Private Computation with Individual and Joint Privacy

Anoosheh Heidarzadeh and Alex Sprintson

Abstract— **This paper considers the problem of single-server Private Computation (PC) in the presence of Side Information (SI). In this problem, there is a server that stores K i.i.d. messages, and a user who has a subset of M uncoded messages or a coded linear combination of them as side information, where the identities of these messages are unknown to the server. The user wants to privately compute a linear combination of a subset of D other messages by downloading information from the server, where the identities of these messages must be kept private individually or jointly. For each setting, we define the capacity as the supremum of all achievable download rates.**

We characterize the capacity of both PC with coded and uncoded SI when individual privacy is required, for all K, M, D. Our results indicate that both settings have the same capacity. In addition, we establish a non-trivial lower bound on the capacity of PC with coded SI when joint privacy is required, for a range of parameters K, M, D. This lower bound is the same as the lower bound we previously established on the capacity of PC with uncoded SI when joint privacy is required.

I. INTRODUCTION

In this work, we consider the problem of Private Computation (PC) in the presence of side information. In this problem, there is a single (or multiple) remote server(s) storing (identical copies of) a database of i.i.d. messages; and there is a user who initially has a *side information* about some subset of messages in the database, where the identities of the messages in the support of the side information are initially unknown to the server. The user wishes to privately compute a linear combination of a different subset of database messages by downloading the minimum possible amount of information from the server(s).

We consider two different types of side information: (i) *uncoded side information* - where the user knows a subset of database messages, and (ii) *coded side information* - where the user holds a linear combination of a subset of database messages. These settings are referred to as *PC with Side Information (PC-SI)* and *PC with Coded Side Information (PC-CSI)*, respectively. We also consider two different privacy conditions: (i) *individual privacy* - where the identity of each message in the support set of the demanded linear combination needs to be kept private individually, and (ii) *joint privacy* - where the identities of all messages in the support set of the demanded linear combination must be kept private jointly. Note that neither individual nor joint privacy requires the coefficients of messages in the demanded linear combination to be kept private. When the condition (i) or (ii) is required, we refer to the PC problem as *Individually-Private Computation (IPC)* or *Jointly-Private Computation (JPC)*, respectively. In both problems, the goal is to design a protocol for generating the user's query and the server(s)' answer(s) such that the entropy of the answer(s) is minimized, while the query satisfies the underlying privacy condition.

Both IPC and JPC settings are related to the problem of Private Computation, introduced in [1], where the goal is to compute a linear combination of the messages in the database, while hiding both the identities and the coefficients of these messages. Several variants of this problem were also studied in [2]–[5]. These works consider neither individual nor joint privacy, nor any type of side information.

The JPC-SI setting, initially introduced in [6], is closely related to the problem of Private Information Retrieval with Side Information (PIR-SI), which was initially introduced in [7], [8] and later extended in several works, e.g., [9]–[12]. In the PIR-SI problem, a user wishes to retrieve a subset of database messages with the help of an uncoded side information, while achieving joint privacy. Several variants of PIR with different types of side information and privacy conditions were also studied in [13]–[20]. The IPC-SI setting is an extension of the PIR-SI problem when individual privacy is required. This problem, known as IPIR-SI, was introduced in [21]. The JPC-CSI and IPC-CSI settings are two generalizations of PIR with Coded Side Information (PIR-CSI), previously studied in [22] and [23].

A. Main Contributions

In this work, we focus on the single-server case. For each type of side information (coded or uncoded) and each privacy condition (individual or joint), the *capacity* of the setting being considered is defined as the supremum of all achievable download rates, where the *download rate* is the ratio of the entropy of a message to the entropy of the server's answer.

We characterize the capacity of both the IPC-SI and IPC-CSI settings, for all parameters. These results subsume several existing results in the PIR literature. The converse proof is information-theoretic, and the achievability scheme is a non-trivial generalization of our recently proposed scheme in [24] for the PIR-CSI setting. Our results show that the capacity of both settings are the same. This implies that, when individual privacy is required, having only *one* linear combination of a subset of messages as side information is as efficient as having them all separately. In addition, we establish a non-trivial lower bound on the capacity of the JPC-CSI setting for a range of parameters. Interestingly, this lower bound is the same as the lower bound we previously established in [6] on the capacity of the JPC-SI setting.

The authors are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (E-mail: {anoosheh, spalex}@tamu.edu).

This material is based upon work supported by the National Science Foundation (NSF) under Grants No. 1718658 and 1642983.

978-1-7281-6432-8/20/$31.00 ©2020 IEEE

The proof of achievability is based on a modification of the scheme we proposed in [6] for the JPC-SI setting.

Our results for both IPC and JPC settings, when compared to the existing results in the PIR literature, indicate that one can privately compute a linear combination of multiple messages much more efficiently than privately retrieving multiple messages, and linearly combining them locally. In addition, comparing our results with those in [1], one can see that hiding only the identities of the messages (either individually or jointly) and not their coefficients —which may still provide a satisfactory level of privacy in many applications, can be done with much less cost, even if there is only one server and/or the user has no side information.

II. Problem Formulation

Throughout, random variables and their realizations are denoted by bold-face letters and regular letters, respectively.

Let $\mathbb{F}_q$ be a finite field of size q, and let $\mathbb{F}_{q^\ell}$ be an extension field of $\mathbb{F}_q$ for a positive integer ℓ. Let K, M, D be non-negative integers such that $K \geq M + D$. We denote $\{1, \dots, K\}$ by $\mathcal{K}$, and let $\mathcal{K}_M$ (or $\mathcal{K}_D$) be the set of all M-subsets (or D-subsets) of $\mathcal{K}$. We also denote by $\mathcal{C}$ the set of all nonzero elements in $\mathbb{F}_q$, and let $\mathcal{C}_M$ (or $\mathcal{C}_D$) be the set of all ordered multisets of $\mathcal{C}$ of size M (or D).

Consider a single server that stores a dataset of K messages, $X_{\mathcal{K}} \triangleq \{X_1, \dots, X_K\}$, where each message X_i is independently and uniformly distributed over $\mathbb{F}_{q^\ell}$. That is, $H(\mathbf{X}_i) = L$ for $i \in \mathcal{K}$, and $H(\mathbf{X}_{\mathcal{K}}) = KL$, where $\mathbf{X}_{\mathcal{K}} \triangleq \{\mathbf{X}_1, \dots, \mathbf{X}_K\}$, and $L \triangleq \ell \log_2 q$. Consider a user that knows a linear combination $Y^{[S,U]} \triangleq \sum_{i \in S} u_i X_i$ of M messages $X_S \triangleq \{X_i\}_{i \in S}$ for some $S \in \mathcal{K}_M$ and some $U \triangleq \{u_i\}_{i \in S} \in \mathcal{C}_M$, and wishes to retrieve a linear combination $Z^{[W,V]} \triangleq \sum_{i \in W} v_i X_i$ from the server for some $W \in \mathcal{K}_D$, $W \cap S = \emptyset$, and some $V \triangleq \{v_i\}_{i \in W} \in \mathcal{C}_D$. We refer to $Y^{[S,U]}$ as the *side information*, X_S as the *side information support set*, S as the *side information support index set*, M as the *side information support size*, $Z^{[W,V]}$ as the *demand*, X_W as the *demand support set*, W as the *demand support index set*, and D as the *demand support size*.

We assume that $\mathbf{S}$, $\mathbf{U}$, $\mathbf{V}$ are distributed uniformly over $\mathcal{K}_M$, $\mathcal{C}_M$, $\mathcal{C}_D$, respectively, and $\mathbf{W}$, given $\mathbf{S} = S$, is uniformly distributed over all $W \in \mathcal{K}_D$, $W \cap S = \emptyset$. Also, we assume that the server initially knows M, D, and the joint distribution of $(\mathbf{W}, \mathbf{V}, \mathbf{S}, \mathbf{U})$, whereas the realization (W, V, S, U) is not initially known to the server.

For any given (W, V, S, U), the user sends to the server a query $Q^{[W,V,S,U]}$, which is a (potentially stochastic) function of (W, V, S, U), in order to retrieve $Z^{[W,V]}$. Note that in this work we focus on queries that are "universal" in the sense that they do not depend on the content of any message, and hence are applicable for all realizations of messages. For simplifying the notation, we denote $\mathbf{Q}^{[\mathbf{W},\mathbf{V},\mathbf{S},\mathbf{U}]}$ by $\mathbf{Q}$. The query must satisfy one of the following privacy conditions:

(i) *Individual Privacy:* every message in $X_{\mathcal{K}}$ must be equally likely to be in the user's demand support set, i.e., for all $i \in \mathcal{K}$, it must hold that

$$\Pr(i \in \mathbf{W} | \mathbf{Q} = Q^{[W,V,S,U]}) = \Pr(i \in \mathbf{W}).$$

(ii) *Joint Privacy:* every D-subset of messages in $X_{\mathcal{K}}$ must be equally likely to be the user's demand support set, i.e., for all $W^* \in \mathcal{K}_D$, it must hold that

$$\Pr(\mathbf{W} = W^* | \mathbf{Q} = Q^{[W,V,S,U]}) = \Pr(\mathbf{W} = W^*).$$

Note that joint privacy implies individual privacy, but not vice versa. The main difference between these two privacy conditions is that for joint privacy which is a stronger notion of privacy, the query must protect the correlation between the indices in the demand support index set, whereas for individual privacy some information about this correlation may be leaked, and hence a weaker notion of privacy.

Neither individual nor joint privacy requires the privacy of the coefficients in the demand to be protected. This is in contrast to the privacy condition being considered in [1], and as a result of this relaxation one can expect more efficient private computation schemes in our settings. In particular, for single-server private computation without any side information, the user must download the entire dataset in order to protect the privacy of both the identities of the messages in the demand support set and their coefficients in the demand [1]. However, for neither of the two privacy conditions being considered here the entire dataset needs to be downloaded, even when the user has no side information.

Upon receiving $Q^{[W,V,S,U]}$, the server sends to the user an answer $A^{[W,V,S,U]}$, which is a (deterministic) function of the query $Q^{[W,V,S,U]}$ and the messages in $X_{\mathcal{K}}$. We denote $\mathbf{A}^{[\mathbf{W},\mathbf{V},\mathbf{S},\mathbf{U}]}$ by $\mathbf{A}$. Note that $H(\mathbf{A}|\mathbf{Q}, \mathbf{X}_{\mathcal{K}}) = 0$. The collection of $A^{[W,V,S,U]}$, $Q^{[W,V,S,U]}$, $Y^{[S,U]}$, and (W, V, S, U) must enable the user to retrieve the demand $Z^{[W,V]}$, i.e.,

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

We refer to this condition as the *recoverability condition*.

For each type of privacy, the problem is to design a protocol for generating a query $Q^{[W,V,S,U]}$ (and the corresponding answer $A^{[W,V,S,U]}$, given $Q^{[W,V,S,U]}$ and $X_{\mathcal{K}}$) for any given (W, V, S, U), such that both the privacy and recoverability conditions are satisfied. We refer to this problem as *single-server Individually-Private Computation with Coded Side Information (IPC-CSI)* or *Jointly-Private Computation with Coded Side Information (JPC-CSI)*, when individual or joint privacy is required, respectively. We similarly define the *IPC-SI* and *JPC-SI* problems for the settings in which the user's side information is the support set X_S itself, instead of a linear combination of the messages in X_S.

A protocol that generates query/answer for the IPC-CSI or JPC-CSI setting is referred to as an *IPC-CSI* or a *JPC-CSI protocol*, respectively. The *rate* of an IPC-CSI or a JPC-CSI protocol is defined as the ratio of the entropy of a message, i.e., L, to the entropy of the answer $\mathbf{A}$. The *capacity* of the IPC-CSI or JPC-CSI setting is defined as the supremum of rates over all IPC-CSI or JPC-CSI protocols, respectively. An IPC-SI or a JPC-SI protocol, its rate, and the capacity of the IPC-SI or JPC-SI setting are defined similarly.

Our goal in this work is to establish lower and/or upper bounds on the capacity of IPC-CSI, JPC-CSI, IPC-SI, and JPC-SI settings, in terms of the parameters K, M, D.

III. Necessary Conditions

The following two lemmas provide a necessary condition for individual and joint privacy, for both types of side information. The proofs are straightforward by the way of contradiction, and hence omitted for brevity.

Lemma 1 (Individual Privacy). *For any $i \in \mathcal{K}$, there exist $W^* \in \mathcal{K}_D$, $V^* \in \mathcal{C}_D$, and $S^* \in \mathcal{K}_M$ where $i \in W^*$ and $S^* \cap W^* = \emptyset$, such that*

$$H(\mathbf{Z}^{[W^*,V^*]}|\mathbf{A},\mathbf{Q},\mathbf{X}_{S^*}) = 0.$$

Lemma 2 (Joint Privacy). *For any $W^* \in \mathcal{K}_D$, there exist $V^* \in \mathcal{C}_D$ and $S^* \in \mathcal{K}_M$ where $S^* \cap W^* = \emptyset$, such that*

$$H(\mathbf{Z}^{[W^*,V^*]}|\mathbf{A},\mathbf{Q},\mathbf{X}_{S^*}) = 0.$$

Thinking of scalar-linear IPC or JPC protocols —where the answer consists only of scalar-linear combinations of the messages in $X_{\mathcal{K}}$, the necessary conditions in Lemmas 1 and 2 imply the need for linear codes that satisfy certain combinatorial requirements. (Recently, in [12], we made a similar connection between single-server PIR with side information and locally recoverable codes.) In particular, for constructing a scalar-linear IPC-CSI (or IPC-SI) protocol one requires a (linear) code of length K that satisfies the following requirement: for any $i \in \mathcal{K}$, there is a codeword of (Hamming) weight D or $M+D$ (or at least D and at most $M+D$) whose support includes the index i. Minimizing the entropy of the answer to maximize the rate of the protocol translates into minimizing the dimension of the code. In this work, we design optimal codes with minimum dimension for all K, M, D for the IPC-CSI setting. These codes naturally serve also as optimal codes for the IPC-SI setting.

The problem of designing a scalar-linear JPC-CSI (or JPC-SI) protocol reduces to the problem of designing a code of length K with minimum dimension satisfying the following requirement: for any D-subset $W \subseteq \mathcal{K}$, there is a codeword of weight D or $M+D$ (or at least D and at most $M+D$) whose support includes the D-subset W. The design of optimal codes satisfying this requirement remains an open problem. In [6], we initiated the study of the JPC-SI setting, and established a non-trivial upper bound on the dimension of optimal codes for this setting. In this work, we make the first attempt towards characterizing the dimension of optimal codes for the JPC-CSI setting; and provide a non-trivial upper bound for a range of parameters K, M, D.

IV. Main Results

Our main results for the IPC and JPC settings with both coded and uncoded side information are summarized in Sections IV-A and IV-B, respectively.

A. IPC-SI and IPC-CSI

The capacity of IPC-SI and IPC-CSI for arbitrary K, M, D are characterized in Theorems 1 and 2, respectively.

Theorem 1. *For the IPC-SI setting with K messages, side information of size M, and demand support size D, the capacity is given by $\lceil \frac{K}{M+D} \rceil^{-1}$.*

Theorem 2. *For the IPC-CSI setting with K messages, side information support size M, and demand support size D, the capacity is given by $\lceil \frac{K}{M+D} \rceil^{-1}$.*

For the converse proof, we use information-theoretic arguments relying primarily on the result of Lemma 1, to upper bound the rate of any IPC-SI protocol (see Section V-A). This upper bound obviously holds for any IPC-CSI protocol. For the proof of achievability, we construct a new scalar-linear IPC-CSI protocol, termed the *Generalized Modified Partition-and-Code (GMPC) protocol*, which achieves the rate upper bound (see Section V-B). This protocol naturally serves also as an IPC-SI protocol. The GMPC protocol is based on the idea of non-uniform randomized partitioning, and generalizes our recently proposed protocol in [24] for the PIR-CSI setting. Examples of the GMPC protocol can be found in a longer version of this work, [25].

Remark 1. The matching capacity of the IPC-SI and IPC-CSI settings shows that achieving individual privacy comes at no loss in capacity if the user has only *one* random linear combination of M random messages, instead of M random messages separately as their side information.

Remark 2. As shown in [21], for the IPIR-SI setting, the normalized download cost of $K - M\lfloor \frac{K}{M+D} \rfloor$ or $D\lceil \frac{K}{M+D} \rceil$ (depending on K, M, D) is achievable, where the *normalized download cost* is defined as the download cost normalized by the entropy of a message. Comparing this with the result of Theorem 1, one can see that, when individual privacy is required, one can privately compute a linear combination of multiple messages much more efficiently than retrieving them privately and linearly combining them locally.

Remark 3. For the case of $M = 0$, the capacity of both IPC-SI and IPC-CSI settings is equal to $\lceil \frac{K}{D} \rceil^{-1}$. Depending on the value of D, the capacity can be substantially larger than $\frac{1}{K}$, which was shown to be the capacity of single-server private computation where the privacy of both the demand support index set and the coefficients in the demand must be preserved [1]. For the case of $D = 1$, both the IPC-SI and IPC-CSI problems reduce to the problems of PIR-SI [7] and PIR-CSI where the demanded message does not lie in the support of the side information [22], respectively. The capacity of these settings were shown to be equal to $\lceil \frac{K}{M+1} \rceil^{-1}$, matching the results of Theorems 1 and 2.

B. JPC-SI and JPC-CSI

Theorem 3 lower bounds the capacity of JPC-SI for all K, M, D, and Theorem 4 establishes a lower bound on the capacity of JPC-CSI for some values of K, M, D.

Theorem 3 ([6]). *For the JPC-SI setting with K messages, side information of size M, and demand support size D, the capacity is lower bounded by $(\lceil \frac{K-M-D}{\lfloor M/D \rfloor + 1} \rceil + 1)^{-1}$.*

Theorem 4. *For the JPC-CSI setting with K messages, side information support size M, and demand support size D, the capacity is lower bounded by $(\frac{K-M-D}{\lfloor M/D \rfloor + 1} + 1)^{-1}$ when $\lfloor \frac{M}{D} \rfloor + 1$ divides $K - M - D$.*

The capacity lower bound in Theorem 3 is achievable by a scalar-linear JPC-SI protocol, called *Partition-and-Code with Interference Alignment (PC-IA)*, which we recently proposed in [6]. The PC-IA protocol is applicable for all K, M, D, and relies on the idea of a probabilistic partitioning that allows the parts to overlap and have multiple blocks of interference that are aligned (for details, see [6]).

Theorem 4, which appears without proof, follows from an observation that the PC-IA protocol (with a slight modification in the choice of coefficients in the linear combinations that constitute the server's answer to the user's query) serves also as a scalar-linear JPC-CSI protocol for some values of K, M, D, particularly when the divisibility condition in the theorem's statement holds. Notwithstanding, the PC-IA protocol is not a JPC-CSI protocol in general, and the construction of JPC-CSI protocols for arbitrary K, M, D is a challenging open problem, and the focus of an ongoing work. Examples of the PC-IA protocol can be found in [25].

Remark 4. As was shown in [6], when joint privacy is required, with the help of an uncoded side information the download cost for the private computation of one linear combination of multiple messages can be much lower than that of privately retrieving multiple messages and computing the linear combination of them. For instance, for K even, when the user has $M = 2$ messages as side information, for privately computing a linear combination of $D = 2$ messages the normalized download cost is equal to $\frac{K}{2} - 1$ (see Theorem 3); whereas, privately retrieving $D = 2$ messages incurs a normalized download cost of $\min\{K-2, K-\lfloor\frac{K}{3}\rfloor\}$, which is significantly higher than $\frac{K}{2}-1$ (see [9, Theorem 2]). Surprisingly, the result of Theorem 4 shows that for some values of K, M, D (e.g., K even and $M = D = 2$), only *one* linear combination of M messages suffices to achieve the same normalized download cost (e.g., $\frac{K}{2} - 1$). This is interesting because regardless of the values of M and D, when joint privacy is required, with the help of only one linear combination of M messages the normalized download cost for retrieving D messages is equal to $K - 1$, which is much higher than, for instance, $\frac{K}{2} - 1$.

Remark 5. The capacity lower bounds in Theorems 3 and 4 are tight for the cases of $D = 1$ and $M = 0$ (see [7], [22]). We have been able to prove the tightness of these bounds for small values of K, M, D, particularly for $M = D = 2$ and several values of K. Nevertheless, it remains open whether these lower bounds are tight for all K, M, D in general.

Remark 6. The matching capacity lower bounds in Theorems 3 and 4 raises an intriguing question whether, similar to the IPC-SI and IPC-CSI settings, the capacity of the JPC-SI and JPC-CSI settings are the same. We conjecture that the answer is affirmative for both linear and non-linear protocols.

V. Proofs of Theorems 1 and 2

Since any IPC-CSI protocol is an IPC-SI protocol, for the converse we only need to upper bound the rate of any IPC-SI protocol, whereas for the achievability it suffices to design an IPC-CSI protocol that achieves the rate upper bound.

A. Converse

Lemma 3. *The rate of any IPC-SI protocol for K messages, side information of size M, and demand support size D, is upper bounded by $\lceil\frac{K}{M+D}\rceil^{-1}$.*

Proof: We need to show that $H(\mathbf{A}) \geq \lceil\frac{K}{M+D}\rceil L$. Fix arbitrary $W \in \mathcal{K}_D$, $V \in \mathcal{C}_D$, $S \in \mathcal{K}_M$ such that $S \cap W = \emptyset$. Let $\mathbf{Z} \triangleq \mathbf{Z}^{[W,V]}$. By the recoverability condition, we have $H(\mathbf{Z}|\mathbf{A},\mathbf{Q},\mathbf{X}_S) = 0$. By a simple application of the chain rule of entropy, we have

$$\begin{aligned} H(\mathbf{A}) &\geq H(\mathbf{A}|\mathbf{Q},\mathbf{X}_S) + H(\mathbf{Z}|\mathbf{A},\mathbf{Q},\mathbf{X}_S)\\ &= H(\mathbf{Z}|\mathbf{Q},\mathbf{X}_S) + H(\mathbf{A}|\mathbf{Q},\mathbf{X}_S,\mathbf{Z})\\ &= H(\mathbf{Z}) + H(\mathbf{A}|\mathbf{Q},\mathbf{X}_S,\mathbf{Z}), \end{aligned} \tag{1}$$

noting that $H(\mathbf{Z}|\mathbf{Q},\mathbf{X}_S) = H(\mathbf{Z})$ because $\mathbf{Z}$ is only a linear combination of messages $\mathbf{X}_W$, and hence independent of $\mathbf{X}_S$ since $W \cap S = \emptyset$, and $\mathbf{Q}$ is independent of all messages $\mathbf{X}_{\mathcal{K}}$.

We consider two cases: (i) $W \cup S = \mathcal{K}$, and (ii) $W \cup S \neq \mathcal{K}$. In the case (i), $M + D = K$, and $\lceil\frac{K}{M+D}\rceil L = L$. Hence, (1) yields $H(\mathbf{A}) \geq H(\mathbf{Z}) = L$, as was to be shown.

In the case (ii), we further lower bound $H(\mathbf{A}|\mathbf{Q},\mathbf{X}_S,\mathbf{Z})$ as follows. Choose an arbitrary message, say $\mathbf{X}_{i_1}$, for some $i_1 \notin W \cup S$. By the result of Lemma 1, there exist $W_1 \in \mathcal{K}_D$, $i_1 \in W_1$, $V_1 \in \mathcal{C}_D$, and $S_1 \in \mathcal{K}_M$, $S_1 \cap W_1 = \emptyset$, such that $H(\mathbf{Z}_1|\mathbf{A},\mathbf{Q},\mathbf{X}_{S_1}) = 0$, or in turn, $H(\mathbf{Z}_1|\mathbf{A},\mathbf{Q},\mathbf{X}_S,\mathbf{Z},\mathbf{X}_{S_1}) = 0$, where $\mathbf{Z}_1 \triangleq \mathbf{Z}^{[W_1,V_1]}$. Thus,

$$\begin{aligned} H(\mathbf{A}|\mathbf{Q},\mathbf{X}_S,\mathbf{Z}) &\geq H(\mathbf{A}|\mathbf{Q},\mathbf{X}_S,\mathbf{Z},\mathbf{X}_{S_1})\\ &\quad + H(\mathbf{Z}_1|\mathbf{A},\mathbf{Q},\mathbf{X}_S,\mathbf{Z},\mathbf{X}_{S_1})\\ &= H(\mathbf{Z}_1|\mathbf{Q},\mathbf{X}_S,\mathbf{Z},\mathbf{X}_{S_1})\\ &\quad + H(\mathbf{A}|\mathbf{Q},\mathbf{X}_S,\mathbf{Z},\mathbf{X}_{S_1},\mathbf{Z}_1)\\ &= H(\mathbf{Z}_1) + H(\mathbf{A}|\mathbf{Q},\mathbf{X}_S,\mathbf{Z},\mathbf{X}_{S_1},\mathbf{Z}_1) \end{aligned} \tag{2}$$

where $\mathbf{Z}_1$ and $(\mathbf{Q},\mathbf{X}_S,\mathbf{Z},\mathbf{X}_{S_1})$ are independent because $i_1 \in W_1$ and $i_1 \notin W \cup S \cup S_1$. Let $n \triangleq \lceil\frac{K}{M+D}\rceil$. Using Lemma 1 recursively, it follows that for all $1 \leq k < n$ there exist $i_1,\dots,i_k \in \mathcal{K}$, $W_1,\dots,W_k \in \mathcal{K}_D$, $V_1,\dots,V_k \in \mathcal{C}_D$, and $S_1,\dots,S_k \in \mathcal{K}_M$ satisfying $i_l \in W_l$, $S_l \cap W_l = \emptyset$, $i_l \notin \cup_{j=1}^{l-1}(W_j \cup S_j) \cup (W \cup S)$ for all $1 \leq l \leq k$, such that

$$H(\mathbf{Z}_k|\mathbf{A},\mathbf{Q},\mathbf{X}_S,\mathbf{Z},\mathbf{X}_{S_1},\mathbf{Z}_1,\dots,\mathbf{X}_{S_{k-1}},\mathbf{Z}_{k-1},\mathbf{X}_{S_k}) = 0,$$

where $\mathbf{Z}_l \triangleq \mathbf{Z}^{[W_l,V_l]}$ for all $1 \leq l \leq k$. Obviously, $\left|\cup_{j=1}^{k-1}(W_j \cup S_j) \cup (W \cup S)\right| \leq (M+D)k$ for all $1 \leq k < n$. Applying the same technique as in (2), one can see that for all $1 \leq k < n$, we have

$$\begin{aligned} &H(\mathbf{A}|\mathbf{Q},\mathbf{X}_S,\mathbf{Z},\mathbf{X}_{S_1},\mathbf{Z}_1,\dots,\mathbf{X}_{S_{k-1}},\mathbf{Z}_{k-1})\\ &\quad \geq H(\mathbf{Z}_k) + H(\mathbf{A}|\mathbf{Q},\mathbf{X}_S,\mathbf{Z},\mathbf{X}_{S_1},\mathbf{Z}_1,\dots,\mathbf{X}_{S_k},\mathbf{Z}_k). \end{aligned}$$

Putting together these lower bounds for all k, we have

$$H(\mathbf{A}|\mathbf{Q},\mathbf{X}_S,\mathbf{Z}) \geq \sum_{k=1}^{n-1} H(\mathbf{Z}_k) = (n-1)L, \tag{3}$$

since $\mathbf{Z}_1,\dots,\mathbf{Z}_{n-1}$ are independent by the choice of $i_1,\dots,i_{n-1}$ in the construction. Combining (1) and (3), we get $H(\mathbf{A}) \geq nL = \lceil\frac{K}{M+D}\rceil L$, as was to be shown. □

B. Achievability

For the ease of notation, we define $n \triangleq \lceil \frac{K}{M+D} \rceil$, $m \triangleq n(M+D)-K$, and $r \triangleq M+D-m$.

Generalized Modified Partition-and-Code (GMPC) Protocol: This protocol consists of three steps as follows:

Step 1: For any $1 \leq l < n$, let $I_l \triangleq \{(l-1)(M+D)+1,\dots,l(M+D)\}$, and let $I_n \triangleq \{1,\dots,m,(n-1)(M+D)+1,\dots,K\}$. Note that $I_1 \cap I_n = \{1,\dots,m\}$.

First, the user constructs a random permutation π on $\mathcal{K}$ as follows. With probability $\alpha \triangleq \frac{m+2r}{K}$, the user chooses $l^* \in \{1,n\}$ uniformly at random; otherwise, with probability $1-\alpha$, the user randomly chooses $l^* \in \{2,\dots,n-1\}$.

If $l^* \in \{1,n\}$, with probability β (or $1-\beta$) where the choice of β will be specified shortly, the user assigns $\mu \triangleq \min\{D,m\}$ (or $D-\rho \triangleq D-\min\{D,r\}$) randomly chosen indices from W and $m-\mu$ (or $m-D+\rho$) randomly chosen indices from S to $\{\pi(j): 1 \leq j \leq m\}$ at random, and randomly assigns the rest of the indices in $W \cup S$ to $\{\pi(j): j \in I_{l^*} \setminus \{1,\dots,m\}\}$. Otherwise, if $l^* \in \{2,\dots,n-1\}$, the user randomly assigns the $M+D$ indices in $W \cup S$ to $\{\pi(j): j \in I_{l^*}\}$. Then, the user assigns the (not-yet-assigned) indices in $\mathcal{K} \setminus (W \cup S)$ to $\{\pi(j): j \notin I_{l^*}\}$.

The value of β, which depends on the values of K, M, D, is carefully chosen to satisfy the individual privacy condition:

$$\beta \triangleq \begin{cases} \frac{m}{m+2r}, & D \leq m, D \leq r, \\ \frac{D}{m+2r}, & D > m, D \leq r, \\ 1-\frac{2D}{m+2r}, & D \leq m, D > r, \\ \frac{r}{M}\left(1-\frac{2D}{m+2r}\right), & D > m, D > r. \end{cases}$$

Next, the user constructs n ordered sets $Q'_1,\dots,Q'_n$, each of size $M+D$, defined as $Q'_k \triangleq \{\pi(j): j \in I_l\}$; and constructs an ordered multiset Q'' of size $M+D$, defined as $Q'' \triangleq \{c_j : j \in I_{l^*}\}$ where $c_j = v_{\pi(j)}$ or $c_j = u_{\pi(j)}$ when $\pi(j) \in W$ or $\pi(j) \in S$, respectively. Recall that $v_{\pi(j)}$ or $u_{\pi(j)}$ is the coefficient of the message $X_{\pi(j)}$ in the user's demand or side information, respectively.

The user then constructs $Q_l = (Q'_l, Q'')$ for $1 \leq l \leq n$, and sends the query $Q^{[W,V,S,U]} = \{Q_1,\dots,Q_n\}$ to the server.

Step 2: By using $Q_l = (Q'_l, Q'')$'s, the server computes A_l's, defined as $A_l \triangleq \sum_{j=1}^{M+D} c_{i_j} X_{i_j}$ where $Q'_l = \{i_1,\dots,i_{M+D}\}$ and $Q'' = \{c_{i_1},\dots,c_{i_{M+D}}\}$, and sends the answer $A^{[W,V,S,U]} = \{A_1,\dots,A_n\}$ back to the user.

Step 3: Upon receiving the server's answer, the user retrieves the demand $Z^{[W,V]}$ by subtracting off the contribution of the side information $Y^{[S,U]}$ from $A_{l^*} = Z^{[W,V]} + Y^{[S,U]}$.

Lemma 4. *The GMPC protocol is a scalar-linear IPC-CSI protocol, and achieves the rate* $\lceil \frac{K}{M+D} \rceil^{-1}$.

Proof: The rate and the scalar-linearity of the GMPC protocol are obvious from the construction. Clearly, the recoverability condition is also satisfied.

To prove that the GMPC protocol satisfies the individual privacy condition, we need to show that for any given query Q generated by the protocol, for all $i \in \mathcal{K}$, it holds that

$$\Pr(i \in \mathbf{W} | \mathbf{Q} = Q) = \Pr(i \in \mathbf{W}) = \frac{D}{K},$$

noting that $\mathbf{W}$ is distributed uniformly over $\mathcal{K}_D$.

Fix an arbitrary $i \in \mathcal{K}$. We consider the following three different cases separately: (i) $\pi^{-1}(i) \in \{1,\dots,m\}$; (ii) $\pi^{-1}(i) \in I_l \setminus \{1,\dots,m\}$ for some $l \in \{1,n\}$; and (iii) $\pi^{-1}(i) \in I_l$ for some $l \notin \{1,n\}$, where $\pi^{-1}(i) = j$ if and only if $\pi(j) = i$.

First, consider the case (i). In this case, we have

$$\begin{aligned} &\Pr(i \in \mathbf{W} | \mathbf{Q} = Q) \\ &\quad = \sum_{l \in \{1,n\}} \Pr(i \in \mathbf{W}, l^* = l | \mathbf{Q} = Q) \\ &\quad = \sum_{l \in \{1,n\}} \Pr(l^* = l | \mathbf{Q} = Q) \times \Pr(i \in \mathbf{W} | \mathbf{Q} = Q, l^* = l) \\ &\quad = 2 \Bigg(\frac{1}{2} \times \alpha \Bigg(\beta \times \frac{\binom{m-1}{\mu-1}}{\binom{m}{\mu}} + (1-\beta) \times \frac{\binom{m-1}{D-\rho-1}}{\binom{m}{D-\rho}} \Bigg)\Bigg)\Bigg) \\ &\quad = \begin{cases} \alpha\beta\left(\frac{D}{m}\right), & D \leq m, D \leq r, \\ \alpha\beta, & D > m, D \leq r, \\ \alpha\left(\beta\left(\frac{D}{m}\right) + (1-\beta)\left(\frac{D-r}{m}\right)\right), & D \leq m, D > r, \\ \alpha\left(\beta + (1-\beta)\left(\frac{D-r}{m}\right)\right)\Big), & D > m, D > r, \end{cases} \\ &\quad = \frac{D}{K}, \end{aligned}$$

for our choice of β for each range of values of m and r.

Next, consider the case (ii). In this case, we have

$$\begin{aligned} &\Pr(i \in \mathbf{W} | \mathbf{Q} = Q) \\ &\quad = \Pr(i \in \mathbf{W}, l^* = l | \mathbf{Q} = Q) \\ &\quad = \Pr(l^* = l | \mathbf{Q} = Q) \times \Pr(i \in \mathbf{W} | \mathbf{Q} = Q, l^* = l) \\ &\quad = \frac{1}{2} \times \alpha \left(\beta \times \frac{\binom{r-1}{D-\mu-1}}{\binom{r}{D-\mu}} + (1-\beta) \times \frac{\binom{r-1}{\rho-1}}{\binom{r}{\rho}} \right) \\ &\quad = \begin{cases} \frac{\alpha}{2}(1-\beta)\left(\frac{D}{r}\right), & D \leq m, D \leq r, \\ \frac{\alpha}{2}\left(\beta\left(\frac{D-m}{r}\right) + (1-\beta)\left(\frac{D}{r}\right)\right), & D > m, D \leq r, \\ \frac{\alpha}{2}(1-\beta), & D \leq m, D > r, \\ \frac{\alpha}{2}\left(\beta\left(\frac{D-m}{r}\right) + (1-\beta)\right), & D > m, D > r, \end{cases} \\ &\quad = \frac{D}{K}, \end{aligned}$$

for our choices of β specified earlier.

Lastly, consider the case (iii). In this case, we have

$$\begin{aligned} &\Pr(i \in \mathbf{W} | \mathbf{Q} = Q) \\ &\quad = \Pr(i \in \mathbf{W}, l^* = l | \mathbf{Q} = Q) \\ &\quad = \Pr(l^* = l | \mathbf{Q} = Q) \Pr(i \in \mathbf{W} | \mathbf{Q} = Q, l^* = l) \\ &\quad = \frac{1}{n-2} \times (1-\alpha) \left(\frac{D}{M+D}\right) \\ &\quad = \left(\frac{M+D}{K-m-2r}\right)\left(\frac{K-m-2r}{K}\right)\left(\frac{D}{M+D}\right) \\ &\quad = \frac{D}{K}. \end{aligned}$$

This completes the proof. □

References

[1] H. Sun and S. A. Jafar, "The capacity of private computation," *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3880–3897, 2019.

[2] S. A. Obead and J. Kliewer, "Achievable rate of private function retrieval from MDS coded databases," *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 2117–2121, 2018.

[3] S. A. Obead, H.-Y. Lin, E. Rosnes, and J. Kliewer, "Capacity of private linear computation for coded databases," *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 813–820, 2018.

[4] M. Mirmohseni and M. A. Maddah-Ali, "Private function retrieval," in *2018 Iran Workshop on Communication and Information Theory (IWCIT)*, April 2018, pp. 1–6.

[5] Z. Chen, Z. Wang, and S. A. Jafar, "The asymptotic capacity of private search," *IEEE Transactions on Information Theory*, pp. 1–1, 2020.

[6] 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.

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

[8] S. Kadhe, B. Garcia, A. Heidarzadeh, S. El Rouayheb, and A. Sprintson, "Private information retrieval with side information," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2032–2043, 2020.

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

[10] 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.

[11] 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.

[12] S. Kadhe, A. Heidarzadeh, A. Sprintson, and O. O. Koyluoglu, "On an equivalence between single-server pir with side information and locally recoverable codes," in *2019 IEEE Information Theory Workshop (ITW)*, 2019, pp. 1–5.

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

[14] S. P. Shariatpanahi, M. J. Siavoshani, and M. A. Maddah-Ali, "Multi-message private information retrieval with private side information," in *2018 IEEE Information Theory Workshop (ITW)*, 2018, pp. 1–5.

[15] R. Tandon, "The capacity of cache aided private information retrieval," in *55th Annual Allerton Conf. on Commun., Control, and Computing*, Oct 2017, pp. 1078–1082.

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

[17] ——, "Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching," *IEEE Transactions on Information Theory*, pp. 1–1, 2018.

[18] Y. Wei and S. Ulukus, "The capacity of private information retrieval with private side information under storage constraints," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2023–2031, 2020.

[19] A. Heidarzadeh, F. Kazemi, and A. Sprintson, "Capacity of single-server single-message private information retrieval with private coded side information," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 1662–1666.

[20] 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)*, 2019, pp. 1098–1104.

[21] A. Heidarzadeh, S. Kadhe, S. E. Rouayheb, and A. Sprintson, "Single-server multi-message individually-private information retrieval with side information," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 1042–1046.

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

[23] F. Kazemi, E. Karimi, A. Heidarzadeh, and A. Sprintson, "Multi-server private information retrieval with coded side information," in *2019 16th Canadian Workshop on Information Theory (CWIT)*, 2019, pp. 1–6.

[24] A. Heidarzadeh, F. Kazemi, and A. Sprintson, "The role of coded side information in single-server private information retrieval," Oct 2019. [Online]. Available: arXiv:1910.07612

[25] A. Heidarzadeh and A. Sprintson, "Private computation with individual and joint privacy," Jan 2020. [Online]. Available: arXiv:2001.04545