) \geq 0$ holds trivially, we only need to show $I(A_{\mathcal{N}}^{[k]}; \mathcal{Q}|W_{\mathcal{K}}, Q_{\mathcal{N}}^{[k]}) \leq 0$. The proof is presented next.

$$I(A_{\mathcal{N}}^{[k]}; \mathcal{Q}|W_{\mathcal{K}}, Q_{\mathcal{N}}^{[k]})$$
$$\leq I(A_{\mathcal{N}}^{[k]}, W_{[1:K]}, S; \mathcal{Q}|W_{\mathcal{K}}, Q_{\mathcal{N}}^{[k]}) \quad (68)$$
$$= I(W_{[1:K]}, S; \mathcal{Q}|W_{\mathcal{K}}, Q_{\mathcal{N}}^{[k]})$$
$$\quad + I(A_{\mathcal{N}}^{[k]}; \mathcal{Q}|W_{[1:K]}, S, W_{\mathcal{K}}, Q_{\mathcal{N}}^{[k]}) \quad (69)$$
$$\overset{(a)}{=} I(W_{[1:K]}, S; \mathcal{Q}|W_{\mathcal{K}}, Q_{\mathcal{N}}^{[k]}) \quad (70)$$
$$= H(W_{[1:K]}, S|W_{\mathcal{K}}, Q_{\mathcal{N}}^{[k]}) - H(W_{[1:K]}, S|W_{\mathcal{K}}, \mathcal{Q}) \quad (71)$$
$$\overset{(b)}{\leq} H(W_{[1:K]\backslash\mathcal{K}}, S) - H(W_{[1:K]\backslash\mathcal{K}}, S|\mathcal{Q}) \quad (72)$$
$$= I(W_{[1:K]\backslash\mathcal{K}}, S; \mathcal{Q}) \quad (73)$$
$$\overset{(c)}{=} 0, \quad (74)$$

where (a) holds because the answers are deterministic functions of the messages, common randomness, and queries. (b) and (c) follow from the independence of the messages, queries and common randomness (refer to (5)(6)).

Note that we allow $\mathcal{K}$ to be the empty set. (67) follows from (66) and is the form that will be used in the remaining proofs. $\qquad\square$

The following lemma allows us to split the answers from all N servers into two parts, one from E servers and the other from the remaining $N - E$ servers.

Lemma 2: For any node set $\mathcal{E} \subset [1:N]$ and $|\mathcal{E}| = E$,

$$\left(1 - \frac{E}{N}\right) H(A_{[1:N]}^{[1]}|Q_{[1:N]}^{[1]})$$
$$\geq L - o(L) + H(A_{[1:N]\backslash\mathcal{E}}^{[1]}|W_1, A_{\mathcal{E}}^{[1]}, Q_{[1:N]}^{[1]}). \quad (75)$$

Proof: From the correctness constraint, $H(W_1|A_{[1:N]}^{[1]}, Q_{[1:N]}^{[1]}) = o(L)$, we have

$$H(A_{[1:N]}^{[1]}|Q_{[1:N]}^{[1]}) \quad (76)$$
$$= H(W_1|Q_{[1:N]}^{[1]}) + H(A_{[1:N]}^{[1]}|W_1, Q_{[1:N]}^{[1]})$$
$$\quad - H(W_1|A_{[1:N]}^{[1]}, Q_{[1:N]}^{[1]}) \quad (77)$$
$$= L + H(A_{[1:N]}^{[1]}|W_1, Q_{[1:N]}^{[1]}) - o(L) \quad (78)$$
$$= L - o(L) + H(A_{\mathcal{E}}^{[1]}|W_1, Q_{[1:N]}^{[1]})$$
$$\quad + H(A_{[1:N]\backslash\mathcal{E}}^{[1]}|W_1, A_{\mathcal{E}}^{[1]}, Q_{[1:N]}^{[1]}) \quad (79)$$
$$\overset{(a)}{=} L - o(L) + H(A_{\mathcal{E}}^{[1]}|Q_{[1:N]}^{[1]})$$
$$\quad + H(A_{[1:N]\backslash\mathcal{E}}^{[1]}|W_1, A_{\mathcal{E}}^{[1]}, Q_{[1:N]}^{[1]}) \quad (80)$$
$$\overset{(b)}{\geq} L - o(L) + \frac{E}{N} H(A_{[1:N]}^{[1]}|Q_{[1:N]}^{[1]})$$
$$\quad + H(A_{[1:N]\backslash\mathcal{E}}^{[1]}|W_1, A_{\mathcal{E}}^{[1]}, Q_{[1:N]}^{[1]}), \quad (81)$$

where in (a), $\mathcal{E}$ can be any node set with size E. (b) follows by averaging over all $\mathcal{E} \in [1:N]$ with size E, and applying Han's inequality [14]. $\qquad\square$

For the answers from $N - E$ servers, we have the following recursive relationship.

Lemma 3: For any $k \in [1:K-1]$, and for any node set $\mathcal{E} \subset [1:N]$ and $|\mathcal{E}| = E$,

$$H(A_{[1:N]\backslash\mathcal{E}}^{[k]}|W_{[1:k]}, A_{\mathcal{E}}^{[k]}, Q_{[1:N]}^{[k]})$$
$$\geq \frac{T-E}{N-E} \cdot \Big(H(A_{[1:N]\backslash\mathcal{E}}^{[k+1]}|W_{[1:k+1]}, A_{\mathcal{E}}^{[k+1]},$$
$$Q_{[1:N]}^{[k+1]}) + L - o(L) \Big). \quad (82)$$

Proof: For any set $\mathcal{T}' \in [1:N]\backslash\mathcal{E}$ with size $|\mathcal{T}'| = T-E$,

$$H(A_{[1:N]\backslash\mathcal{E}}^{[k]}|W_{[1:k]}, A_{\mathcal{E}}^{[k]}, Q_{[1:N]}^{[k]}) \quad (83)$$
$$\geq H(A_{\mathcal{T}'}^{[k]}|W_{[1:k]}, A_{\mathcal{E}}^{[k]}, Q_{[1:N]}^{[k]}) \quad (84)$$
$$\overset{(a)}{=} H(A_{\mathcal{T}'}^{[k]}|W_{[1:k]}, A_{\mathcal{E}}^{[k]}, Q_{\mathcal{T}'\cup\mathcal{E}}^{[k]}) \quad (85)$$
$$\overset{(b)}{=} H(A_{\mathcal{T}'}^{[k+1]}|W_{[1:k]}, A_{\mathcal{E}}^{[k+1]}, Q_{\mathcal{T}'\cup\mathcal{E}}^{[k+1]}) \quad (86)$$
$$\overset{(c)}{=} H(A_{\mathcal{T}'}^{[k+1]}|W_{[1:k]}, A_{\mathcal{E}}^{[k+1]}, Q_{[1:N]}^{[k+1]}), \quad (87)$$

where (a) and (c) follow from Lemma 1, and (b) is due to the privacy constraint (10).

Averaging over all $\mathcal{T}' \in [1 : N] \setminus \mathcal{E}$ with size $T - E$ and using Han's inequality [14], we have

$$H(A_{[1:N]\setminus\mathcal{E}}^{[k]}|W_{[1:k]}, A_{\mathcal{E}}^{[k]}, Q_{[1:N]}^{[k]}) \tag{88}$$

$$\geq \frac{1}{\binom{N-E}{T-E}} \sum_{\substack{\mathcal{T}'\subset[1:N]\setminus\mathcal{E} \\ |\mathcal{T}'|=T-E}} H(A_{\mathcal{T}'}^{[k+1]}|W_{[1:k]}, A_{\mathcal{E}}^{[k+1]}, Q_{[1:N]}^{[k+1]}) \tag{89}$$

$$\geq \frac{T-E}{N-E} \cdot H(A_{[1:N]\setminus\mathcal{E}}^{[k+1]}|W_{[1:k]}, A_{\mathcal{E}}^{[k+1]}, Q_{[1:N]}^{[k+1]}) \tag{90}$$

$$= \frac{T-E}{N-E} \cdot \Big(H(A_{[1:N]\setminus\mathcal{E}}^{[k+1]}|W_{[1:k+1]}, A_{\mathcal{E}}^{[k+1]}, Q_{[1:N]}^{[k+1]})$$
$$+ H(W_{k+1}|W_{[1:k]}, A_{\mathcal{E}}^{[k+1]}, Q_{[1:N]}^{[k+1]})$$
$$- H(W_{k+1}|W_{[1:k]}, A_{[1:N]}^{[k+1]}, Q_{[1:N]}^{[k+1]}) \Big) \tag{91}$$

$$\overset{(a)}{=} \frac{T-E}{N-E} \cdot \Big(H(A_{[1:N]\setminus\mathcal{E}}^{[k+1]}|W_{[1:k+1]}, A_{\mathcal{E}}^{[k+1]}, Q_{[1:N]}^{[k+1]})$$
$$+ L - o(L) \Big). \tag{92}$$

In (a), the second term follows from the security constraint (9) and the independence of the messages, and the third term follows from the correctness constraint (11). □

A. The Proof of $R \leq \left(1 - \frac{E}{N}\right)\left(1 + \frac{T-E}{N-E} + \cdots + \left(\frac{T-E}{N-E}\right)^{K-1}\right)^{-1}$

Applying Lemma 3 iteratively, we have

$$H(A_{[1:N]\setminus\mathcal{E}}^{[1]}|W_1, A_{\mathcal{E}}^{[1]}, Q_{[1:N]}^{[1]}) \tag{93}$$

$$\geq \frac{T-E}{N-E}\Big(H(A_{[1:N]\setminus\mathcal{E}}^{[2]}|W_1, W_2, A_{\mathcal{E}}^{[2]}, Q_{[1:N]}^{[2]})$$
$$+ L - o(L) \Big) \tag{94}$$

$$\geq (L - o(L))\left(\frac{T-E}{N-E} + \left(\frac{T-E}{N-E}\right)^2\right)$$
$$+ \left(\frac{T-E}{N-E}\right)^2 H(A_{[1:N]\setminus\mathcal{E}}^{[3]}|W_{[1:3]}, A_{\mathcal{E}}^{[3]}, Q_{[1:N]}^{[3]}) \tag{95}$$

$$\geq (L - o(L))\left(\frac{T-E}{N-E} + \cdots + \left(\frac{T-E}{N-E}\right)^{K-1}\right)$$
$$+ \left(\frac{T-E}{N-E}\right)^{K-1} H(A_{[1:N]\setminus\mathcal{E}}^{[K]}|W_{[1:K]}, A_{\mathcal{E}}^{[K]}, Q_{[1:N]}^{[K]}) \tag{96}$$

$$\geq (L - o(L))\left(\frac{T-E}{N-E} + \cdots + \left(\frac{T-E}{N-E}\right)^{K-1}\right). \tag{97}$$

Combining with Lemma 2, we have

$$\left(1 - \frac{E}{N}\right) H(A_{[1:N]}^{[1]}|Q_{[1:N]}^{[1]})$$

$$\geq (L - o(L))\left(1 + \frac{T-E}{N-E} + \cdots + \left(\frac{T-E}{N-E}\right)^{K-1}\right). \tag{98}$$

From the definition of the PIR rate and by letting $L \to \infty$,

$$R = \frac{H(W_1)}{\sum_{n=1}^N H(A_n^{[1]}|Q_n^{[1]})} \tag{99}$$

$$\leq \frac{L}{H(A_{[1:N]}^{[1]}|Q_{[1:N]}^{[1]})} \tag{100}$$

$$\leq \left(1 - \frac{E}{N}\right)\left(1 + \frac{T-E}{N-E} + \cdots + \left(\frac{T-E}{N-E}\right)^{K-1}\right)^{-1}. \tag{101}$$

B. The Proof of $\rho \geq \frac{E}{N-E}\left(1 + \frac{T-E}{N-E} + \cdots + \left(\frac{T-E}{N-E}\right)^{K-1}\right)$

For any set of nodes $\mathcal{E} \subset [1 : N]$ with size $|\mathcal{E}| = E$, from the security constraint (9),

$$0 = I(W_{[1:K]}; A_{\mathcal{E}}^{[k]}, Q_{\mathcal{E}}^{[k]}) \tag{102}$$

$$= H(A_{\mathcal{E}}^{[k]}|Q_{\mathcal{E}}^{[k]}) - H(A_{\mathcal{E}}^{[k]}|W_{[1:K]}, Q_{\mathcal{E}}^{[k]}) \tag{103}$$

$$\overset{(a)}{=} H(A_{\mathcal{E}}^{[k]}|Q_{\mathcal{E}}^{[k]}) - H(A_{\mathcal{E}}^{[k]}|W_{[1:K]}, Q_{\mathcal{E}}^{[k]})$$
$$+ H(A_{\mathcal{E}}^{[k]}|W_{[1:K]}, Q_{\mathcal{E}}^{[k]}, S) \tag{104}$$

$$= H(A_{\mathcal{E}}^{[k]}|Q_{\mathcal{E}}^{[k]}) - I(S; A_{\mathcal{E}}^{[k]}|W_{[1:K]}, Q_{\mathcal{E}}^{[k]}) \tag{105}$$

$$\geq H(A_{\mathcal{E}}^{[k]}|Q_{\mathcal{E}}^{[k]}) - H(S) \tag{106}$$

$$\overset{(b)}{=} H(A_{\mathcal{E}}^{[k]}|Q_{[1:N]}^{[k]}) - H(S). \tag{107}$$

(a) holds because the answers are deterministic functions of the queries, messages, and common randomness. By setting $\mathcal{K} = \emptyset$ in Lemma 1, we have (b).

Averaging over all $\mathcal{E}$,

$$H(S) \geq \frac{1}{\binom{N}{E}} \sum_{\substack{\mathcal{E}\in[1:N] \\ |\mathcal{E}|=E}} H(A_{\mathcal{E}}^{[k]}|Q_{[1:N]}^{[k]}) \tag{108}$$

$$\geq \frac{E}{N} \cdot H(A_{[1:N]}^{[k]}|Q_{[1:N]}^{[k]}). \tag{109}$$

From (98), we have

$$H(A_{[1:N]}^{[k]}|Q_{[1:N]}^{[k]})$$

$$\geq \frac{N}{N-E} \cdot \left(1 + \frac{T-E}{N-E} + \cdots + \left(\frac{T-E}{N-E}\right)^{K-1}\right)$$
$$\times (L - o(L)). \tag{110}$$

Hence, by letting $L \to \infty$, $\rho = \frac{H(S)}{H(W_k)} \geq \frac{E}{N-E}\left(1 + \frac{T-E}{N-E} + \cdots + \left(\frac{T-E}{N-E}\right)^{K-1}\right)$.

VII. CONVERSE WHEN $E \geq T$

A. The Proof of $R \leq 1 - \frac{E}{N}$

From the correctness constraint, for any message W_k, $H(W_k|A_{[1:N]}^{[k]}, Q_{[1:N]}^{[k]}) = o(L)$. For any set of nodes

$\mathcal{E} \subset [1 : N]$ with size $|\mathcal{E}| = E$,

$$H(W_k) - o(L)$$

$$= H(W_k|Q_{[1:N]}^{[k]}) - H(W_k|A_{[1:N]}^{[k]}, Q_{[1:N]}^{[k]}) \tag{111}$$

$$= H(A_{[1:N]}^{[k]}|Q_{[1:N]}^{[k]}) - H(A_{[1:N]}^{[k]}|W_k, Q_{[1:N]}^{[k]}) \tag{112}$$

$$\leq H(A_{[1:N]}^{[k]}|Q_{[1:N]}^{[k]}) - H(A_{\mathcal{E}}^{[k]}|W_k, Q_{[1:N]}^{[k]}) \tag{113}$$

$$\overset{(a)}{=} H(A_{[1:N]}^{[k]}|Q_{[1:N]}^{[k]}) - H(A_{\mathcal{E}}^{[k]}|W_k, Q_{\mathcal{E}}^{[k]}) \tag{114}$$

$$\overset{(b)}{=} H(A_{[1:N]}^{[k]}|Q_{[1:N]}^{[k]}) - H(A_{\mathcal{E}}^{[k]}|Q_{\mathcal{E}}^{[k]}) \tag{115}$$

$$\leq H(A_{[1:N]}^{[k]}|Q_{[1:N]}^{[k]}) - H(A_{\mathcal{E}}^{[k]}|Q_{[1:N]}^{[k]}) \tag{116}$$

$$\leq H(A_{[1:N]}^{[k]}|Q_{[1:N]}^{[k]}) - \frac{E}{N} H(A_{[1:N]}^{[k]}|Q_{[1:N]}^{[k]}), \tag{117}$$

where (a) follows from Lemma 1 and (b) follows from the security constraint (9). (117) follows by averaging over all $\mathcal{E}$ with size E and from Han's inequality.

Therefore, by letting $L \to \infty$, $R = \frac{H(W_k)}{\sum_{n=1}^{N} H(A_n^{[k]})} \leq \frac{H(W_k)}{H(A_{[1:N]}^{[k]}|\mathcal{Q})} \leq 1 - \frac{E}{N}$.

B. The Proof of $\rho \geq \frac{E}{N-E}$

Note that (109) holds when $E \geq T$ as well. Then we have $H(S) \geq \frac{E}{N} \cdot H(A_{[1:N]}^{[k]}|Q_{[1:N]}^{[k]})$. From (117), we have $H(A_{[1:N]}^{[k]}|Q_{[1:N]}^{[k]}) \geq \frac{N}{N-E}(H(W_k) - o(L))$. Hence, by letting $L \to \infty$, $\rho = \frac{H(S)}{H(W_k)} \geq \frac{E}{N-E}$.

VIII. CONCLUSIONS

We show that $C_{\text{ETPIR}} = \left(1 - \frac{E}{N}\right) \cdot \left(1 + \frac{T-E}{N-E} + \left(\frac{T-E}{N-E}\right)^2 + \cdots + \left(\frac{T-E}{N-E}\right)^{K-1}\right)^{-1}$ when $E < T$, and $\left(1 - \frac{E}{N}\right)$ when $E \geq T$. It is surprising that by adding eavesdroppers and allowing servers to collude in the PIR problem, there is a synergistic effect and the problem behaves differently in two regimes. In particular, when $E \geq T$, the PIR problem behaves like SPIR. The ETPIR problem includes many previous works as special cases (refer to Figure 1), and the present work capitalizes on all ideas used in these works. Further, new challenges arise that go beyond previously studied problems, and new ideas are proposed to resolve the corresponding challenges. In particular, in the achievability proof when $E < T$ (Section IV), we require a layered MDS structure that simultaneously handles 3 correlated constraints – security, privacy and correctness.

Regarding future work, it will be interesting to see whether including other elements would result in similar synergistic interactions, such as coded databases [15], [16], non-responsive servers [2], active adversaries [5], [11], [17], multi-round protocols [18], arbitrary message length [19], [20] and secure storage [21], [22] *etc*. Another promising direction is to find explicit precoding matrices when $E < T$ as we have only shown the existence in our schemes. In particular, it is interesting to see if Vandermonde matrices are sufficient. Also, our correctness conditions are over-constrained such that relaxations might lead to field size reduction.

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. El 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 Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.

[5] K. Banawan and S. Ulukus, "The capacity of private information retrieval from Byzantine and colluding databases," *IEEE Trans. Inf. Theory*, to be published, doi: 10.1109/TIT.2018.2869154.

[6] Q. Wang and M. Skoglund, "Secure private information retrieval from colluding databases with eavesdroppers," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 2456–2460.

[7] Q. Wang and M. Skoglund, "On PIR and symmetric PIR from colluding databases with adversaries and eavesdroppers," *IEEE Trans. Inf. Theory*, to be published, doi: 10.1109/TIT.2018.2878034.

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

[9] H. Sun and S. A. Jafar, "The capacity of symmetric private information retrieval," *IEEE Trans. Inf. Theory*, to be published, doi: 10.1109/TIT.2018.2848977.

[10] Q. Wang and M. Skoglund, "Linear symmetric private information retrieval for MDS coded distributed storage with colluding servers," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2017, pp. 71–75.

[11] Q. Wang and M. Skoglund, "Secure symmetric private information retrieval from colluding databases with adversaries," in *Proc. 55th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2017, pp. 1083–1090.

[12] J. T. Schwartz, "Fast probabilistic algorithms for verification of polynomial identities," *J. ACM*, vol. 27, no. 4, pp. 701–717, Oct. 1980.

[13] R. Zippel, "Probabilistic algorithms for sparse polynomials," in *Symbolic and Algebraic Computation*. Berlin, Germany: Springer, 1979, pp. 216–226.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.

[15] 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.

[16] Q. Wang and M. Skoglund, "Symmetric private information retrieval for MDS coded distributed storage," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[17] Q. Wang, H. Sun, and M. Skoglund. (2018). "The ϵ-error capacity of symmetric PIR with Byzantine adversaries." [Online]. Available: https://arxiv.org/abs/1809.03988

[18] 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.

[19] 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.

[20] C. Tian, H. Sun, and J. Chen. (2018). "Capacity-achieving private information retrieval codes with optimal message size and upload cost." [Online]. Available: https://arxiv.org/abs/1808.07536

[21] 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.

[22] Z. Jia, H. Sun, and S. A. Jafar. (2018). "Cross subspace alignment and the asymptotic capacity of X-secure T-private information retrieval." [Online]. Available: https://arxiv.org/abs/1808.07457

Qiwen Wang (M'18) received her B.Sc. degree in Mathematics and B.Eng. degree in Information Engineering in 2010, and Ph.D. degree in Information Engineering in 2015, all from the Chinese University of Hong Kong, Hong Kong. She is currently a postdoctoral researcher in the Department of Information Science and Engineering, KTH Royal Institute of Technology, Stockholm, Sweden.

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.

Mikael Skoglund (S'93–M'97–SM'04–F'19) received the Ph.D. degree in 1997 from Chalmers University of Technology, Sweden. In 1997, he joined the Royal Institute of Technology (KTH), Stockholm, Sweden, where he was appointed to the Chair in Communication Theory in 2003. At KTH, he heads the Department of Information Science and Engineering.

Dr. Skoglund has worked on problems in source-channel coding, coding and transmission for wireless communications, Shannon theory, information and control, and statistical signal processing. He has authored and co-authored some 150 journal and 350 conference papers.

Dr. Skoglund has served on numerous technical program committees for IEEE sponsored conferences. During 2003–08 he was an associate editor with the IEEE TRANSACTIONS ON COMMUNICATIONS and during 2008–12 he was on the editorial board for the IEEE TRANSACTIONS ON INFORMATION THEORY. He is general co-chair for IEEE ITW 2019.