

Multi-Server Private Linear Transformation with Joint Privacy

Fatemeh Kazemi and Alex Sprintson

Dept. of ECE, Texas A&M University, USA (E-mail: {fatemeh.kazemi, spalex}@tamu.edu)

Abstract—This paper focuses on the Private Linear Transformation (PLT) problem in the multi-server scenario. In this problem, there are N servers, each of which stores an identical copy of a database consisting of K independent messages, and there is a user who wishes to compute L independent linear combinations of a subset of D messages in the database while leaking no information to the servers about the identity of the entire set of these D messages required for the computation. We focus on the setting in which the coefficient matrix of the desired L linear combinations generates a Maximum Distance Separable (MDS) code. We characterize the capacity of the PLT problem, defined as the supremum of all achievable download rates, for all parameters $N, K, D \geq 1$ and $L = 1$, i.e., when the user wishes to compute one linear combination of D messages. We show that the capacity in this case is $\Phi(1/N, K - D + 1)$, where $\Phi(A, B) \triangleq (1 + A + A^2 + \dots + A^{B-1})^{-1}$. Moreover, we establish an upper bound and lower bounds on the capacity of the PLT problem for all parameters $N, K, D, L \geq 1$. We show that for the regime where $K - D$ is a non-negative integer multiple of L , the proposed capacity upper bound is equal to $\Phi(1/N, (K - D + L)/L)$. Also, leveraging some known capacity results, we show the tightness of this bound in the following regimes: (i) the case when there is a single server (i.e., $N = 1$), (ii) the case when $L = 1$, and (iii) the case when $L = D$.

I. INTRODUCTION

This work focuses on the Private Linear Transformation (PLT) problem which was recently introduced in [1], [2]. In the PLT problem, there are N servers, each of which stores an identical copy of a database consisting of K independent messages. Also, there is a user who wishes to compute L independent linear combinations of a subset of D messages in the database, without revealing any information to the servers about the identities of the D messages required for the computation, while downloading the minimum possible amount of information from the servers.

The PLT problem can be seen as an interesting extension of the Private Information Retrieval (PIR) (e.g., [3]–[21]) and Private Linear Computation (PLC) (e.g., [22]–[25]) problems, which have been extensively studied in the literature. To be more specific, for $L = D$, the PLT problem reduces to the multi-message PIR problem in which the goal is to privately retrieve a subset of D messages in the database. Moreover, for $L = 1$, the PLT problem reduces to the PLC problem in which the goal is to privately compute one linear combination of a D -subset of messages. The PLT problem can be motivated by several practical scenarios such as linear transformation technique applied for dimensionality reduction in Machine Learning (ML) applications (see [2]).

This material is based upon work supported by the National Science Foundation under Grants No. 1718658.

A. Previous and Related Work

In the classical PIR problem, a user wants to download a message from a database replicated over N non-colluding servers, without leaking any information about the identity of the desired message to any individual server. The capacity of the information-theoretic PIR was derived in [4]. Then, the PIR problem has been extended in various directions, such as coded PIR (e.g., [5]–[7]), multi-message PIR (e.g., [8]–[11]), and PIR with side information (e.g., [15]–[21]).

The PLC problem, which was initially introduced in [22], is an interesting generalization of the PIR problem, in which the user wishes to compute one arbitrary linear combination of the messages in the database, while revealing no information about the identities and the coefficients of the messages in the demanded linear combination to any server. Several variants of the PLC problem were also studied in [23]–[28]. In [25], a variation of the PLC problem was considered in which it is only required to protect the identities of the messages in the demanded linear combination, while the coefficients used to construct the linear combination do not need to be hidden from the server.

This work focuses on the PLT problem recently introduced in [1], [2], which is also closely related to the PIR and PLC problems. Indeed, one naive protocol for the PLT problem is to privately retrieve all the D messages required for computation using a multi-message PIR scheme, and then compute the demanded linear combinations. Another simple approach for the PLT problem is to compute each required linear combination separately using a PLC protocol.

Although there is a significant body of literature on PIR and PLC problems, there are only a few studies on PLT. In particular, the PLT problem was studied in the single-server setting considering the following two privacy requirements: (i) the individual privacy, where the identity of each individual message in the support set of the demanded linear combinations needs to be kept private [1]; and (ii) the joint privacy, in which the identity of the entire set of messages in the support set of the demanded linear combinations must be kept private [2]. All variants of the PIR and PLC problems, can be considered for the PLT problem which opens several interesting directions for future work.

In [2], the authors recently proved that the capacity of the PLT with a single server and joint privacy is $L/(K - D + L)$. However, the capacity of the PLT in the multi-server scenario was left as an open problem in [2]. Remarkably, neither a general achievability scheme nor a converse was known in this case. This work is motivated by this open problem.

B. Main Contributions

This paper studies the multi-server setting of PLT problem under the joint privacy guarantee. We consider the setting in which the coefficient matrix related to the required linear combinations generates MDS code. This setting is motivated by several practical scenarios, e.g., the coefficient matrix may be chosen randomly over the field of real numbers [2]. As the first contribution of this work, we characterize the capacity of the PLT for the case of $L = 1$ (i.e., PLC problem), thus settling the open problem mentioned above for the case of $L = 1$. We show that the capacity in this case is $\Phi(1/N, K - D + 1)$, where $\Phi(A, B) = (1 + A + A^2 + \dots + A^{B-1})^{-1}$. Moreover, we establish an upper bound on the capacity of PLT problem for all parameters $N, K, D, L \geq 1$, and based on some known capacity results, we show the tightness of our proposed upper bound in some special cases of the problem: (i) the single-server case (i.e., $N = 1$), (ii) the case of $L = 1$, and (iii) the case of $L = D$. An independent and concurrent work [29] which mainly focuses on the special case of $L = 1$ (i.e., PLC problem), characterizes the capacity of this case under the individual and joint privacy guarantees. Also, for a specific regime when L divides $K - D$, an upper bound on the capacity of PLT was presented in [29] with no detailed proof. In this paper, we present a detailed converse proof for the established upper bound on the capacity of PLT for all parameters $N, K, D, L \geq 1$. Our converse proofs are based on a reduction technique and leverage the capacity results for two settings of the PIR problem with side information. Our achievability proof for the PLC ($L = 1$) is inspired by our recently proposed scheme of [20] for single-server PIR with private coded side information, and the multi-server private computation scheme of [22]. We propose some trivial PLT protocols based on multi-message PIR and PLC protocols, which provide lower bounds (does not match our proposed upper bound) on the capacity of the multi-server PLT.

II. PROBLEM FORMULATION

A. Basic Notation

Throughout this paper, we denote random variables by bold letters and their realizations by regular letters. The functions $\mathbb{P}(\cdot)$, $\mathbb{P}(\cdot|\cdot)$, $\mathbb{H}(\cdot)$, $\mathbb{H}(\cdot|\cdot)$, and $I(\cdot;\cdot)$ denote probability, conditional probability, entropy, conditional entropy, and conditional mutual information, respectively. Let $\mathbb{Z}_{\geq 0}$ and $\mathbb{N}$ denote the set of non-negative integers and the set of positive integers, respectively. For any $i \in \mathbb{N}$, let $[i] \triangleq \{1, \dots, i\}$. Let $\mathbb{F}_q$ be a finite field for some prime q , $\mathbb{F}_q^\times \triangleq \mathbb{F}_q \setminus \{0\}$ be the multiplicative group of $\mathbb{F}_q$, and $\mathbb{F}_q^S$ be the S -dimensional vector space over $\mathbb{F}_q$ for some integer $S \geq 1$. Let $B \triangleq S \log_2 q$. Let $K, D, L \geq 1$ be integers such that $L \leq D \leq K$. Let $\mathcal{K} \triangleq [K]$. Let $\mathbb{W}$ denote the set of all D -subsets (i.e., subsets of size D) $\mathcal{W}$ of $\mathcal{K}$, and $\mathbb{V}$ denote the set of all MDS matrices V of dimension $L \times D$ with entries in $\mathbb{F}_q$ (i.e., every $L \times L$ submatrix of V is full-rank). We denote the cardinality of a set $\mathcal{S}$ by $|\mathcal{S}|$. For a positive real number A and a positive integer number B , let $\Phi(A, B) = (1 + A + A^2 + \dots + A^{B-1})^{-1}$.

B. Setup and Assumptions

Consider N non-colluding servers, each stores an identical copy of a database consisting of K messages, $X_{\mathcal{K}} = \{X_1, \dots, X_K\}$, where each message X_i is a row vector of length S . Let $X \triangleq [X_1^\top, \dots, X_K^\top]^\top$ be a matrix of dimension $K \times S$. For some $\mathcal{R} \triangleq \{i_1, \dots, i_r\} \subset \mathcal{K}$, let $X_{\mathcal{R}}$ be the submatrix of X of size $|\mathcal{R}| \times S$, restricted to its rows indexed by the set $\mathcal{R}$, i.e., $X_{\mathcal{R}} = [X_{i_1}^\top, \dots, X_{i_r}^\top]^\top$.

Suppose that there is a user who wishes to compute L linear combinations of D messages $\{X_i : i \in \mathcal{W}\}$, as $V_1 X_{\mathcal{W}}, \dots, V_L X_{\mathcal{W}}$, where $\mathcal{W} \in \mathbb{W}$ is the index set of the D messages required for the computation, and V_ℓ , $\ell \in [L]$, denoting the coefficient vector of the ℓ th desired linear combination, is the ℓ th row of an $L \times D$ MDS matrix V with entries in $\mathbb{F}_q$, i.e., $V = [V_1^\top, \dots, V_L^\top]^\top$, $V \in \mathbb{V}$. Indeed, the user wants to compute the $L \times S$ matrix $Z^{[\mathcal{W}, V]} \triangleq V X_{\mathcal{W}}$ whose rows are the L required linear combinations. We refer to $Z^{[\mathcal{W}, V]}$ as the *demand*, $\mathcal{W}$ as the *demand's index set*, V as the *demand's coefficient matrix*, L as the *demand's dimension*, and D as the *demand's support size*.

We assume $\mathbf{X}_1, \dots, \mathbf{X}_K$ are independently and uniformly distributed over $\mathbb{F}_q^S$, i.e., $H(\mathbf{X}_i) = B$ for $i \in \mathcal{K}$. Thus, $H(\mathbf{X}) = KB$, $H(\mathbf{X}_{\mathcal{R}}) = |\mathcal{R}|B$ for every $\mathcal{R} \subset \mathcal{K}$, and $H(\mathbf{Z}^{[\mathcal{W}, V]}) = LB$. We also assume that $\mathcal{W}$, V , and $\mathbf{X}$ are independent random variables such that $\mathcal{W}$ and V are uniformly distributed over $\mathbb{W}$ and $\mathbb{V}$, respectively. Moreover, we assume that the servers initially know the distributions of $\mathcal{W}$ and V , whereas the servers have no information about the realizations $\mathcal{W}$ and V in advance.

C. Privacy and Recoverability Conditions

To retrieve the demand $Z^{[\mathcal{W}, V]}$ for any given $\mathcal{W}$ and V , the user generates N queries $\{Q_n^{[\mathcal{W}, V]}\}_{n \in [N]}$, and sends $Q_n^{[\mathcal{W}, V]}$ (potentially stochastic function of $\mathcal{W}$ and V) to server n . Let $Q^{[\mathcal{W}, V]} \triangleq \{Q_n^{[\mathcal{W}, V]}\}_{n \in [N]}$ and $\mathbf{Q}^{[\mathcal{W}, V]} \triangleq \{Q_n^{[\mathcal{W}, V]}\}_{n \in [N]}$.

Once server n receives the query $Q_n^{[\mathcal{W}, V]}$, it responds with an answer $A_n^{[\mathcal{W}, V]}$ (deterministic function of $Q_n^{[\mathcal{W}, V]}$ and $\mathbf{X}$, i.e., $H(A_n^{[\mathcal{W}, V]} | Q_n^{[\mathcal{W}, V]}, \mathbf{X}) = 0$) to the user. For clarity, let $A^{[\mathcal{W}, V]} \triangleq \{A_n^{[\mathcal{W}, V]}\}_{n \in [N]}$ and $\mathbf{A}^{[\mathcal{W}, V]} \triangleq \{A_n^{[\mathcal{W}, V]}\}_{n \in [N]}$.

Recoverability Condition: The answers $A^{[\mathcal{W}, V]}$ from all the servers along with the queries $Q^{[\mathcal{W}, V]}$, and the realizations $\mathcal{W}, V$ must enable the user to retrieve the demand $Z^{[\mathcal{W}, V]}$. This condition is referred to as the *recoverability condition*, as formally stated in the following

$$H(\mathbf{Z}^{[\mathcal{W}, V]} | \mathbf{A}^{[\mathcal{W}, V]}, \mathbf{Q}^{[\mathcal{W}, V]}, \mathcal{W}, V) = 0,$$

Privacy Condition: The queries $Q^{[\mathcal{W}, V]}$ should be designed such that the servers infer no information about the user's demand index set $\mathcal{W}$. This condition is referred to as the *(joint) privacy condition*, formally stated as follows

$$I(\mathcal{W}; \mathbf{Q}_n^{[\mathcal{W}, V]}, \mathbf{A}_n^{[\mathcal{W}, V]}, \mathbf{X}_{\mathcal{K}}) = 0 \quad \forall n \in [N].$$

Equivalently, from the perspective of each server, every D -subset of indices $\mathcal{K}$ must be equally likely to be the demand's index set, i.e., for any given $\tilde{\mathcal{W}} \in \mathbb{W}$, it must hold that

$$\mathbb{P}(\mathcal{W} = \tilde{\mathcal{W}} | \mathbf{Q}_n^{[\mathcal{W}, V]} = Q_n^{[\mathcal{W}, V]}) = \mathbb{P}(\mathcal{W} = \tilde{\mathcal{W}}) \quad \forall n \in [N].$$

D. Problem Statement

The problem is to design a protocol for generating queries $Q^{[\mathcal{W}, \mathcal{V}]}$ and their corresponding answers $A^{[\mathcal{W}, \mathcal{V}]}$ (for any given $\mathcal{W}$ and $\mathcal{V}$) such that both the privacy and recoverability conditions are satisfied. We refer to this problem as *Private Linear Transformation (PLT)*. A protocol for generating queries/answers for PLT is referred to as a *PLT protocol*.

The *rate* of a PLT protocol is defined as the ratio of the entropy of the demand, i.e., $H(\mathbf{Z}^{[\mathcal{W}, \mathcal{V}]}) = LB$, to the total entropy of answers from the servers, i.e., $\sum_{n=1}^N H(\mathbf{A}_n^{[\mathcal{W}, \mathcal{V}]})$. The *capacity* of the PLT problem, denoted by $C^{PLT}(N, K, L, D)$, is defined as the supremum of rates over all PLT protocols, i.e.,

$$C^{PLT}(N, K, L, D) \triangleq \sup \frac{LB}{\sum_{n=1}^N H(\mathbf{A}_n^{[\mathcal{W}, \mathcal{V}]})}$$

In this work, our goal is to characterize (or derive non-trivial bounds on) the capacity of the PLT problem, i.e., $C^{PLT}(N, K, L, D)$, and to design a PLT protocol that is capacity-achieving.

III. MAIN RESULTS

In this section, we present our main results. Theorem 1 establishes an upper bound on the capacity of the PLT problem for all parameters $N, K, L, D \geq 1$. Leveraging some known capacity results, we show that the presented upper bound is tight in the following regimes: (i) the case where there is a single server (i.e., $N = 1$), (ii) the case where $L = 1$, and (iii) the case where $L = D$. Theorem 2 characterizes the capacity of the PLT problem for all parameters $N, K, D \geq 1$ and $L = 1$, i.e., the case where the user wishes to privately compute *one* linear combination of D messages in the database. The proofs of theorems 1 and 2 are given in sections IV and V, respectively.

Theorem 1. *The capacity of the PLT problem with N non-colluding and replicated servers, K messages, demand's support size D , and demand's dimension L , (i) if $\frac{K-D}{L} \leq 1$, is upper bounded by*

$$C^{PLT}(N, K, L, D) \leq \left(1 + \frac{K-D}{LN}\right)^{-1},$$

(ii) and if $\frac{K-D}{L} \geq 1$, is upper bounded by

$$C^{PLT}(N, K, L, D) \leq \left(\frac{1 - \left(\frac{1}{N}\right)^{\lfloor \theta \rfloor}}{1 - \frac{1}{N}} + \frac{(\theta - \lfloor \theta \rfloor)}{N^{\lfloor \theta \rfloor}} \right)^{-1}.$$

where $\theta \triangleq \frac{K-D+L}{L}$.

The converse proof is provided in Section IV-A, which is based on a reduction argument and leverages the capacity result for multi-message PIR with private side information problem, introduced in [14].

Corollary 1. *If $\frac{K-D}{L} \in \mathbb{Z}_{\geq 0}$, the capacity upper bounds provided in Theorem 1, can be written as*

$$C^{PLT}(N, K, L, D) \leq \Phi\left(\frac{1}{N}, \frac{K-D+L}{L}\right).$$

Remark 1. The capacity upper bounds in Theorem 1 are tight for the case when $N = 1$ (i.e., when there is a single server), which is equal to $L/(K-D+L)$ as was shown in [2, Theorem 2]. Moreover, in Theorem 2, we prove the tightness of this upper bound for the case of $L = 1$.

Remark 2. Notably, for the case of $L = D$, where the user wishes to privately compute D independent linear combinations of D -subset of messages in the database (which is equivalent to privately retrieving these D messages), the capacity upper bound in Theorem 1, i.e., (i) $(1 + (K-D)/DN)^{-1}$ if $K/D \leq 2$, and (ii) $\Phi(1/N, K/D)$ if $K/D \geq 2$ and $K/D \in \mathbb{N}$, is tight as was shown in [9]. Note that in this case, an optimal capacity-achieving multi-message PIR protocol proposed in [9, Theorems 1, 2] is an optimal protocol that achieves the capacity upper bound in Theorem 1.

Theorem 2. *The capacity of the PLT problem with N non-colluding and replicated servers, K messages, demand's support size D , and demand's dimension $L = 1$, is given by*

$$C^{PLT}(N, K, 1, D) = \Phi\left(\frac{1}{N}, K-D+1\right).$$

The converse proof follows directly from the result of Theorem 1 for $L = 1$. Also, an alternative proof of converse, similar to that of Theorem 1, is provided in Section V. For the achievability proof, we design a PLT protocol that achieves the proposed upper bound provided by converse, and is inspired by both our recently proposed scheme of [20] for the single-server PIR with private coded side information problem, and the scheme proposed in [22] for the private computation problem.

Remark 3. The result of Theorem 2 generalizes the previous finding reported in [2] for the PLT problem with a single server, without any prior side information, when joint privacy is required, and $L = 1$. As was shown in [2], the capacity of this setting is equal to $K-D+1$, which is consistent with the result of Theorem 2 for $N = 1$. Also, evidently it can be observed that for the case of $D = 1$, the result of Theorem 2 reduces to the known capacity result of [4] for the classical PIR problem where the user wants to privately download one message in the database, which is $\Phi(1/N, K)$.

Remark 4. It is worthwhile to compare the result of Theorem 2 with the capacity result of [22] for the related PLC problem where the user wishes to compute one arbitrary linear combination of K independent messages in a database replicated at N non-colluding servers, while hiding both the identities and the coefficients of the messages participating in the demand. As was shown in [22], the capacity of this setting is equal to $\Phi(1/N, K)$. Unlike the privacy requirements in the private computation problem introduced in [22], in the PLT problem, the goal is to hide only the identities of the D messages participating in the user's demand and not necessarily the values of their coefficients, which based on the result of Theorem 2, it can be fulfilled more efficiently

with much higher rate, i.e., $\Phi(1/N, K - D + 1)$. This is interesting since this type of access privacy are motivated by many practical scenarios such as linear transformation technique used for dimensionality reduction in Machine Learning (ML) applications (see, e.g. [2], [30] and references therein). By comparing the capacity results of these two problems, one can readily conclude that the advantage of PLT protocols over the repeated use of a PLC protocol becomes more tangible when the demand's support size D increases.

Remark 5. It is noteworthy that for¹ $D \geq 2$, a trivial PLT protocol for $L = 1$ would be privately retrieving the D messages required for the linear computation using an optimal multi-message PIR scheme satisfying privacy of demand messages jointly, introduced in [9], and then computing the required linear combination. As was shown in [9, Theorems 1, 2], the optimal rate that can be achieved leveraging this approach, is upper bounded by $D^{-1} \leq 1/2$. The result of Theorem 2 indicates that the PLT problem in general can be addressed much more efficiently with the rate of $\Phi(1/N, K - D + 1) \geq 1/2$.

Remark 6. Interestingly, in the PLT problem, a simple approach of computing each of the required linear combinations separately through applying an optimal PLT scheme introduced in Theorem 2, cannot achieve the capacity upper bound presented in Theorem 1 for all parameters N, K, L, D .

IV. PROOF OF THEOREM 1

A. Converse proof

The proof of converse follows from the capacity result for the problem of multi-message PIR with private side information, referred to as M-PIR-PSI, introduced in [14, Theorem 1]. In this problem, there is a database of K independent messages whose copies are replicated across N servers, and there is a user who has access to M messages from the database as side information. The user wishes to retrieve P messages from the database while leaking no information about the identities of both the desired messages and the side information messages, to any individual server. As was shown in [14, Theorem 1], the capacity of this setting, denoted by $C^{MPIR-PSI}(N, K, P, M)$,

(i) if $\frac{K-M}{P} \leq 2$ is given by

$$C^{MPIR-PSI}(N, K, P, M) = \left(1 + \frac{K - M - P}{PN}\right)^{-1}, \quad (1)$$

(ii) if $\frac{K-M}{P} \geq 2$ is upper bounded by

$$C^{MPIR-PSI}(N, K, P, M) \leq \left(\frac{1 - (\frac{1}{N})^{\lfloor \rho \rfloor}}{1 - \frac{1}{N}} + \frac{(\rho - \lfloor \rho \rfloor)}{N^{\lfloor \rho \rfloor}} \right)^{-1}, \quad (2)$$

where $\rho \triangleq \frac{K-M}{P}$. In case (ii), as was shown [14, Corollary 1], if $\frac{K-M}{P} \in \mathbb{N}$, the capacity is given by

$$C^{MPIR-PSI}(N, K, P, M) = \Phi\left(\frac{1}{N}, \frac{K - M}{P}\right). \quad (3)$$

¹Note that for the case of $D = 1$, the PLT problem reduces to the classical single-message PIR problem introduced in [4].

In the following, we want to show that any PLT protocol designed for the problem with N servers, K messages, demand's support size D , and demand's dimension L , can be used as a protocol that satisfies both the recoverability and the privacy conditions of the M-PIR-PSI problem with demand size $P = L$ and side information size $M = D - L$. Specifically, for a given instance of the M-PIR-PSI problem with the set of demand indices $\mathcal{P}$ of size L , (i.e., $P = L$), and the set of side information indices $\mathcal{S}$ of size $D - L$, (i.e., $M = D - L$), the user can construct a random $L \times D$ MDS matrix V and forms the set $\mathcal{W} = \mathcal{P} \cup \mathcal{S}$. Then, for the given $\mathcal{W}$ and V , the user and the servers can apply a PLT protocol for generating queries $Q^{[\mathcal{W}, V]}$ and their corresponding answers $A^{[\mathcal{W}, V]}$, such that the user can privately compute L MDS coded linear combinations of the D messages indexed by the set $\mathcal{W}$ (i.e., union of demands and side information messages). The user can then retrieve the L desired messages by subtracting off the contribution of the $D - L$ side information messages from the computed L linear combinations.

Now, we need to prove that the PLT-based protocol described above satisfies both the recoverability and the joint privacy conditions of the M-PIR-PSI problem. It should be noted that since the PLT protocol enables the user to compute L MDS coded linear combinations of D messages, based on the property of MDS codes², one can readily verify that the user can always retrieve the L desired messages by subtracting off the contribution of $D - L$ side information messages from the L computed linear equations, and solving the resulting system of L linear equations with L unknowns. Thus, the recoverability condition is satisfied.

It is easy to verify that by applying the PLT protocol, the identities of all the D messages (i.e., the union of the demand messages and side information messages) participating in the L linear combinations, will be jointly protected from each server as a result of the privacy guarantees of the PLT protocol. Indeed, from the perspective of each server, every D -subset of K messages is equally likely to be the union of the demand messages and side information messages. Moreover, due to the property of MDS codes, within each D -subset of messages, every subset of size L can be considered as the set of demand messages (i.e., the remaining $D - L$ as the set of side information messages) with equal probability. This ensures that the described PLT-based protocol satisfies the privacy condition in the M-PIR-PSI problem.

Thus, we conclude that any achievable rates in the PLT problem with N servers, K messages, demand's support size D , and demand's dimension L , would be also achievable (using the PLT-based protocol) in the M-PIR-PSI problem with N servers, K messages, demand size $P = L$, and side information size $M = D - L$. Thus, the capacity of PLT problem with parameters N, K, D, L , i.e., $C^{PLT}(N, K, L, D)$, is upper bounded by the capacity of the M-PIR-PSI problem with parameters $N, K, P = L, M = D - L$, i.e., $C^{MPIR-PSI}(N, K, L, D - L)$.

²Every $L \times L$ submatrix of an $L \times D$ MDS matrix is invertible.

Thus, the proof is completed by substituting P with L , and M with $D - L$ in equations 1, 2.

V. PROOF OF THEOREM 2

Here, we prove the converse by showing that the capacity for the case of $L = 1$, i.e., $C^{PLT}(N, K, 1, D)$, is upper bounded by the capacity of PIR with private side information problem, referred to as PIR-PSI, in which a database of K independent messages is replicated across N servers, and the user has access to M messages from the database as side information. The user wants to retrieve one message from the database while hiding jointly the identities of the desired message and the side information messages, from any individual server. This problem was introduced by Chen et al. [12]. As was shown in [12, Theorem 1], the capacity of PIR-PSI problem, denoted by $C^{PIR-PSI}(N, K, M)$, is equal to $\Phi(\frac{1}{N}, K - M)$.

Any PLT protocol designed for the problem with N servers, K messages, demand's support size D , and demand's dimension $L = 1$, enables the user to compute one linear combination of a subset of D messages while hiding the identities of these messages from any server. So, based on a similar reasoning used in the converse proof of Theorem 1, one can easily confirm that such PLT protocol would also be a protocol satisfying the recoverability and the privacy conditions in the PIR-PSI problem with side information size $M = D - 1$. Thus, any achievable rate in the PLT problem with N servers, K messages, demand's support size D , and demand's dimension $L = 1$, can be also achieved for the PIR-PSI problem with N servers, K messages, and side information size $M = D - 1$. Thus, we have $C^{PLT}(N, K, 1, D) \leq C^{PIR-PSI}(N, K, D - 1) = \Phi(\frac{1}{N}, K - D + 1)$.

A. Achievability proof

In this section, we complete the proof of Theorem 2 by designing a PLT protocol for the setting with N servers, K messages, demand's support size D , and demand's dimension $L = 1$, such that it achieves the upper bound provided by converse on the rate of any such PLT protocols, i.e., $\Phi(1/N, K - D + 1)$. The proposed protocol, referred to as the *Modified GRS Code*, leverages ideas from a modified version of the Specialized GRS Code Protocol proposed for the problem of single-server PIR with private coded side information in [20], as well as the PLC scheme proposed for the PLC problem in [22].

Modified GRS Code protocol: Assume $q \geq K$, and let each message consists of $S = N \binom{K}{D}$ symbols from $\mathbb{F}_q$. Suppose the user wishes to privately compute one linear combination of D messages indexed by a set $\mathcal{W}$, as $V_1 X_{\mathcal{W}} = \sum_{i \in \mathcal{W}} v_i X_i$ where V_1 is a row vector of length D . This protocol consists of four steps as follows:

Step 1: By using the Modified Specialized GRS Code protocol proposed in [20], the user first constructs a polynomial $p(x) = \sum_{i=0}^{K-D} p_i x^i \triangleq \prod_{i \notin \mathcal{W}} (x - \omega_i)$ where $\omega_1, \dots, \omega_K$ are K distinct arbitrarily chosen elements from $\mathbb{F}_q$. The user then constructs $r \triangleq K - D + 1$ vectors $Q_1, \dots, Q_r$, each

of length K , such that $Q_i = [\alpha_1 \omega_1^{i-1}, \dots, \alpha_K \omega_K^{i-1}]$, $i \in [r]$, where $\alpha_j = \frac{v_j}{p(\omega_j)}$ for any $j \in \mathcal{W}$, and α_j is chosen randomly from $\mathbb{F}_q^\times$ for any $j \notin \mathcal{W}$.

Step 2: Let $\hat{X}_i \triangleq \sum_{j=1}^K \alpha_j \omega_j^{i-1} X_j$ for $i \in [r]$. We refer to $\hat{X}_i$ as a *super-message*. Note that the vector Q_i , constructed in Step 1, is the vector of coefficients of the messages $\{X_i\}_{i \in \mathcal{K}}$ in the super-message $\hat{X}_i$. Let $F \triangleq \binom{K}{D}$, and let $W_1, W_2, \dots, W_F$ be the collection of all D -subsets of $\mathcal{K}$ in a lexicographical order. The structure of the Specialized GRS Code protocol [20] ensures that for each W_f , $f \in [F]$, there exist exactly $q - 1$ linear combinations $Y_f^1, Y_f^2, \dots, Y_f^{q-1}$ of the messages $\{X_i\}_{i \in W_f}$ with (non-zero) coefficients from $\mathbb{F}_q^\times$, such that for every $k \in [q - 1]$, Y_f^k can be written as a linear combination of the super-messages $\hat{X}_1, \dots, \hat{X}_r$. Let $\beta_f^k \triangleq [\beta_{f,1}^k, \dots, \beta_{f,r}^k]$ be a vector of length r such that $Y_f^k = \sum_{i=1}^r \beta_{f,i}^k \hat{X}_i$. It should be noted that, for each $f \in [F]$, $Y_f^1, Y_f^2, \dots, Y_f^{q-1}$ are the same up to a scalar multiple, i.e., for each $k \in [q - 1]$, $Y_f^k = \delta_k Y_f^1$, or equivalently, $\beta_f^k = \delta_k \beta_f^1$, for some distinct $\delta_k \in \mathbb{F}_q^\times$. The user then constructs F vectors $\beta_1, \dots, \beta_F$, each of length r , such that $\beta_f = \beta_f^{k_f}$ for $f \in [F]$, is chosen arbitrarily from the set of vectors $\{\beta_f^k\}_{k \in [q-1]}$. Let $Y_f \triangleq Y_f^{k_f}$ for $f \in [F]$. Each Y_f is referred to as a (linear) *function*. Note that β_f is the vector of coefficients of the super-messages $\{\hat{X}_i\}_{i \in [r]}$ in the function Y_f .

Step 3: The user then sends to all servers the vectors $Q_1, \dots, Q_r$, associated with the super-messages $\hat{X}_1, \dots, \hat{X}_r$, and the vectors $\beta_1, \dots, \beta_F$, associated with $Y_1, \dots, Y_F$.

Step 4: Then, the user and the servers leverage the PLC scheme of [22] with r (independent) messages and F (linear) functions of these messages such that the user can privately retrieve one of these functions. Indeed, the $r = K - D + 1$ super-messages $\{\hat{X}_i\}_{i \in [r]}$ and the F functions $\{Y_f\}_{f \in [F]}$, respectively, play the role of the original messages and the functions in the PLC scheme, and the user is interested in retrieving the function Y_{f^*} privately, where Y_{f^*} is a linear combination with non-zero coefficients of the messages $\{X_i\}_{i \in \mathcal{W}}$. Note that by construction, there exists only one function Y_{f^*} among $Y_1, \dots, Y_F$ such that Y_{f^*} is a linear combination (with only non-zero coefficients) of the messages $\{X_i\}_{i \in \mathcal{W}}$, and the user's demand is a scalar multiple of Y_{f^*} . More specifically, each server first constructs the super-messages $\{\hat{X}_i\}_{i \in [r]}$ by using the coefficient vectors $\{Q_i\}_{i \in [r]}$ as described in Step 2, and then constructs the functions $\{Y_f\}_{f \in [F]}$ by utilizing the super-messages $\{\hat{X}_i\}_{i \in [r]}$ and the coefficient vectors $\{\beta_f\}_{f \in [F]}$ as explained in Step 2. Note that each function Y_f for $f \in [F]$ consists of $S = N^F$ symbols (from $\mathbb{F}_q$) where N is the number of servers. Then, each server sends to the user $S(1/N + 1/N^2 + \dots + 1/N^{K-D+1})$ carefully designed linear combinations of all symbols associated with all functions $\{Y_f\}_{f \in [F]}$. The details of the design of the user's query to each server and each server's transmitted linear combinations (which also depend on the query of the user) can be found in [22, Section 4].

An example of the Modified GRS Code protocol is provided in [31].

Lemma 1. *The Modified GRS Code protocol is a PLT protocol, and achieves the rate $(\frac{1}{N}, K - D + 1)$.*

Proof. Since the messages $\mathbf{X}_{[K]}$ are uniformly and independently distributed over $\mathbb{F}_q^S$, and $\{\hat{X}_1, \dots, \hat{X}_r\}$ are linearly independent combinations of the messages in $X_{[K]}$, thus $\{\hat{\mathbf{X}}_1, \dots, \hat{\mathbf{X}}_r\}$ are uniformly and independently distributed over $\mathbb{F}_q^S$ as well, i.e., $H(\hat{\mathbf{X}}_1) = \dots = H(\hat{\mathbf{X}}_r) = S \log q = B$. Hence, the rate of the Modified GRS Code protocol is the same as the rate of the PLC protocol for N servers and $K - D + 1$ messages, which is given by $\Phi(\frac{1}{N}, K - D + 1)$ (see [22, Theorem 1]).

From the step 4 of the Modified GRS Code protocol, it is evident that the recoverability condition is satisfied. For the joint privacy of the proposed protocol, the proof is as follows. The PLC protocol protects the privacy of the function requested by the user (i.e., no server can infer any information about the index of the function requested by the user upon receiving the query). Consider an arbitrary server $n \in [N]$, which receives an arbitrary query $Q_n^{[\mathbf{W}, \mathbf{V}]}$, generated by the proposed protocol. Given $\mathbf{Q}_n^{[\mathbf{W}, \mathbf{V}]} = Q_n^{[\mathbf{W}, \mathbf{V}]}$, from the perspective of server n , every function Y_f for $f \in [F]$, is equally likely to be the user's desired function. We denote the support of Y_f by $\mathcal{Y}_f$, i.e., $\mathcal{Y}_f$ is the set of all indices $i \in [K]$ such that X_i has a non-zero coefficient in the linear combination Y_f . Note that for any $\tilde{\mathbf{W}} \in \mathbb{W}$, in the proposed protocol, there exists only one function Y_{f^*} among $Y_1, \dots, Y_F$ with $\mathcal{Y}_{f^*} = \tilde{\mathbf{W}}$. Thus, for any $\tilde{\mathbf{W}} \in \mathbb{W}$ and every $n \in [N]$, the following holds

$$\begin{aligned} \mathbb{P}(\mathbf{W} = \tilde{\mathbf{W}} | \mathbf{Q}_n^{[\mathbf{W}, \mathbf{V}]} = Q_n^{[\mathbf{W}, \mathbf{V}]}) \\ = \Pr(\mathbf{W} = \mathcal{Y}_{f^*} | \mathbf{Q}_n^{[\mathbf{W}, \mathbf{V}]} = Q_n) = \frac{1}{\binom{K}{D}} = \mathbb{P}(\mathbf{W} = \tilde{\mathbf{W}}). \end{aligned}$$

This completes the proof. $\square$

REFERENCES

- [1] A. Heidarzadeh, N. Esmati, and A. Sprintson, "Single-server private linear transformation: The individual privacy case," *arXiv preprint arXiv:2106.05222*, 2021.
- [2] —, "Single-server private linear transformation: The joint privacy case," *arXiv preprint arXiv:2106.05220*, 2021.
- [3] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," in *IEEE Symposium on Foundations of Computer Science*, 1995.
- [4] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4075–4088, July 2017.
- [5] N. B. Shah, K. Rashmi, and K. Ramchandran, "One extra bit of download ensures perfectly private information retrieval," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 856–860.
- [6] R. Tajeddine, O. W. Gnilke, and S. El Rouayheb, "Private information retrieval from mds coded data in distributed storage systems," *IEEE Trans. on Information Theory*, vol. 64, no. 11, pp. 7081–7093, 2018.
- [7] K. Banawan and S. Ulukus, "The capacity of private information retrieval from coded databases," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1945–1956, 2018.
- [8] —, "Multi-message private information retrieval," in *IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 1898–1902.
- [9] K. Banawan and S. Ulukus, "Multi-message private information retrieval: Capacity results and near-optimal schemes," *IEEE Transactions on Information Theory*, vol. 64, no. 10, pp. 6842–6862, Oct 2018.
- [10] A. Heidarzadeh, S. Kadhe, S. E. Rouayheb, and A. Sprintson, "Single-server multi-message individually-private information retrieval with side information," in *IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 1042–1046.
- [11] S. Li and M. Gastpar, "Single-server multi-message private information retrieval with side information," in *2018 56th Annual Allerton Conf. on Commun., Control, and Computing*, Oct 2018.
- [12] Z. Chen, Z. Wang, and S. A. Jafar, "The capacity of t-private information retrieval with private side information," *IEEE Transactions on Information Theory*, vol. 66, no. 8, pp. 4761–4773, 2020.
- [13] C. Tian, H. Sun, and J. Chen, "Capacity-achieving private information retrieval codes with optimal message size and upload cost," *IEEE Trans. on Information Theory*, vol. 65, no. 11, pp. 7613–7627, 2019.
- [14] M. J. Siavoshani, S. P. Shariatpanahi, and M. A. Maddah-Ali, "Private information retrieval for a multi-message scenario with private side information," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3235–3244, 2021.
- [15] S. Kadhe, B. Garcia, A. Heidarzadeh, S. Rouayheb, and A. Sprintson, "Private information retrieval with side information," *IEEE Trans. on Information Theory*, vol. 66, no. 4, pp. 2032–2043, 2019.
- [16] A. Heidarzadeh, F. Kazemi, and A. Sprintson, "The role of coded side information in single-server private information retrieval," *IEEE Transactions on Information Theory*, vol. 67, no. 1, pp. 25–44, 2020.
- [17] F. Kazemi, E. Karimi, A. Heidarzadeh, and A. Sprintson, "Single-server single-message online private information retrieval with side information," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 350–354.
- [18] A. Heidarzadeh, F. Kazemi, and A. Sprintson, "Capacity of single-server single-message private information retrieval with coded side information," in *IEEE Information Theory Workshop (ITW)*, Nov 2018.
- [19] F. Kazemi, E. Karimi, A. Heidarzadeh, and A. Sprintson, "Multi-server private information retrieval with coded side information," in *2019 IEEE 16th Canadian Workshop on Information Theory (CWIT)*, 2019.
- [20] A. Heidarzadeh, F. Kazemi, and A. Sprintson, "Capacity of single-server single-message private information retrieval with private coded side information," in *IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 1662–1666.
- [21] F. Kazemi, E. Karimi, A. Heidarzadeh, and A. Sprintson, "Private information retrieval with private coded side information: The multi-server case," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2019, pp. 1098–1104.
- [22] H. Sun and S. A. Jafar, "The capacity of private computation," *IEEE Trans. on Information Theory*, vol. 65, no. 6, pp. 3880–3897, 2018.
- [23] M. Mirmohseni and M. A. Maddah-Ali, "Private function retrieval," in *2018 Iran Workshop on Communication and Information Theory (IWCIT)*. IEEE, 2018, pp. 1–6.
- [24] A. Heidarzadeh and A. Sprintson, "Private computation with side information: The single-server case," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 1657–1661.
- [25] A. Heidarzadeh and A. Sprintson, "Private computation with individual and joint privacy," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 1112–1117.
- [26] S. A. Obead and J. Kliewer, "Achievable rate of private function retrieval from mds coded databases," in *IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 2117–2121.
- [27] B. Tahmasebi and M. A. Maddah-Ali, "Private sequential function computation," in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019, pp. 1667–1671.
- [28] S. A. Obead, H.-Y. Lin, E. Rosnes, and J. Kliewer, "Private function computation for noncolluding coded databases," *arXiv:2003.10007*, 2020. [Online]. Available: arXiv:2003.10007
- [29] N. Esmati and A. Heidarzadeh, "Multi-server private linear computation with joint and individual privacy guarantees," *arXiv:2108.09271*, Aug 2021.
- [30] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: applications to image and text data," in *Proc. of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 2001, pp. 245–250.
- [31] F. Kazemi and A. Sprintson, "Multi-server single-message private linear transformation with joint privacy," *arXiv:2108.09843*, Aug 2021.