Symmetric Private Information Retrieval at the Private Information Retrieval Rate

Zhusheng Wang, *Student Member, IEEE*, and Sennur Ulukus, *Fellow, IEEE*

***Abstract*—We consider the problem of *symmetric* private information retrieval (SPIR) with user-side common randomness. In SPIR, a user retrieves a message out of K messages from N non-colluding and replicated databases in such a way that no single database knows the retrieved message index (user privacy), and the user gets to know nothing further than the retrieved message (database privacy), i.e., the privacy constraint between the user and the databases is *symmetric*. SPIR has the following three properties: its capacity is smaller than the capacity of PIR which requires only user privacy; it is infeasible in the case of a single database; and it requires presence of shared common randomness among the databases. We introduce a new variant of SPIR where the user is provided with a random subset of the shared database common randomness, which is unknown to the databases. We determine the exact capacity region of the triple (d, ρ_S, ρ_U), where d is the download cost, ρ_S is the amount of shared database (server) common randomness, and ρ_U is the amount of available user-side common randomness. We show that with a suitable amount of ρ_U, this new SPIR achieves the capacity of the conventional PIR. As a corollary, single-database SPIR becomes feasible. Further, the presence of user-side ρ_U reduces the amount of required server-side ρ_S.**

***Index Terms*—Private information retrieval, symmetric private information retrieval, user-side common randomness.**

I. Introduction

PRIVATE information retrieval (PIR) is a fundamental problem, where a user downloads a message out of K possible messages stored in N non-colluding and replicated databases in such a way that no single database can know which message the user has downloaded [1]. This privacy requirement is referred to as *user privacy*. Symmetric PIR (SPIR) is an extended version of PIR, where additionally, the user learns nothing about the remaining messages stored in the databases while downloading its desired message [2]. This is referred to as *database privacy*. While PIR can be achieved with no shared common randomness, it is well-known that information-theoretic SPIR is possible only when the databases share a certain minimum amount of common randomness that is unknown to the user.

Manuscript received 31 December 2021; revised 20 April 2022 and 14 June 2022; accepted 22 June 2022. Date of publication 25 October 2022; date of current version 14 November 2022. This work was supported in part by Army Research Office (ARO) under Grant W911NF2010142, and in part by NSF under Grant CCF 17-13977 and Grant ECCS 18-07348. This article was presented in part at IEEE ISIT 2021 [DOI: 10.1109/ISIT45174.2021.9517907]. *(Corresponding author: Sennur Ulukus.)*

The authors are with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: zhusheng@umd.edu; ulukus@umd.edu).

Digital Object Identifier 10.1109/JSAIT.2022.3188610

The information-theoretic capacity of PIR and SPIR have been found in [3] and [4] as

$$C_{\text{PIR}} = \frac{1-\frac{1}{N}}{1-\left(\frac{1}{N}\right)^K} \quad \text{and} \quad C_{\text{SPIR}} = 1-\frac{1}{N} \tag{1}$$

First, C_{SPIR} is smaller than C_{PIR}, since SPIR is a more constrained problem than PIR, as it requires not only user privacy but also database privacy. Second, single-database SPIR is infeasible as $C_{\text{SPIR}} = 0$ for $N = 1$, while single-database PIR is feasible as $C_{\text{PIR}} = \frac{1}{K}$ for $N = 1$. Our goal in this paper is two-fold: To explore ways to increase SPIR capacity to the level of PIR capacity, and as importantly, to make single-database SPIR feasible.

Our motivation to focus on SPIR comes from constantly growing importance of privacy, not only the privacy of the retrieving user, but also the privacy of the databases, as the stored information in the databases may belong to other users. In addition, recent papers have shown that other important privacy primitives, such as private set intersection (PSI) [5], [6], can be recast as versions of the SPIR problem, e.g., multi-message SPIR. Thus, here, we investigate ways to increase the SPIR capacity. Further, in practical applications, enforcing non-collusion could be difficult, as in some cases, all databases may naturally belong to the same entity, e.g., in various multi-party secure computation problems [5]–[13]. If all databases collude or belong to the same entity, the system essentially becomes a single-database system [14], [15]. The single-database PIR problem has been studied under extended conditions, e.g., side-information [16]–[19]. Here, we investigate single-database SPIR under extended conditions, with the goal of making it feasible. Other important variants of PIR and SPIR problem have also been investigated, see, e.g., [20]–[58], especially relevant to us being those with side information (SI), with or without privacy of SI (PSI) [16]–[26].

In this paper, we introduce SPIR with user-side common randomness, which solves the above two issues.[1] In this model,

[1] In the PIR with (P)SI papers, the main objective of utilizing (P)SI is to further increase the capacity. In the SPIR problem, the databases share two sets of information systems: a message information system and a common randomness information system. As stated in [25, Th. 2], the capacity of SPIR-PSI is exactly equal to the capacity of SPIR without any PSI. This means that the message information system cannot be utilized as PSI to improve the capacity of SPIR. As an alternative, we turn to the common randomness information system, i.e., private side common randomness (called user-side common randomness in this paper), where the common randomness information system is utilized as PSI. Through this, the capacity of SPIR can possibly be further increased and the single-database SPIR can possibly be made feasible using user-side common randomness. We study this possibility in depth in this paper.

the user obtains a part of the common randomness shared by the databases. The databases know the size of the user-side common randomness, but they do not know what the user possesses exactly. One way to implement this is for the user to fetch a part of the common randomness from the databases uniformly randomly, i.e., without the user knowing what it will get and without the databases knowing what it got, except for its cardinality. That is, all subsets of a certain size are equally likely to be obtained by the user.[2] Another practical implementation could be for an external helper to distribute common randomness to the user and the databases randomly.

For database-side (server-side) common randomness of amount ρ_S and user-side common randomness of amount ρ_U, we determine the exact capacity region of the triple (d, ρ_S, ρ_U), where d is the download cost which is the inverse of the capacity. We show that with a suitable ρ_U, SPIR capacity becomes equal to the conventional PIR capacity. For the single-database case, since the conventional PIR capacity is $\frac{1}{K}$, this implies that single-database SPIR with user-side common randomness is feasible. In addition, the presence of user-side ρ_U reduces the amount of required server-side ρ_S.

II. Problem Formulation

We consider a system of $N \geq 1$ non-colluding databases each storing the same set of $K \geq 2$ i.i.d. messages each of which consisting of L i.i.d. symbols uniformly selected from a sufficiently large finite field $\mathbb{F}_q$, i.e.,

$$H(W_k) = L, \quad k \in [K] \tag{2}$$

$$H(W_{1:K}) = H(W_1) + \cdots + H(W_K) = KL \tag{3}$$

For convenience, we denote a random variable and its realization by using the same general uppercase letter when distinction is clear from the context. We address this issue additionally whenever clarification is needed. As in [4], we use a random variable $\mathcal{F}$ to denote the randomness in the retrieval strategy implemented by the user. Due to the user privacy constraint, the realization of $\mathcal{F}$ is only known to the user, and is unknown to any of the databases. Due to the database privacy constraint, databases need to share some amount of common randomness $\mathcal{R}_S$; we will call this *server-side* common randomness. The server-side common randomness $\mathcal{R}_S$ with size M is a set of i.i.d. symbols $\{S_1, S_2, \ldots, S_M\}$ uniformly selected from $\mathbb{F}_q$, i.e.,

$$H(S_m) = 1, \quad m \in [M] \tag{4}$$

$$H(S_{1:M}) = H(S_1) + \cdots + H(S_M) = M \tag{5}$$

Moreover, the set of indices $\{1, 2, \ldots, M\}$ forms an alphabet $\mathcal{A}$, i.e., $\mathcal{A} = \{1, 2, \ldots, M\}$. Before the retrieval process starts, the user obtains a partial knowledge of $\mathcal{R}_S$. We denote it by $\mathcal{R}_U$, and call it *user-side* common randomness. Therefore, we introduce a new random variable $\mathcal{A}_U$ corresponding to the uniform selection of elements without replacement from $\mathcal{A}$ (the sample space of $\mathcal{A}_U$ is the power set of $\mathcal{A}$). User-side common randomness $\mathcal{R}_U$ is a set of i.i.d. symbols from $\mathbb{F}_q$ where the indices of the symbols are constituted by $\mathcal{A}_U$. Further, we assume that $\mathcal{A}_U$ is not known to any individual database and also is kept private throughout the retrieval process,[3] although the cardinality of $\mathcal{A}_U$ can be public information to the databases. In addition, we introduce another new random variable $\bar{\mathcal{A}}_U$, which is the complement of $\mathcal{A}_U$ with respect to the universe $\mathcal{A}$, i.e., $\bar{\mathcal{A}}_U = \mathcal{A} \backslash \mathcal{A}_U$. Likewise, $\mathcal{R}_S \backslash \mathcal{R}_U$ is also a set of i.i.d. symbols from $\mathbb{F}_q$ where the indices of symbols are constituted by $\bar{\mathcal{A}}_U$. Thus, after determining the selection $\mathcal{A}_U$, $\bar{\mathcal{A}}_U$ is also deterministic; see Fig. 1 for the specific system model.

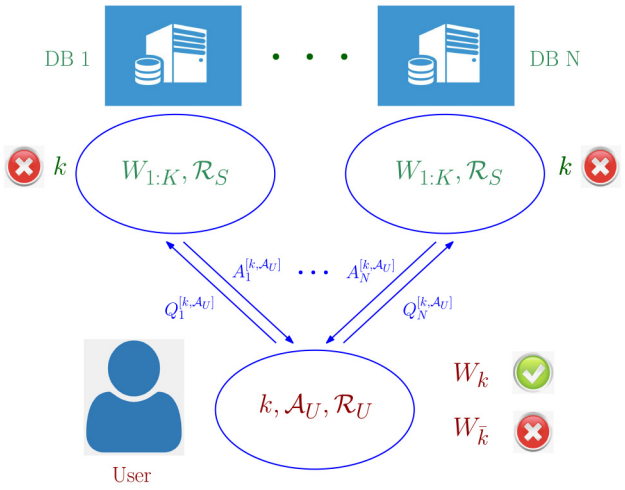

Fig. 1. System model for SPIR with user-side common randomness.

The server-side common randomness $\mathcal{R}_S$ is generated independently of the stored message set in the databases. The desired message index k, the random selection $\mathcal{A}_U$ and the retrieval strategy randomness $\mathcal{F}$, are all determined at the user-side before the retrieval process starts. Moreover, all these random variables are mutually independent, thus,

$$\begin{aligned} H(W_{1:K}, \mathcal{R}_S, k, \mathcal{A}_U, \mathcal{F}) &= H(W_{1:K}) + H(\mathcal{R}_S) + H(k) \\ &\quad + H(\mathcal{A}_U) + H(\mathcal{F}) \end{aligned} \tag{6}$$

Using the desired message index and the user-side common randomness indices, the user generates a query for each database according to the retrieval strategy randomness $\mathcal{F}$. Hence, the queries $Q_n^{[k,\mathcal{A}_U]}, n \in [N]$ are deterministic functions of $\mathcal{F}$,

$$H\left(Q_1^{[k,\mathcal{A}_U]}, Q_2^{[k,\mathcal{A}_U]}, \ldots, Q_N^{[k,\mathcal{A}_U]} | \mathcal{F}\right) = 0, \quad \forall k, \ \forall \mathcal{A}_U \tag{7}$$

During the independent query generation stage, (6) and (7) lead to the following relationship,

$$I\left(Q_{1:N}^{[k,\mathcal{A}_U]}; W_{1:K}, \mathcal{R}_S\right) = 0, \quad \forall k, \ \forall \mathcal{A}_U \tag{8}$$

[2]This random fetching problem can be viewed as randomized (multi-message) SPIR. The only difference of it from the conventional (multi-message) SPIR is that there is no input at the user side.

[3]We note that this assumption is with some loss of generality. There could be a version of the problem where we do not care about the privacy of $\mathcal{A}_U$ against the databases during the retrieval process. This version of the problem could potentially have a higher retieval rate. This choice is akin to enforcing "W-privacy" versus "W-S privacy" (see [16]–[26], especially [16], [25]), where W-privacy stands for message privacy only and W-S privacy stands for message and side-information privacy in a PIR setting with side information.

After receiving a query from the user, each database generates a truthful answer based on the stored message set $W_{1:K}$ and the server-side common randomness $\mathcal{R}_S$,

$$H\left(A_n^{[k,\mathcal{A}_U]}|Q_n^{[k,\mathcal{A}_U]}, W_{1:K}, \mathcal{R}_S\right) = 0, \quad \forall n, \ \forall k, \ \forall \mathcal{A}_U \quad (9)$$

After collecting all N answers from the databases, the user should be able to decode the desired message W_k reliably,

$$\begin{aligned}&\text{[reliability]} \quad H\left(W_k|Q_{1:N}^{[k,\mathcal{A}_U]}, A_{1:N}^{[k,\mathcal{A}_U]}, \mathcal{R}_U\right)\\ &\overset{(7)}{=} H\left(W_k|\mathcal{F}, A_{1:N}^{[k,\mathcal{A}_U]}, \mathcal{R}_U\right) = 0, \ \forall k, \forall \mathcal{A}_U \end{aligned} \quad (10)$$

Due to the user privacy constraint, the query generated to retrieve the desired message should be statistically indistinguishable from other queries. Specifically, for all k, k', all n, and all $\mathcal{A}_U$, there exists some $\mathcal{A}'_U$ with $|\mathcal{A}'_U| = |\mathcal{A}_U|$, i.e., $H(\mathcal{R}'_U) = H(\mathcal{R}_U)$,[4] such that,

$$\begin{aligned}&\text{[user privacy]} \quad \left(Q_n^{[k,\mathcal{A}_U]}, A_n^{[k,\mathcal{A}_U]}, W_{1:K}, \mathcal{R}_S\right)\\ &\sim \left(Q_n^{[k',\mathcal{A}'_U]}, A_n^{[k',\mathcal{A}'_U]}, W_{1:K}, \mathcal{R}_S\right)\end{aligned} \quad (11)$$

As in [41], [51], the joint probability distribution of all random variables at the databases can be factorized in the following way,

$$\begin{aligned}&P\left(\left(Q_n^{[k,\mathcal{A}_U]}, A_n^{[k,\mathcal{A}_U]}, W_{1:K}, \mathcal{R}_S\right) = (q, a, w_{1:K}, r_S)\right)\\ &= P\left(Q_n^{[k,\mathcal{A}_U]} = q\right)\\ &\quad \cdot P\left((W_{1:K}, \mathcal{R}_S) = (w_{1:K}, r_S)|Q_n^{[k,\mathcal{A}_U]} = q\right)\\ &\quad \cdot P\left(A_n^{[k,\mathcal{A}_U]} = a|\left(Q_n^{[k,\mathcal{A}_U]}, W_{1:K}, \mathcal{R}_S\right) = (q, w_{1:K}, r_S)\right) \quad (12)\\ &= P\left(Q_n^{[k,\mathcal{A}_U]} = q\right) \cdot P((W_{1:K}, \mathcal{R}_S) = (w_{1:K}, r_S)) \cdot c \quad (13)\end{aligned}$$

where the second term in (13) comes from the independent query generation of message set as well as server-side common randomness (8) and becomes a constant depending on the realizations of the pair $(W_{1:K}, \mathcal{R}_S)$, and the third term c in (13) is also a constant either taking the value of 0 or 1 depending on the choice of a because of the fact that the generated answer in a database is a deterministic function of the information that database possesses (9). As a consequence, we obtain the following equivalent expression for user privacy for all potential query realizations q,

$$\text{[user privacy]} \quad P\left(Q_n^{[k,\mathcal{A}_U]} = q\right) = P\left(Q_n^{[k',\mathcal{A}'_U]} = q\right) \quad (14)$$

Due to the database privacy constraint, the user should learn nothing about $W_{\bar{k}}$ which is the complement of W_k, i.e., $W_{\bar{k}} = \{W_1, \ldots, W_{k-1}, W_{k+1}, \ldots, W_K\}$,

$$\text{[database privacy]} \quad I\left(W_{\bar{k}}; \mathcal{F}, A_{1:N}^{[k,\mathcal{A}_U]}, \mathcal{R}_U\right) = 0, \ \forall k, \ \forall \mathcal{A}_U \quad (15)$$

Again due to the database privacy, the user should not learn all the information about the remaining common randomness among the databases when the retrieval is complete. However, in order to formulate the problem in an easier and clearer way, we add an additional requirement that the user should not gain any knowledge about the remaining common randomness among the databases even after retrieving the desired message,

$$I\left(\mathcal{R}_S \backslash \mathcal{R}_U; \mathcal{F}, A_{1:N}^{[k,\mathcal{A}_U]}, W_k, \mathcal{R}_U\right) = 0, \quad \forall k, \ \forall \mathcal{A}_U \quad (16)$$

An achievable SPIR scheme is a scheme that satisfies the reliability constraint (10), the user privacy constraint (14) and the database privacy constraint (15). The efficiency of a scheme is measured in terms of the number of downloaded bits by the user from all databases denoted by D. We define the normalized download cost d, the normalized server-side common randomness ρ_S, and the normalized user-side common randomness ρ_U, as

$$d = \frac{D}{L}, \qquad \rho_S = \frac{H(\mathcal{R}_S)}{L}, \qquad \rho_U = \frac{H(\mathcal{R}_U)}{L} \quad (17)$$

where L is the message length. Thus, the triple (d, ρ_S, ρ_U) is said to be achievable if all three values can be realized simultaneously by a valid achievable scheme. Our goal in this paper is to determine the capacity region over all achievable triples (d, ρ_S, ρ_U).

III. Main Result

We state the main result of our paper in the following theorem which is the *capacity region* for the triple (d, ρ_S, ρ_U).

Theorem 1: With user-side common randomness, the multi-database SPIR capacity region for $N \geq 2$ and $K \geq 2$ is

$$d \geq 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1}} \quad (18)$$

$$\rho_S - \rho_U \geq \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1}} \quad (19)$$

$$\frac{N-1}{N} d + \rho_U \geq 1 \quad (20)$$

$$\frac{N}{N-1} \rho_U + N\rho_S \geq \frac{N}{N-1} \quad (21)$$

Remark 1: The capacity region is defined in the form of a triple (d, ρ_S, ρ_U), where d is the reciprocal of the capacity defined in [3], [4], ρ_S is the required amount of common randomness shared among the databases relative to the message size, ρ_U is the total amount of common randomness obtained by the user before the retrieval starts relative to the message size. Theorem 1 gives the optimal tradeoff among these three variables and determines the exact capacity region. There are two corner points in this capacity region. The first corner point is $(\frac{N}{N-1}, \frac{1}{N-1}, 0)$, which is an intersection point among (19), (20) and the implicit constraint $\rho_U \geq 0$. The second corner point is $(1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}}, \frac{1}{N}+\frac{1}{N^2}+\cdots+\frac{1}{N^K}, \frac{1}{N^K})$, which is an intersection point among (17), (18) and (20). The achievability of the first corner point is provided in the existing paper [4]. The new achievability of the second corner point is introduced

[4]In the single-database case, $\mathcal{A}_U$ and $\mathcal{A}'_U$ can not be exactly the same although some overlap is allowed, nor can $\mathcal{R}_U$ and $\mathcal{R}'_U$. Otherwise, user-privacy, database-privacy and reliability constraints jointly form a contradiction, and as a consequence, the problem degenerates to the infeasible conventional single-database SPIR problem, which is trivial. However, this constraint on the strict difference between $\mathcal{A}_U$ and $\mathcal{A}'_U$ (also $\mathcal{R}_U$ and $\mathcal{R}'_U$) does not apply to the multi-database case. This is because its accompanying reliability constraint requires the user to collect the answers from all the databases, not only an individual one. Moreover, in the remaining content of this paper, we always assume that $\mathcal{A}'_U$ has the same cardinality as $\mathcal{A}_U$ and $\mathcal{R}'_U$ has the same entropy as $\mathcal{R}_U$.

in Section VI. Furthermore, any point on the line segment joining these two corner points can be achieved by time-sharing between these two different schemes. Any other remaining point in the capacity region can be achieved by adding extra common randomness at the user- and server-side simultaneously, or by increasing the server-side common randomness and the download cost.

Remark 2: The right hand side of (18) is the optimum normalized download cost of classical PIR, $d_{\text{PIR}} = 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1}}$ [3]. Thus, (18) states that $d \geq d_{\text{PIR}}$.

When $\rho_U = 0$, i.e., when there is no user-side common randomness, (20) becomes $d \geq d_{\text{SPIR}}$, where $d_{\text{SPIR}} = \frac{N}{N-1}$ is the optimum normalized download cost of classical SPIR [4], (21) gives $\rho_S \geq \frac{1}{N-1}$, and (18) and (19) are non-binding. Note that $d_{\text{SPIR}} > d_{\text{PIR}}$ for all N. Therefore, when $\rho_U = 0$, Theorem 1 reduces to the capacity of classical SPIR [4], and it corresponds to the first corner point.

When $\rho_U = \frac{1}{N^K}$, both (18) and (20) become equivalent to $d \geq d_{\text{PIR}}$, and the new SPIR download cost achieves $d = d_{\text{PIR}}$. In addition, from (19) and (21), we deduce that $\rho_S \geq \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1}} + \frac{1}{N^K} = \frac{1}{N} d_{\text{PIR}}$, which implies that the minimum amount of required server-side common randomness must be no smaller than the download cost in each database with symmetry across databases. Therefore, when $\rho_U = \frac{1}{N^K}$, the new SPIR download cost equals the download cost of the traditional PIR, and it corresponds to the second corner point.

Corollary 1: When $\rho_U = 0$, Theorem 1 reduces to the capacity of classical SPIR. That is, $d \geq \frac{N}{N-1} = d_{\text{SPIR}}$ and $\rho_S \geq \frac{1}{N-1} = \frac{1}{N} d_{\text{SPIR}}$.

Corollary 2: When $\rho_U = \frac{1}{N^K}$, new SPIR download cost equals the download cost of the traditional PIR, $d = d_{\text{PIR}}$. The required amount of server-side common randomness becomes $\rho_S \geq \frac{1}{N} d_{\text{PIR}}$.

Remark 3: The gap between ρ_S and ρ_U must be no smaller than a specific value as a function of N and K as given on the right hand side of (19). This comes from the database privacy constraint, where part of the common randomness, i.e., $\mathcal{R}_S \backslash \mathcal{R}_U$, is utilized to hide the undesired messages.

Remark 4: From (21), we observe that the existence of user-side common randomness can help reduce the required amount of server-side common randomness. In fact, from Corollary 1, when $\rho_U = 0$, we need $\rho_S \geq \frac{1}{N} d_{\text{SPIR}}$, whereas from Corollary 2, when $\rho_U = \frac{1}{N^K}$, we need $\rho_S \geq \frac{1}{N} d_{\text{PIR}}$. Noting that $d_{\text{PIR}} \leq d_{\text{SPIR}}$, the required server-side common randomness in Corollary 2 is smaller compared to Corollary 1. For instance, for $N = 2$ databases and $K = 2$ messages, classical SPIR optimum download cost $d = d_{\text{SPIR}} = 2$ is achieved by $\rho_S = 1$ [4]. In Theorem 1, $d = 2$ can be achieved by $\rho_S = \frac{3}{4}$ with $\rho_U = \frac{1}{4}$.

Remark 5: It is well-known that, for $N = 1$, classical SPIR is not feasible [4]. With user-side common randomness, single-database SPIR becomes feasible. The following corollary states the capacity region of this case as a reduction from Theorem 1.

Corollary 3: With user-side common randomness, the single-database SPIR capacity region for $N = 1$ and $K \geq 2$ is

$$d \geq K \tag{22}$$

$$\rho_S - \rho_U \geq K - 1 \tag{23}$$

$$\rho_U \geq 1 \tag{24}$$

Remark 6: The optimal normalized download cost for single-database PIR is $d = K$ [3], [16], which is achieved by downloading all messages from the database. One of the difficulties of single-database SPIR is that downloading all messages is not a valid SPIR scheme. Corollary 3 shows that single-database PIR capacity can be achieved for single-database SPIR by means of user-side common randomness.

Remark 7: The first two terms in Corollary 3 follow from the first two terms in Theorem 1. The third term in Corollary 3 follows from the last two terms in Theorem 1 by multiplying both sides of the fourth term in Theorem 1 by $N - 1$.

Remark 8: Like multi-database SPIR, in the single-database SPIR as well, the gap between ρ_S and ρ_U must be no smaller than a specific value as a function of K as given in (23) to avoid information leakage on undesired messages.

Remark 9: From Corollary 3, the minimum download cost for single-database SPIR with user-side common randomness is $d = K$, the minimum required server-side common randomness is $\rho_S = K$, of which $\rho_U = 1$ must be acquired randomly by the user.

IV. Motivating Examples

Example 1: We consider a single-database case $N = 1$, $K = 3$ and $L = 1$. We use W_1, W_2 and W_3 to denote the three message symbols uniformly selected from a finite field $\mathbb{F}_q$. The common randomness S_1, S_2 and S_3 stored in the database are also three uniformly selected symbols from $\mathbb{F}_q$. Our new achievable scheme is given in Table I. In Table I, we go from the table on the top to the table on the bottom by compact denotation of the queries as $q_1 = [W_1 + S_1, W_2 + S_2, W_3 + S_3]$, $q_2 = [W_1 + S_2, W_2 + S_3, W_3 + S_1]$ and $q_3 = [W_1 + S_3, W_2 + S_1, W_3 + S_2]$. This compact notation makes it more apparent that queries q_1, q_2, q_3 are used for all user-side common randomness settings, e.g., S_1, S_2, S_3 and for all desired messages, e.g., W_1, W_2, W_3, with equal probability.

The reliability constraint follows from the fact that the user can always decode the desired message by using its own common randomness. The database privacy constraint follows from the fact that the undesired messages are always mixed with unknown common randomness. For the user-privacy constraint, we have for all $k, k' \in [3], k' \neq k$ and a random selection $\mathcal{A}_U \in \{\{1\}, \{2\}, \{3\}\}$ under a uniform distribution, there exists another different $\mathcal{A}'_U \in \{\{1\}, \{2\}, \{3\}\}$, such that,

$$P\left(Q^{[k,\mathcal{A}_U]} = q\right) = P\left(Q^{[k',\mathcal{A}'_U]} = q\right) = \frac{1}{3} \tag{25}$$

where $q \in \{q_1, q_2, q_3\}$. Specifically from the point of view of the database, the same set of queries can be invoked for any desired message $W_i, i \in [3]$ with the same probability distribution. This scheme achieves $d = 3$, $\rho_U = 1$ and $\rho_S = 3$, which exactly matches the boundary of the SPIR capacity region for $N = 1$ and $K = 3$ in Corollary 3.

Example 2: We consider a multi-database case $N = 2$, $K = 2$ and $L = 4$. We use W_1 and W_2 to denote the two messages each consisting of 4 symbols that are uniformly selected

TABLE I
THE QUERY TABLE FOR THE CASE $N = 1, K = 3, L = 1$. THE TABLE AT THE BOTTOM DENOTES THE QUERY SETS COMPACTLY AS q_1, q_2 AND q_3

$\mathcal{R}_U$	desired message		
	W_1	W_2	W_3
S_1	W_1+S_1 W_2+S_2 W_3+S_3	W_2+S_1 W_3+S_2 W_1+S_3	W_3+S_1 W_1+S_2 W_2+S_3
S_2	W_1+S_2 W_2+S_3 W_3+S_1	W_2+S_2 W_3+S_3 W_1+S_1	W_3+S_2 W_1+S_3 W_2+S_1
S_3	W_1+S_3 W_2+S_1 W_3+S_2	W_2+S_3 W_3+S_1 W_1+S_2	W_3+S_3 W_1+S_1 W_2+S_2

$\mathcal{R}_U$	desired message		
	W_1	W_2	W_3
S_1	q_1	q_3	q_2
S_2	q_2	q_1	q_3
S_3	q_3	q_2	q_1

TABLE II
THE QUERY TABLE FOR THE CASE $N = 2, K = 2, L = 4$

$\mathcal{R}_U$	desired message: W_1		desired message: W_2	
	DB1	DB2	DB1	DB2
S_1	a_1+S_1 b_1+S_2 $a_3+b_2+S_3$	a_2+S_1 b_2+S_3 $a_4+b_1+S_2$	b_1+S_1 a_1+S_2 $b_3+a_2+S_3$	b_2+S_1 a_2+S_3 $b_4+a_1+S_2$
S_2	a_1+S_2 b_1+S_3 $a_3+b_2+S_1$	a_2+S_2 b_2+S_1 $a_4+b_1+S_3$	b_1+S_2 a_1+S_3 $b_3+a_2+S_1$	b_2+S_2 a_2+S_1 $b_4+a_1+S_3$
S_3	a_1+S_3 b_1+S_1 $a_3+b_2+S_2$	a_2+S_3 b_2+S_2 $a_4+b_1+S_1$	b_1+S_3 a_1+S_1 $b_3+a_2+S_2$	b_2+S_3 a_2+S_2 $b_4+a_1+S_1$

from a finite field $\mathbb{F}_q$. The common randomness S_1, S_2 and S_3 shared between the two databases are also uniformly selected symbols from $\mathbb{F}_q$. Then, we use $[a_1, a_2, a_3, a_4]$ as a random uniform permutation of the symbols in the first message W_1, and independently, $[b_1, b_2, b_3, b_4]$ as another random uniform permutation of the symbols in the second message W_2. Our new achievable scheme is given in Table II. Each set of queries shown in Table II (e.g., $a_1+S_1, b_1+S_2, a_3+b_2+S_3, a_2+S_1, b_2+S_3, a_4+b_1+S_2$) is one possible choice after performing message symbol index permutation and unknown server-side common randomness index permutation. Due to space limitations, we use one particular permutation to represent all possible permutation outcomes. During an actual implementation, the user should uniformly randomly select one random permutation out of all possible permutations.

Verification that this proposed scheme achieves reliability, user privacy and database privacy constraints is similar to the one analyzed in Example 1. Specifically with regard to the user privacy, for any $n \in [2]$, given a random selection $\mathcal{A}_U \in \{\{1\},\{2\},\{3\}\}$ under a uniform distribution, a user wishes to retrieve W_k, that database can always find $\mathcal{A}'_U \in \{\{1\},\{2\},\{3\}\}$ such that $Q_n^{[k',\mathcal{A}'_U]} = Q_n^{[k,\mathcal{A}_U]}$ for all $k' \neq k$. In other words, that database is not able to recognize the desired message index from the query taking into consideration message symbol index permutation and unknown server-side common randomness index permutation. This scheme achieves $d = \frac{3}{2}$, $\rho_U = \frac{1}{4}$ and $\rho_S = \frac{3}{4}$. This is the second corner point of the capacity region in Theorem 1 where all inequalities are satisfied with equality, i.e., here $\rho_U = \frac{1}{N^K}$, $d = d_{\text{PIR}} = 1 + \frac{1}{N}$, and $\rho_S = \frac{1}{N} d_{\text{PIR}}$.

Example 3: We consider a multi-database case $N = 3$, $K = 2$, $L = 36$ and $\rho_U = \frac{1}{18}$. We use W_1 and W_2 to denote the two messages each consisting of 36 symbols that are uniformly selected from a finite field $\mathbb{F}_q$. The common randomness $S_1, \ldots, S_{17}$ shared among the three database are also uniformly selected symbols from $\mathbb{F}_q$. Then, We use $[a_1, \ldots, a_{36}]$ as a random uniform permutation of the symbols in the first message W_1, and independently, $[b_1, \ldots, b_{36}]$ as another random uniform permutation of the symbols in the second message W_2. For the first 9 bits of the desired message after message symbol index permutation, i.e., $[a_1, \ldots, a_9]$, we utilize server-side common randomness $\{S_1, S_2, S_3, S_4\}$ and then our new achievable scheme for one random selection of unknown server-side common randomness index permutation is given in Table III. For the next 9 bits, i.e., $[a_{10}, \ldots, a_{18}]$, we select another set of server-side common randomness, e.g., $\{S_5, S_6, S_7, S_8\}$, and then use our scheme in Table III once more. For the last 18 bits, we use the classical SPIR scheme in [4].

We observe from Table III that, for the first 9 bits, this scheme achieves $D = 12$, $H(\mathcal{R}_U) = 1$ and $H(\mathcal{R}_S) = 4$. Doubling these, for the first 18 bits, this scheme achieves $D = 24$, $H(\mathcal{R}_U) = 2$ and $H(\mathcal{R}_S) = 8$. For the last 18 bits, we use the classical SPIR scheme in [4], which achieves $D = 27$, $H(\mathcal{R}_U) = 0$ and $H(\mathcal{R}_S) = 9$. Thus, by combining these two different schemes in a time-sharing manner, we ultimately have $d = \frac{24+27}{36} = \frac{17}{12}$, $\rho_U = \frac{2+0}{36} = \frac{1}{18}$, and $\rho_S = \frac{8+9}{36} = \frac{17}{36}$, which corresponds to a point on the line segment joining the first corner point where $\rho_U = 0$ and the second corner point where $\rho_U = \frac{1}{9}$ of the capacity region in Theorem 1.

V. CONVERSE PROOF

In this section, we provide the converse proof of Theorem 1. The four inequalities in Theorem 1 are proved in Lemmas 3, 4, 9 and 10 below. Towards proving these four lemmas, we need Lemmas 1-2 and Lemmas 5-8 below. We note that Lemmas 1-2 extend [3, Lemmas 5-6], and Lemmas 5 and 8 extend [4, Lemmas 1 and 2, eq. (39)]. These extensions are needed because we have two additional sets of random variables in our system model: $\mathcal{R}_S$ and $\mathcal{R}_U$ with respect to techniques in [3], and $\mathcal{R}_U$ with respect to techniques in [4].

Lemma 1 (Messages Dependence Upper Bound):

$$I\Big(W_{2:K}; Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, \mathcal{R}_S | W_1\Big) \leq D - L \tag{26}$$

Proof:

$$I\Big(W_{2:K}; Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, \mathcal{R}_S | W_1\Big)$$
$$= I\Big(W_{2:K}; Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, \mathcal{R}_S | W_1\Big) + I(W_{2:K}; W_1) \tag{27}$$

TABLE III
THE QUERY TABLE FOR THE FIRST 9 BITS IN THE CASE $N = 3, K = 2, L = 36$

$\mathcal{R}_U$	desired message: W_1			desired message: W_2		
	DB1	DB2	DB3	DB1	DB2	DB3
S_1	$a_1 + S_1$ $b_1 + S_2$ $a_4 + b_2 + S_3$ $a_5 + b_3 + S_4$	$a_2 + S_1$ $b_2 + S_3$ $a_6 + b_1 + S_2$ $a_7 + b_3 + S_4$	$a_3 + S_1$ $b_3 + S_4$ $a_8 + b_1 + S_2$ $a_9 + b_2 + S_3$	$b_1 + S_1$ $a_1 + S_2$ $b_4 + a_2 + S_3$ $b_5 + a_3 + S_4$	$b_2 + S_1$ $a_2 + S_3$ $b_6 + a_1 + S_2$ $b_7 + a_3 + S_4$	$b_3 + S_1$ $a_3 + S_4$ $b_8 + a_1 + S_2$ $b_9 + a_2 + S_3$
S_2	$a_1 + S_2$ $b_1 + S_3$ $a_4 + b_2 + S_4$ $a_5 + b_3 + S_1$	$a_2 + S_2$ $b_2 + S_4$ $a_6 + b_1 + S_3$ $a_7 + b_3 + S_1$	$a_3 + S_2$ $b_3 + S_1$ $a_8 + b_1 + S_3$ $a_9 + b_2 + S_4$	$b_1 + S_2$ $a_1 + S_3$ $b_4 + a_2 + S_4$ $b_5 + a_3 + S_1$	$b_2 + S_2$ $a_2 + S_4$ $b_6 + a_1 + S_3$ $b_7 + a_3 + S_1$	$b_3 + S_2$ $a_3 + S_1$ $b_8 + a_1 + S_3$ $b_9 + a_2 + S_4$
S_3	$a_1 + S_3$ $b_1 + S_4$ $a_4 + b_2 + S_1$ $a_5 + b_3 + S_2$	$a_2 + S_3$ $b_2 + S_1$ $a_6 + b_1 + S_4$ $a_7 + b_3 + S_2$	$a_3 + S_3$ $b_3 + S_2$ $a_8 + b_1 + S_4$ $a_9 + b_2 + S_1$	$b_1 + S_3$ $a_1 + S_4$ $b_4 + a_2 + S_1$ $b_5 + a_3 + S_2$	$b_2 + S_3$ $a_2 + S_1$ $b_6 + a_1 + S_4$ $b_7 + a_3 + S_2$	$b_3 + S_3$ $a_3 + S_2$ $b_8 + a_1 + S_4$ $b_9 + a_2 + S_1$
S_4	$a_1 + S_4$ $b_1 + S_1$ $a_4 + b_2 + S_2$ $a_5 + b_3 + S_3$	$a_2 + S_4$ $b_2 + S_2$ $a_6 + b_1 + S_1$ $a_7 + b_3 + S_3$	$a_3 + S_4$ $b_3 + S_3$ $a_8 + b_1 + S_1$ $a_9 + b_2 + S_2$	$b_1 + S_4$ $a_1 + S_1$ $b_4 + a_2 + S_2$ $b_5 + a_3 + S_3$	$b_2 + S_4$ $a_2 + S_2$ $b_6 + a_1 + S_1$ $b_7 + a_3 + S_3$	$b_3 + S_4$ $a_3 + S_3$ $b_8 + a_1 + S_1$ $b_9 + a_2 + S_2$

$$= I\Big(W_{2:K}; Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, W_1, \mathcal{R}_S\Big) \tag{28}$$

$$= I\Big(W_{2:K}; Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, \mathcal{R}_S\Big) + I\Big(W_{2:K}; W_1 | Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, \mathcal{R}_S\Big) \tag{29}$$

$$= I\Big(W_{2:K}; Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, \mathcal{R}_S\Big) \tag{30}$$

$$= I\Big(W_{2:K}; A_{1:N}^{[1,\mathcal{A}_U]} | Q_{1:N}^{[1,\mathcal{A}_U]}, \mathcal{R}_S\Big) + I\Big(W_{2:K}; Q_{1:N}^{[1,\mathcal{A}_U]}, \mathcal{R}_S\Big) \tag{31}$$

$$= I\Big(W_{2:K}; A_{1:N}^{[1,\mathcal{A}_U]} | Q_{1:N}^{[1,\mathcal{A}_U]}, \mathcal{R}_S\Big) \tag{32}$$

$$= H\Big(A_{1:N}^{[1,\mathcal{A}_U]} | Q_{1:N}^{[1,\mathcal{A}_U]}, \mathcal{R}_S\Big) - H\Big(A_{1:N}^{[1,\mathcal{A}_U]} | Q_{1:N}^{[1,\mathcal{A}_U]}, W_{2:K}, \mathcal{R}_S\Big) \tag{33}$$

$$= H\Big(A_{1:N}^{[1,\mathcal{A}_U]} | Q_{1:N}^{[1,\mathcal{A}_U]}, \mathcal{R}_S\Big) - H\Big(A_{1:N}^{[1,\mathcal{A}_U]} | Q_{1:N}^{[1,\mathcal{A}_U]}, W_{2:K}, \mathcal{R}_S\Big) - H\Big(W_1 | Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, W_{2:K}, \mathcal{R}_S\Big) \tag{34}$$

$$= H\Big(A_{1:N}^{[1,\mathcal{A}_U]} | Q_{1:N}^{[1,\mathcal{A}_U]}, \mathcal{R}_S\Big) - H\Big(W_1, A_{1:N}^{[1,\mathcal{A}_U]} | Q_{1:N}^{[1,\mathcal{A}_U]}, W_{2:K}, \mathcal{R}_S\Big) \tag{35}$$

$$\leq H\Big(A_{1:N}^{[1,\mathcal{A}_U]}\Big) - H\Big(W_1, A_{1:N}^{[1,\mathcal{A}_U]} | Q_{1:N}^{[1,\mathcal{A}_U]}, W_{2:K}, \mathcal{R}_S\Big) \tag{36}$$

$$\leq D - H\Big(W_1, A_{1:N}^{[1,\mathcal{A}_U]} | Q_{1:N}^{[1,\mathcal{A}_U]}, W_{2:K}, \mathcal{R}_S\Big) \tag{37}$$

$$= D - H\Big(W_1 | Q_{1:N}^{[1,\mathcal{A}_U]}, W_{2:K}, \mathcal{R}_S\Big) - H\Big(A_{1:N}^{[1,\mathcal{A}_U]} | Q_{1:N}^{[1,\mathcal{A}_U]}, W_1, W_{2:K}, \mathcal{R}_S\Big) \tag{38}$$

$$= D - H\Big(W_1 | Q_{1:N}^{[1,\mathcal{A}_U]}, W_{2:K}, \mathcal{R}_S\Big) \tag{39}$$

$$= D - L \tag{40}$$

where (27) follows from the i.i.d. message setting in the databases (3), (30) follows from the reliable decoding of the first message (10), (32) follows from the independence of the message set (6) and the independent query generation (8), (34) follows from the reliable decoding of the first message (10) again, (39) follows from the truthful deterministic answer generation at each database (9), (40) follows from the joint application of (2), (3), (6) and (8). ■

Lemma 2: (Messages Dependence Lower Bound):

$$I\Big(W_{k:K}; Q_{1:N}^{[k-1,\mathcal{A}_U]}, A_{1:N}^{[k-1,\mathcal{A}_U]}, \mathcal{R}_S | W_{1:k-1}\Big) \geq \frac{1}{N} I\Big(W_{k+1:K}; Q_{1:N}^{[k,\mathcal{A}'_U]}, A_{1:N}^{[k,\mathcal{A}'_U]}, \mathcal{R}_S | W_{1:k}\Big) + \frac{L}{N}, \ \forall k \in [2:K] \tag{41}$$

Proof:

$$NI\Big(W_{k:K}; Q_{1:N}^{[k-1,\mathcal{A}_U]}, A_{1:N}^{[k-1,\mathcal{A}_U]}, \mathcal{R}_S | W_{1:k-1}\Big) \geq \sum_{n=1}^{N} I\Big(W_{k:K}; Q_n^{[k-1,\mathcal{A}_U]}, A_n^{[k-1,\mathcal{A}_U]}, \mathcal{R}_S | W_{1:k-1}\Big) \tag{42}$$

$$= \sum_{n=1}^{N} I\Big(W_{k:K}; Q_n^{[k,\mathcal{A}'_U]}, A_n^{[k,\mathcal{A}'_U]}, \mathcal{R}_S | W_{1:k-1}\Big) \tag{43}$$

$$\geq \sum_{n=1}^{N} I\Big(W_{k:K}; A_n^{[k,\mathcal{A}'_U]} | Q_n^{[k,\mathcal{A}'_U]}, W_{1:k-1}, \mathcal{R}_S\Big) \tag{44}$$

$$= \sum_{n=1}^{N} \Big(H\Big(A_n^{[k,\mathcal{A}'_U]} | Q_n^{[k,\mathcal{A}'_U]}, W_{1:k-1}, \mathcal{R}_S\Big) - H\Big(A_n^{[k,\mathcal{A}'_U]} | Q_n^{[k,\mathcal{A}'_U]}, W_{1:K}, \mathcal{R}_S\Big)\Big) \tag{45}$$

$$= \sum_{n=1}^{N} H\Big(A_n^{[k,\mathcal{A}'_U]} | Q_n^{[k,\mathcal{A}'_U]}, W_{1:k-1}, \mathcal{R}_S\Big) \tag{46}$$

$$\geq \sum_{n=1}^{N} H\left(A_n^{[k,\mathcal{A}'_U]}|Q_{1:N}^{[k,\mathcal{A}'_U]}, A_{1:n-1}^{[k,\mathcal{A}'_U]}, W_{1:k-1}, \mathcal{R}_S\right) \quad (47)$$

$$= \sum_{n=1}^{N}\Big(H\left(A_n^{[k,\mathcal{A}'_U]}|Q_{1:N}^{[k,\mathcal{A}'_U]}, A_{1:n-1}^{[k,\mathcal{A}'_U]}, W_{1:k-1}, \mathcal{R}_S\right)$$
$$-H\left(A_n^{[k,\mathcal{A}'_U]}|Q_{1:N}^{[k,\mathcal{A}'_U]}, A_{1:n-1}^{[k,\mathcal{A}'_U]}, W_{1:K}, \mathcal{R}_S\right)\Big) \quad (48)$$

$$= \sum_{n=1}^{N} I\left(W_{k:K}; A_n^{[k,\mathcal{A}'_U]}|Q_{1:N}^{[k,\mathcal{A}'_U]}, A_{1:n-1}^{[k,\mathcal{A}'_U]}, W_{1:k-1}, \mathcal{R}_S\right) \quad (49)$$

$$= I\left(W_{k:K}; A_{1:N}^{[k,\mathcal{A}'_U]}|Q_{1:N}^{[k,\mathcal{A}'_U]}, W_{1:k-1}, \mathcal{R}_S\right) \quad (50)$$

$$= I\left(W_{k:K}; A_{1:N}^{[k,\mathcal{A}'_U]}|Q_{1:N}^{[k,\mathcal{A}'_U]}, W_{1:k-1}, \mathcal{R}_S\right)$$
$$+I\left(W_{k:K}; Q_{1:N}^{[k,\mathcal{A}'_U]}|W_{1:k-1}, \mathcal{R}_S\right) \quad (51)$$

$$= I\left(W_{k:K}; Q_{1:N}^{[k,\mathcal{A}'_U]}, A_{1:N}^{[k,\mathcal{A}'_U]}|W_{1:k-1}, \mathcal{R}_S\right) \quad (52)$$

$$= I\left(W_{k:K}; Q_{1:N}^{[k,\mathcal{A}'_U]}, A_{1:N}^{[k,\mathcal{A}'_U]}|W_{1:k-1}, \mathcal{R}_S\right)$$
$$+I\left(W_{k:K}; W_k|Q_{1:N}^{[k,\mathcal{A}'_U]}, A_{1:N}^{[k,\mathcal{A}'_U]}, W_{1:k-1}, \mathcal{R}_S\right) \quad (53)$$

$$= I\left(W_{k:K}; Q_{1:N}^{[k,\mathcal{A}'_U]}, A_{1:N}^{[k,\mathcal{A}'_U]}, W_k|W_{1:k-1}, \mathcal{R}_S\right) \quad (54)$$

$$= I(W_{k:K}; W_k|W_{1:k-1}, \mathcal{R}_S)$$
$$+I\left(W_{k:K}; Q_{1:N}^{[k,\mathcal{A}'_U]}, A_{1:N}^{[k,\mathcal{A}'_U]}|W_{1:k}, \mathcal{R}_S\right) \quad (55)$$

$$= I\left(W_{k:K}; Q_{1:N}^{[k,\mathcal{A}'_U]}, A_{1:N}^{[k,\mathcal{A}'_U]}|W_{1:k}, \mathcal{R}_S\right) + L \quad (56)$$

$$= I\left(W_{k+1:K}; Q_{1:N}^{[k,\mathcal{A}'_U]}, A_{1:N}^{[k,\mathcal{A}'_U]}|W_{1:k}, \mathcal{R}_S\right)$$
$$+I\left(W_k; Q_{1:N}^{[k,\mathcal{A}'_U]}, A_{1:N}^{[k,\mathcal{A}'_U]}|W_{1:k}, W_{k+1:K}, \mathcal{R}_S\right) + L \quad (57)$$

$$= I\left(W_{k+1:K}; Q_{1:N}^{[k,\mathcal{A}'_U]}, A_{1:N}^{[k,\mathcal{A}'_U]}|W_{1:k}, \mathcal{R}_S\right) + L \quad (58)$$

$$= I\left(W_{k+1:K}; Q_{1:N}^{[k,\mathcal{A}'_U]}, A_{1:N}^{[k,\mathcal{A}'_U]}|W_{1:k}, \mathcal{R}_S\right)$$
$$+I(W_{k+1:K}; \mathcal{R}_S|W_{1:k}) + L \quad (59)$$

$$= I\left(W_{k+1:K}; Q_{1:N}^{[k,\mathcal{A}'_U]}, A_{1:N}^{[k,\mathcal{A}'_U]}, \mathcal{R}_S|W_{1:k}\right) + L \quad (60)$$

where (43) follows from the application of user-privacy (11), (46) and (48) both follow from truthful deterministic answer generation by each database (9), (51) follows from (3), (6) and (8), (53) follows from reliability constraint (10), (56) follows from (2), (3) and (6), (59) follows from (3) and (6) again. Dividing both sides by N completes the proof. ■

Lemma 3 (Minimal Download Cost d):

$$d \geq 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1}} \quad (61)$$

Proof: Following steps similar to [3, eq. (62)-(67)] for Lemma 2, we obtain

$$I\left(W_{2:K}; Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, \mathcal{R}_S|W_1\right) \geq \left(\frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1}}\right)L \quad (62)$$

Combining the upper bound in Lemma 1 and the lower bound in (62) completes the proof. ■

Lemma 4 (Minimal Difference Between ρ_S and ρ_U):

$$\rho_S - \rho_U \geq \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1}} \quad (63)$$

Proof: From (62), we have,

$$I\left(W_{2:K}; Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, \mathcal{R}_S|W_1\right)$$
$$= H(W_{2:K}|W_1) - H\left(W_{2:K}|Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, W_1, \mathcal{R}_S\right) \quad (64)$$

$$= (K-1)L - H\left(W_{2:K}|Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, W_1, \mathcal{R}_S\right) \quad (65)$$

$$\geq \left(\frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1}}\right)L \quad (66)$$

Thus, we obtain,

$$H\left(W_{2:K}|Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, W_1, \mathcal{R}_S\right) \leq (K-1)L - \left(\frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1}}\right)L \quad (67)$$

Next, we have the following upper bound,

$$I\left(W_{2:K}; \mathcal{R}_S\backslash\mathcal{R}_U|Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, W_1, \mathcal{R}_U\right)$$
$$= H\left(\mathcal{R}_S\backslash\mathcal{R}_U|Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, W_1, \mathcal{R}_U\right)$$
$$-H\left(\mathcal{R}_S\backslash\mathcal{R}_U|Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, W_{1:K}, \mathcal{R}_U\right) \quad (68)$$

$$\leq H\left(\mathcal{R}_S\backslash\mathcal{R}_U|Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, W_1, \mathcal{R}_U\right) \quad (69)$$

$$= H(\mathcal{R}_S\backslash\mathcal{R}_U) \quad (70)$$

$$= H(\mathcal{R}_S) - H(\mathcal{R}_U) \quad (71)$$

where (70) follows from the presumed independence of the remaining common randomness among the databases when the retrieval is complete (16).

In addition, we have the following lower bound,

$$I\left(W_{2:K}; \mathcal{R}_S\backslash\mathcal{R}_U|Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, W_1, \mathcal{R}_U\right)$$
$$= H\left(W_{2:K}|Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, W_1, \mathcal{R}_U\right)$$
$$-H\left(W_{2:K}|Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, W_1, \mathcal{R}_S\right) \quad (72)$$

$$= (K-1)L - H\left(W_{2:K}|Q_{1:N}^{[1,\mathcal{A}_U]}, A_{1:N}^{[1,\mathcal{A}_U]}, W_1, \mathcal{R}_S\right) \quad (73)$$

$$\geq (K-1)L - (K-1)L + \left(\frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1}}\right)L \quad (74)$$

$$= \left(\frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1}}\right)L \quad (75)$$

where (73) follows from the database privacy constraint (15) in the realization of $k = 1$ and reliability constraint (10), and (74) follows from (67).

Combining (71) and (75) yields the desired result. ■

Lemma 5 (Different Indices Effect on the Same Message):

$$H\left(A_n^{[k',\mathcal{A}'_U]}|Q_n^{[k',\mathcal{A}'_U]}, W_k, \mathcal{R}'_U\right) - H\left(A_n^{[k,\mathcal{A}_U]}|Q_n^{[k,\mathcal{A}_U]}, W_k, \mathcal{R}_U\right) \leq H(\mathcal{R}_U), \quad \forall k' \neq k \quad (76)$$

Proof: From the user privacy constraint (11), we have,

$$H\left(W_k|Q_n^{[k,\mathcal{A}_U]},A_n^{[k,\mathcal{A}_U]},\mathcal{R}_U\right) = H\left(W_k|Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]},\mathcal{R}_U\right) \quad (77)$$

From the deterministic queries relying on the retrieval strategy (7), database privacy constraint (15), and the fact $W_k \in W_{\bar{k}'}$, we have,

$$\begin{aligned} 0 &= I\left(W_{\bar{k}'};Q_{1:N}^{[k',\mathcal{A}'_U]},A_{1:N}^{[k',\mathcal{A}'_U]},\mathcal{R}'_U\right) && (78)\\ &= I\left(W_k;Q_{1:N}^{[k',\mathcal{A}'_U]},A_{1:N}^{[k',\mathcal{A}'_U]},\mathcal{R}'_U\right) && (79)\\ &= I\left(W_k;Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]}|\mathcal{R}'_U\right) && (80)\\ &= H\left(W_k|Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]}\right) - H\left(W_k|Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]},\mathcal{R}'_U\right) && (81)\end{aligned}$$

Using the equations derived above, we derive an upper bound for the following term,

$$\begin{aligned} &H\left(W_k|Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]},\mathcal{R}'_U\right) - H\left(W_k|Q_n^{[k,\mathcal{A}_U]},A_n^{[k,\mathcal{A}_U]},\mathcal{R}_U\right)\\ &= H\left(W_k|Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]}\right) - H\left(W_k|Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]},\mathcal{R}_U\right) && (82)\\ &= I\left(W_k;\mathcal{R}_U|Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]}\right) && (83)\\ &= H\left(\mathcal{R}_U|Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]}\right) - H\left(\mathcal{R}_U|Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]},W_k\right) && (84)\\ &\le H\left(\mathcal{R}_U|Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]}\right) && (85)\\ &\le H(\mathcal{R}_U) && (86)\end{aligned}$$

where (82) follows from (77) and (81). This is different from the equivalence $H(W_k|Q_n^{[k']},A_n^{[k']}) = H(W_k|Q_n^{[k]},A_n^{[k]})$ in the SPIR problem without user-side common randomness and it leads to the difference between our Lemma 5 and [4, Lemma 1].

Once again from the user privacy constraint (11), we have,

$$H\left(Q_n^{[k,\mathcal{A}_U]},A_n^{[k,\mathcal{A}_U]},\mathcal{R}_S\right) = H\left(Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]},\mathcal{R}_S\right) \quad (87)$$

which is gives the following equality,

$$\begin{aligned} &H\left(\mathcal{R}_S\backslash\mathcal{R}_U|Q_n^{[k,\mathcal{A}_U]},A_n^{[k,\mathcal{A}_U]},\mathcal{R}_U\right) + H\left(Q_n^{[k,\mathcal{A}_U]},A_n^{[k,\mathcal{A}_U]},\mathcal{R}_U\right)\\ &= H\left(\mathcal{R}_S\backslash\mathcal{R}'_U|Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]},\mathcal{R}'_U\right) + H\left(Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]},\mathcal{R}'_U\right) \quad (88)\end{aligned}$$

Noting the independence of the remaining common randomness among the databases after the retrieval process (16), thus, we have,

$$\begin{aligned} &H(\mathcal{R}_S\backslash\mathcal{R}_U) + H\left(Q_n^{[k,\mathcal{A}_U]},A_n^{[k,\mathcal{A}_U]},\mathcal{R}_U\right)\\ &= H(\mathcal{R}_S\backslash\mathcal{R}'_U) + H\left(Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]},\mathcal{R}'_U\right) \quad (89)\end{aligned}$$

which leads to

$$H\left(Q_n^{[k,\mathcal{A}_U]},A_n^{[k,\mathcal{A}_U]},\mathcal{R}_U\right) = H\left(Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]},\mathcal{R}'_U\right) \quad (90)$$

Likewise, without taking into consideration the answers, we also have,

$$H\left(Q_n^{[k,\mathcal{A}_U]},\mathcal{R}_U\right) = H\left(Q_n^{[k',\mathcal{A}'_U]},\mathcal{R}'_U\right) \quad (91)$$

As a consequence, we derive the following relation by utilizing the independent message set (6) and also deterministic queries (7),

$$\begin{aligned} H\left(Q_n^{[k,\mathcal{A}_U]},W_k,\mathcal{R}_U\right) &= H\left(Q_n^{[k,\mathcal{A}_U]},\mathcal{R}_U\right) + H(W_k) && (92)\\ &= H\left(Q_n^{[k',\mathcal{A}'_U]},\mathcal{R}'_U\right) + H(W_k) && (93)\\ &= H\left(Q_n^{[k',\mathcal{A}'_U]},W_k,\mathcal{R}'_U\right) && (94)\end{aligned}$$

Now, we are ready to prove the Lemma 5,

$$\begin{aligned} &H\left(A_n^{[k,\mathcal{A}_U]}|Q_n^{[k,\mathcal{A}_U]},W_k,\mathcal{R}_U\right)\\ &= H\left(Q_n^{[k,\mathcal{A}_U]},A_n^{[k,\mathcal{A}_U]},W_k,\mathcal{R}_U\right) - H\left(Q_n^{[k,\mathcal{A}_U]},W_k,\mathcal{R}_U\right) && (95)\\ &= H\left(W_k|Q_n^{[k,\mathcal{A}_U]},A_n^{[k,\mathcal{A}_U]},\mathcal{R}_U\right) + H\left(Q^{[k,\mathcal{A}_U]},A_n^{[k,\mathcal{A}_U]},\mathcal{R}_U\right) - H\left(Q_n^{[k,\mathcal{A}_U]},W_k,\mathcal{R}_U\right) && (96)\\ &\ge H\left(W_k|Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]},\mathcal{R}'_U\right) + H\left(Q_n^{[k,\mathcal{A}_U]},A_n^{[k,\mathcal{A}_U]},\mathcal{R}_U\right) - H\left(Q_n^{[k,\mathcal{A}_U]},W_k,\mathcal{R}_U\right) - H(\mathcal{R}_U) && (97)\\ &= H\left(W_k|Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]},\mathcal{R}'_U\right) + H\left(Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]},\mathcal{R}'_U\right) - H\left(Q_n^{[k',\mathcal{A}'_U]},W_k,\mathcal{R}'_U\right) - H(\mathcal{R}_U) && (98)\\ &= H\left(Q_n^{[k',\mathcal{A}'_U]},A_n^{[k',\mathcal{A}'_U]},W_k,\mathcal{R}'_U\right) - H\left(Q_n^{[k',\mathcal{A}'_U]},W_k,\mathcal{R}'_U\right) - H(\mathcal{R}_U) && (99)\\ &= H\left(A_n^{[k',\mathcal{A}'_U]}|Q_n^{[k',\mathcal{A}'_U]},W_k,\mathcal{R}'_U\right) - H(\mathcal{R}_U) && (100)\end{aligned}$$

where (97) follows from (86), and (98) follows from (90) and (94). ■

Lemma 6 (Symmetry):

$$H\left(A_n^{[k,\mathcal{A}_U]}|Q_n^{[k,\mathcal{A}_U]},\mathcal{R}_U\right) = H\left(A_n^{[k',\mathcal{A}'_U]}|Q_n^{[k',\mathcal{A}'_U]},\mathcal{R}'_U\right), \quad \forall k' \neq k \quad (101)$$

Proof: The proof of Lemma 6 follows from (90) and (91). ■

Lemma 7: (Effect of Conditioning on Retrieval Strategy):

$$H\left(A_n^{[k,\mathcal{A}_U]}|\mathcal{F},Q_n^{[k,\mathcal{A}_U]},W_k,\mathcal{R}_U\right) = H\left(A_n^{[k,\mathcal{A}_U]}|Q_n^{[k,\mathcal{A}_U]},W_k,\mathcal{R}_U\right) \quad (102)$$

Proof: We prove Lemma 7 by showing that the following conditional mutual information is non-positive and thus is zero,

$$I\left(A_n^{[k,\mathcal{A}_U]};\mathcal{F}|Q_n^{[k,\mathcal{A}_U]},W_k,\mathcal{R}_U\right) \le I\left(A_n^{[k,\mathcal{A}_U]},W_{1:K},\mathcal{R}_S\backslash\mathcal{R}_U;\mathcal{F}|Q_n^{[k,\mathcal{A}_U]},W_k,\mathcal{R}_U\right) \quad (103)$$

$$= I\Big(W_{1:K}, \mathcal{R}_S\backslash\mathcal{R}_U; \mathcal{F}|Q_n^{[k,\mathcal{A}_U]}, W_k, \mathcal{R}_U\Big) + I\Big(A_n^{[k,\mathcal{A}_U]}; \mathcal{F}|Q_n^{[k,\mathcal{A}_U]}, W_{1:K}, \mathcal{R}_S\Big) \quad (104)$$

$$= I\Big(W_{1:K}, \mathcal{R}_S\backslash\mathcal{R}_U; \mathcal{F}|Q_n^{[k,\mathcal{A}_U]}, W_k, \mathcal{R}_U\Big) \quad (105)$$

$$\le I\Big(W_{1:K}, \mathcal{R}_S\backslash\mathcal{R}_U; \mathcal{F}|Q_n^{[k,\mathcal{A}_U]}, W_k, \mathcal{R}_U\Big) + I\Big(W_k; \mathcal{F}|Q_n^{[k,\mathcal{A}_U]}, \mathcal{R}_U\Big) \quad (106)$$

$$= I\Big(W_{1:K}, \mathcal{R}_S\backslash\mathcal{R}_U; \mathcal{F}|Q_n^{[k,\mathcal{A}_U]}, \mathcal{R}_U\Big) \quad (107)$$

$$\le I\Big(W_{1:K}, \mathcal{R}_S\backslash\mathcal{R}_U; \mathcal{F}, Q_n^{[k,\mathcal{A}_U]}, \mathcal{R}_U\Big) \quad (108)$$

$$= I\Big(\mathcal{R}_S\backslash\mathcal{R}_U; \mathcal{F}, Q_n^{[k,\mathcal{A}_U]}, \mathcal{R}_U\Big) + I\Big(W_{1:K}; \mathcal{F}, Q_n^{[k,\mathcal{A}_U]}, \mathcal{R}_U|\mathcal{R}_S\backslash\mathcal{R}_U\Big) \quad (109)$$

$$= I\Big(\mathcal{R}_S\backslash\mathcal{R}_U; \mathcal{F}, Q_n^{[k,\mathcal{A}_U]}, \mathcal{R}_U\Big) + H(W_{1:K}|\mathcal{R}_S\backslash\mathcal{R}_U) - H\Big(W_{1:K}|\mathcal{F}, Q_n^{[k,\mathcal{A}_U]}, \mathcal{R}_S\Big) \quad (110)$$

$$= I\Big(\mathcal{R}_S\backslash\mathcal{R}_U; \mathcal{F}, Q_n^{[k,\mathcal{A}_U]}, \mathcal{R}_U\Big) + H(W_{1:K}) - H(W_{1:K}) \quad (111)$$

$$= I\Big(\mathcal{R}_S\backslash\mathcal{R}_U; \mathcal{F}, Q_n^{[k,\mathcal{A}_U]}, \mathcal{R}_U\Big) \quad (112)$$

$$= H(\mathcal{R}_S\backslash\mathcal{R}_U) - H\Big(\mathcal{R}_S\backslash\mathcal{R}_U|\mathcal{F}, Q_n^{[k,\mathcal{A}_U]}, \mathcal{R}_U\Big) \quad (113)$$

$$= H(\mathcal{R}_S\backslash\mathcal{R}_U) - H(\mathcal{R}_S\backslash\mathcal{R}_U) \quad (114)$$

$$= 0 \quad (115)$$

where (105) follows from the fact that the answer is a deterministic function of the corresponding query, message set and server-side common randomness (9), (111) follows from the independence of message set (6) and the query is a deterministic function of the realization of retrieval strategy randomness (7), and (114) follows from the independence of the remaining common randomness among the databases (7) and (16). ■

Lemma 8 (Effect of Conditioning on Undesired Message):

$$H\Big(A_n^{[k',\mathcal{A}'_U]}|Q_n^{[k',\mathcal{A}'_U]}, \mathcal{R}'_U\Big) = H\Big(A_n^{[k',\mathcal{A}'_U]}|Q_n^{[k',\mathcal{A}'_U]}, W_k, \mathcal{R}'_U\Big), \quad \forall k' \neq k \quad (116)$$

Proof: From the database privacy constraint (15) and noting that $W_k \in W_{\bar{k}'}$, we have,

$$0 = I\Big(W_{\bar{k}'}; Q_{1:N}^{[k',\mathcal{A}'_U]}, A_{1:N}^{[k',\mathcal{A}'_U]}, \mathcal{R}'_U\Big) \quad (117)$$

$$= I\Big(W_k; Q_{1:N}^{[k',\mathcal{A}'_U]}, A_{1:N}^{[k',\mathcal{A}'_U]}, \mathcal{R}'_U\Big) \quad (118)$$

$$= I\Big(A_n^{[k',\mathcal{A}'_U]}; W_k|Q_n^{[k',\mathcal{A}'_U]}, \mathcal{R}'_U\Big) \quad (119)$$

$$= H\Big(A_n^{[k',\mathcal{A}'_U]}|Q_n^{[k',\mathcal{A}'_U]}, \mathcal{R}'_U\Big) - H\Big(A_n^{[k',\mathcal{A}'_U]}|Q_n^{[k',\mathcal{A}'_U]}, W_k, \mathcal{R}'_U\Big) \quad (120)$$

which is the desired result. ■

Lemma 9 (Minimal Bound for d and ρ_U):

$$\frac{N-1}{N}d + \rho_U \ge 1 \quad (121)$$

Proof: Starting from the message length assumption (2),

$$L = H(W_k) \quad (122)$$

$$= H(W_k|\mathcal{F}, \mathcal{R}_U) \quad (123)$$

$$= H(W_k|\mathcal{F}, \mathcal{R}_U) - H\Big(W_k|\mathcal{F}, A_{1:N}^{[k,\mathcal{A}_U]}, \mathcal{R}_U\Big) \quad (124)$$

$$= I\Big(W_k; A_{1:N}^{[k,\mathcal{A}_U]}|\mathcal{F}, \mathcal{R}_U\Big) \quad (125)$$

$$= H\Big(A_{1:N}^{[k,\mathcal{A}_U]}|\mathcal{F}, \mathcal{R}_U\Big) - H\Big(A_{1:N}^{[k,\mathcal{A}_U]}|\mathcal{F}, W_k, \mathcal{R}_U\Big) \quad (126)$$

$$= H\Big(A_{1:N}^{[k,\mathcal{A}_U]}|\mathcal{F}, \mathcal{R}_U\Big) - H\Big(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, Q_n^{[k,\mathcal{A}_U]}, W_k, \mathcal{R}_U\Big) \quad (127)$$

$$= H\Big(A_{1:N}^{[k,\mathcal{A}_U]}|\mathcal{F}, \mathcal{R}_U\Big) - H\Big(A_n^{[k,\mathcal{A}_U]}|Q_n^{[k,\mathcal{A}_U]}, W_k, \mathcal{R}_U\Big) \quad (128)$$

$$\le H\Big(A_{1:N}^{[k,\mathcal{A}_U]}|\mathcal{F}, \mathcal{R}_U\Big) - H\Big(A_n^{[k',\mathcal{A}'_U]}|Q_n^{[k',\mathcal{A}'_U]}, W_k, \mathcal{R}'_U\Big) + H(\mathcal{R}_U) \quad (129)$$

$$= H\Big(A_{1:N}^{[k,\mathcal{A}_U]}|\mathcal{F}, \mathcal{R}_U\Big) - H\Big(A_n^{[k',\mathcal{A}'_U]}|Q_n^{[k',\mathcal{A}'_U]}, \mathcal{R}'_U\Big) + H(\mathcal{R}_U) \quad (130)$$

$$= H\Big(A_{1:N}^{[k,\mathcal{A}_U]}|\mathcal{F}, \mathcal{R}_U\Big) - H\Big(A_n^{[k,\mathcal{A}_U]}|Q_n^{[k,\mathcal{A}_U]}, \mathcal{R}_U\Big) + H(\mathcal{R}_U) \quad (131)$$

$$\le H\Big(A_{1:N}^{[k,\mathcal{A}_U]}|\mathcal{F}, \mathcal{R}_U\Big) - H\Big(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, \mathcal{R}_U\Big) + H(\mathcal{R}_U) \quad (132)$$

where (123) follows from the independence of the message set (6), (124) follows from the reliable decoding of message W_k, (127) and (132) both follow from the fact that each query is determined by the retrieval strategy (7), (128) follows from Lemma 7, (129) follows from Lemma 5, (130) follows from Lemma 8, (131) follows from Lemma 6.

By summing (132) over all $n \in [1 : N]$, we obtain the following relationship,

$$NL \le NH\Big(A_{1:N}^{[k,\mathcal{A}_U]}|\mathcal{F}, \mathcal{R}_U\Big) - \sum_{n=1}^{N} H\Big(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, \mathcal{R}_U\Big) + NH(\mathcal{R}_U) \quad (133)$$

$$\le (N-1)H\Big(A_{1:N}^{[k,\mathcal{A}_U]}|\mathcal{F}, \mathcal{R}_U\Big) + NH(\mathcal{R}_U) \quad (134)$$

$$\le (N-1)\sum_{n=1}^{N} H\Big(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, \mathcal{R}_U\Big) + NH(\mathcal{R}_U) \quad (135)$$

$$\le (N-1)D + NH(\mathcal{R}_U) \quad (136)$$

which completes the proof. ■

Lemma 10 (Minimal Bound for ρ_U and ρ_S):

$$\frac{N}{N-1}\rho_U + N\rho_S \ge \frac{N}{N-1} \quad (137)$$

Proof: Starting with the database privacy constraint (15),

$$0 = I\Big(W_{\bar{k}}; \mathcal{F}, A_{1:N}^{[k,\mathcal{A}_U]}, \mathcal{R}_U\Big) \quad (138)$$

$$= I\Big(W_{\bar{k}}; A_{1:N}^{[k,\mathcal{A}_U]}, \mathcal{R}_U|\mathcal{F}\Big) \quad (139)$$

$$= I\Big(W_{\bar{k}}; A_{1:N}^{[k,\mathcal{A}_U]}, \mathcal{R}_U|\mathcal{F}\Big) + I\Big(W_{\bar{k}}; W_k|\mathcal{F}, A_{1:N}^{[k,\mathcal{A}_U]}, \mathcal{R}_U\Big) \quad (140)$$

$$= I\Big(W_{\bar{k}}; A_{1:N}^{[k,\mathcal{A}_U]}, W_k, \mathcal{R}_U|\mathcal{F}\Big) \quad (141)$$

$$= I\Big(W_{\bar{k}}; A_{1:N}^{[k,\mathcal{A}_U]}|\mathcal{F}, W_k, \mathcal{R}_U\Big) + I\big(W_{\bar{k}}; W_k, \mathcal{R}_U|\mathcal{F}\big) \quad (142)$$

$$= I\Big(W_{\bar{k}}; A_{1:N}^{[k,\mathcal{A}_U]}|\mathcal{F}, W_k, \mathcal{R}_U\Big) \quad (143)$$

$$\geq I\left(W_{\bar{k}}; A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, W_k, \mathcal{R}_U\right) \quad (144)$$

$$= H\left(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, W_k, \mathcal{R}_U\right) - H\left(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, W_{1:K}, \mathcal{R}_U\right) + H\left(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, W_{1:K}, \mathcal{R}_S\right) \quad (145)$$

$$\geq H\left(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, W_k, \mathcal{R}_U\right) - H\left(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, W_{1:K}, \mathcal{R}_U\right) + H\left(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, W_{1:K}, \mathcal{R}_S, \mathcal{R}_U\right) \quad (146)$$

$$= H\left(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, W_k, \mathcal{R}_U\right) - I\left(A_n^{[k,\mathcal{A}_U]}; \mathcal{R}_S|\mathcal{F}, W_{1:K}, \mathcal{R}_U\right) \quad (147)$$

$$= H\left(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, W_k, \mathcal{R}_U\right) - H(\mathcal{R}_S|\mathcal{F}, W_{1:K}, \mathcal{R}_U) + H\left(\mathcal{R}_S|\mathcal{F}, A_n^{[k,\mathcal{A}_U]}, W_{1:K}, \mathcal{R}_U\right) \quad (148)$$

$$\geq H\left(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, W_k, \mathcal{R}_U\right) - H(\mathcal{R}_S|\mathcal{F}, W_k, \mathcal{R}_U) \quad (149)$$

$$= H\left(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, W_k, \mathcal{R}_U\right) - H(\mathcal{R}_U, \mathcal{R}_S\backslash\mathcal{R}_U|\mathcal{F}, W_k, \mathcal{R}_U) \quad (150)$$

$$= H\left(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, W_k, \mathcal{R}_U\right) - H(\mathcal{R}_S\backslash\mathcal{R}_U|\mathcal{F}, W_k, \mathcal{R}_U) \quad (151)$$

$$= H\left(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, W_k, \mathcal{R}_U\right) - H(\mathcal{R}_S\backslash\mathcal{R}_U) \quad (152)$$

$$= H\left(A_n^{[k,\mathcal{A}_U]}|\mathcal{F}, Q_n^{[k,\mathcal{A}_U]}, W_k, \mathcal{R}_U\right) - H(\mathcal{R}_S) + H(\mathcal{R}_U) \quad (153)$$

$$= H\left(A_n^{[k,\mathcal{A}_U]}|Q_n^{[k,\mathcal{A}_U]}, \mathcal{R}_U\right) - H(\mathcal{R}_S) \quad (154)$$

where (140) follows from the reliability constraint (10), (143) follows from (3) and (6), (145) follows from the deterministic answer generation by each database (9) and (7), (152) follows from the independent remaining common randomness among the databases (16), (153) follows from the deterministic queries relying on the retrieval strategy (7), and (154) follows from the steps between (128)-(131) by applying Lemma 5 through Lemma 8 again.

By summing (154) over all $n \in [1 : N]$, we obtain the following relationship,

$$0 \geq \sum_{n=1}^{N} H\left(A_n^{[k,\mathcal{A}_U]}|Q_n^{[k,\mathcal{A}_U]}, \mathcal{R}_U\right) - NH(\mathcal{R}_S) \quad (155)$$

$$\geq H\left(A_{1:N}^{[1,\mathcal{A}_U]}|\mathcal{F}, Q_n^{[1,\mathcal{A}_U]}, \mathcal{R}_U\right) - NH(\mathcal{R}_S) \quad (156)$$

$$= H\left(A_{1:N}^{[1,\mathcal{A}_U]}|\mathcal{F}, \mathcal{R}_U\right) - NH(\mathcal{R}_S) \quad (157)$$

$$\geq \frac{N}{N-1}L - \frac{N}{N-1}H(\mathcal{R}_U) - NH(\mathcal{R}_S) \quad (158)$$

where (158) follows from (134), completing the proof. ■

VI. Achievability Proof

Following the critical idea in [59], our new achievable scheme corresponding to the second corner point in Theorem 1 is based on the principle of converting a given PIR scheme into a valid SPIR scheme using the server-side and user-side common randomness in a manner that does not compromise the download cost. To that end, given any existing information-theoretic PIR achievable scheme, we add a new distinct common randomness to each message symbol. The common randomness added to the desired symbols are substracted out as they are available at the user side, and the remaining common randomness unknown to the user are used to protect the undesired messages. There are two main challenges to constructing such an achievable scheme: first is to simultaneously reduce the amount of required server-side and user-side common randomness to the extent possible, and second is to implement this achievable scheme for all possible user-side common randomness realizations which are unknown ahead of time. By means of converting the PIR scheme in [3] to a corresponding valid SPIR scheme, our proposed new achievable scheme consists of the following steps:

1) *Initial PIR query generation:* For given N and K, generate an initial PIR query table for each desired message using the scheme in [3], e.g., Tables I–III without common randomness S_i's.
2) *Server-side common randomness assignment:* Mix all 1-sum symbols from the desired message across all the databases with the same new common randomness. We call it seed common randomness (e.g., S_1 in first three rows of Table II). Assign a new distinct common randomness to every 1-sum symbol from the undesired messages. For every k-sum symbol containing a desired message symbol, mix it with the common randomness from the $(k-1)$-sum symbol having the same $k-1$ undesired message symbols queried at another database. For every k-sum symbol not containing any desired message symbol, assign a new distinct common randomness. Repeat this until k reaches K. We call this whole modified query table a *query cell*.[5]
3) *Server-side common randomness cycling:* While keeping each query cell, create a new one by adding 1 (mod $|\mathcal{A}|$) to each common randomness index (e.g., S_1 becomes S_2 in Table II). Repeat it $|\mathcal{A}|$ times such that each query cell has a different seed common randomness index.
4) *Query cell determination:* The user has $|\mathcal{A}_U|$ server-side common randomness. The user determines the query cell to be invoked, and selects a random permutation within that cell, by matching its user-side common randomness to the seed common randomness of the cell.

Reliability: The reliability follows from the reliability of the PIR achievable scheme in [3]. One of the desired message symbols is coupled with a common randomness that is known to the user in advance. The other desired message symbols are coupled with interference that are downloaded from other databases.

User Privacy: From the perspective of each database, the same query can be adopted for any desired message with

[5]As we did in Example 2, in this step and next step, we use one particular permutation to represent all possible permutation outcomes coming from message symbol index permutation and unknown server-side common randomness index permutation. We do not show all possible permutations for simplicity.

equal probability. Specifically, as in (14), for any $n \in [N]$, any $k \in [K]$, any provided $\mathcal{A}_U$, any selected query q, we always have $P(Q_n^{[k,\mathcal{A}_U]} = q)$ being a constant, which does not depend on the realizations of n, k, $\mathcal{A}_U$ and q.

Database Privacy: From the perspective of the user, every undesired message symbol is always mixed with some unknown common randomness. As a result, no information about undesired message is leaked to the user.

Performance: We compute the performance of the proposed achievable scheme with regard to ρ_S, ρ_U and d. As in [3], the message length L is N^K,[6] and d is $1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1}}$ because the total number of downloaded symbols across all the databases does not change. Combining the first statement of [3, Lemma 1] and our assignment of server-side common randomness in step 2, we calculate the value of $|\mathcal{A}|$,

$$|\mathcal{A}| = 1 + N \cdot \sum_{k=1}^{K-1} (N-1)^{k-1} \binom{K-1}{k} \tag{159}$$

$$= 1 + \frac{N}{N-1} \cdot \sum_{k=1}^{K-1} (N-1)^{k} \binom{K-1}{k} \tag{160}$$

$$= 1 + \frac{N}{N-1} \cdot \left(\sum_{k=0}^{K-1} \binom{K-1}{k} (N-1)^k 1^{K-1-k} - 1 \right) \tag{161}$$

$$= 1 + \frac{N}{N-1} \cdot \left((N-1+1)^{K-1} - 1 \right) \tag{162}$$

$$= 1 + \frac{N}{N-1} \cdot \left(N^{K-1} - 1 \right) \tag{163}$$

$$= \frac{N^K - 1}{N-1} \tag{164}$$

$$= 1 + \cdots + N^{K-1} \tag{165}$$

which implies that $\rho_S = \frac{H(\mathcal{R}_S)}{L} = \frac{|\mathcal{A}|}{L}$ is $\frac{1}{N} + \cdots + \frac{1}{N^K}$ since $L = N^K$. The total amount of required user-side common randomness $|\mathcal{A}_U|$ is 1 since the user only has one seed common randomness before the retrieval takes place. Thus, $\rho_U = \frac{H(\mathcal{R}_U)}{L} = \frac{|\mathcal{A}_U|}{L}$ is $\frac{1}{N^K}$ since $L = N^K$.

VII. Conclusion

We considered SPIR which is a fundamental primitive in cryptography, as an essential building block in many cryptographic applications, such as, oblivious transfer, secure multi-party computation and zero knowledge proofs. Single-database SPIR could be critical in applications where colluding of all databases cannot be ruled out. Further, side-information and/or cached information is a useful dimension to explore to improve private download rates. In this paper, we introduced an extended version of the SPIR problem, where the user randomly fetches a portion of the available shared common randomness at the databases. This fetched database common randomness can be viewed as a form of side-information at the user. We showed that this side-information increases the SPIR rate, and it can increase it to the level of PIR rate. Since single-database SPIR is infeasible while single-database PIR is feasible, the proposed non-trivial use of user-side common randomness makes single-database SPIR feasible. In this paper, we determined the exact capacity region of the download cost, database-side common randomness, and user-side common randomness. Open problems include considering upload cost together with download cost in this system and encoding user- and server-side common randomness as in the coded PIR problem.

[6] In Examples 1-2, the message length is strictly $L = N^K$. However, in Example 3, we note that the classical SPIR scheme achieving d_{SPIR} in [4] requires the message length to be a multiple of $N - 1 = 2$, and our new SPIR scheme achieving d_{PIR} requires the message length to be a multiple of $N^K = 9$. In order to execute an appropriate half-to-half time-sharing between these two different schemes, we set the overall message length to be 36.

References

[1] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private information retrieval," *J. ACM*, vol. 45, no. 6, pp. 965–981, Nov. 1998.
[2] Y. Gertner, Y. Ishai, E. Kushilevitz, and T. Malkin, "Protecting data privacy in private information retrieval schemes," in *Proc. 30th Annu. ACM Symp. Theory Comput.*, May 1998, pp. 151–160.
[3] 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.
[4] H. Sun and S. A. Jafar, "The capacity of symmetric private information retrieval," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 322–329, Jan. 2019.
[5] Z. Wang, K. Banawan, and S. Ulukus, "Private set intersection: A multi-message symmetric private information retrieval perspective," *IEEE Trans. Inf. Theory*, vol. 68, no. 3, pp. 2001–2019, Mar. 2022.
[6] Z. Wang, K. Banawan, and S. Ulukus, "Multi-party private set intersection: An information-theoretic approach," *IEEE J. Sel. Areas Inf. Theory*, vol. 2, no. 1, pp. 366–379, Mar. 2021.
[7] U. Feige, J. Killian, and M. Naor, "A minimal model for secure computation," in *Proc. 26th Annu. ACM Symp. Theory Comput.*, 1994, pp. 554–563.
[8] O. Goldreich, "Secure multi-party computation," Mar. 1999.
[9] W. Du and M. J. Atallah, "Secure multi-party computation problems and their applications: A review and open problems," in *Proc. Workshop New Security Paradigms*, 2001, pp. 13–22.
[10] D. Evans, V. Kolesnikov, and M. Rosulek, "A pragmatic introduction to secure multi-party computation," *Found. Trends Privacy Security*, vol. 2, nos. 2–3, pp. 70–246, Dec. 2018.
[11] M. J. Freedman, K. Nissim, and B. Pinkas, "Efficient private matching and set intersection," in *Advances in Cryptology (EUROCRYPT)*. Berlin, Germany: Springer, 2004, pp. 1–19.
[12] L. Kissner and D. Song, "Privacy-preserving set operations," in *Advances in Cryptology (CRYPTO)*. Berlin, Germany: Springer, 2005, pp. 241–257.
[13] Y. Zhao and H. Sun, "Expand-and-randomize: An algebraic approach to secure computation," *Entropy*, vol. 23, no. 11, p. 1461, Nov. 2021.
[14] 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.
[15] K. Banawan and S. Ulukus, "The capacity of private information retrieval from Byzantine and colluding databases," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 1206–1219, Feb. 2019.
[16] S. Kadhe, B. Garcia, A. Heidarzadeh, S. E. Rouayheb, and A. Sprintson, "Private information retrieval with side information," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2032–2043, Apr. 2020.
[17] S. Kadhe, A. Heidarzadeh, A. Sprintson, and O. O. Koyluoglu, "Single-server private information retrieval schemes are equivalent to locally recoverable coding schemes," *IEEE J. Sel. Areas Inf. Theory*, vol. 2, no. 1, pp. 391–402, Mar. 2021.
[18] A. Heidarzadeh, B. Garcia, S. Kadhe, S. E. Rouayheb, and A. Sprintson, "On the capacity of single-server multi-message private information retrieval with side information," in *Proc. Allerton Conf.*, Oct. 2018, pp. 180–187.
[19] S. Li and M. Gastpar, "Single-server multi-message private information retrieval with side information," in *Proc. Allerton Conf.*, Oct. 2018, pp. 173–179.
[20] R. Tandon, "The capacity of cache aided private information retrieval," in *Proc. Allerton Conf.*, Oct. 2017, pp. 1078–1082.
[21] Y.-P. Wei, K. Banawan, and S. Ulukus, "Cache-aided private information retrieval with partially known uncoded prefetching: Fundamental limits," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 6, pp. 1126–1139, Jun. 2018.
[22] Y.-P. Wei, K. Banawan, and S. Ulukus, "Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 3215–3232, May 2019.
[23] Y.-P. Wei, K. Banawan, and S. Ulukus, "The capacity of private information retrieval with partially known private side information," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8222–8231, Dec. 2019.
[24] Y.-P. Wei and S. Ulukus, "The capacity of private information retrieval with private side information under storage constraints," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2023–2031, Apr. 2020.
[25] Z. Chen, Z. Wang, and S. A. Jafar, "The capacity of T-private information retrieval with private side information," *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 4761–4773, Aug. 2020.

[26] M. J. Siavoshani, S. P. Shariatpanahi, and M. A. Maddah-Ali, "Private information retrieval for a multi-message scenario with private side information," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3235–3244, May 2021.
[27] K. Banawan and S. Ulukus, "Multi-message private information retrieval: Capacity results and near-optimal schemes," *IEEE Trans. Inf. Theory*, vol. 64, no. 10, pp. 6842–6862, Oct. 2018.
[28] N. B. Shah, K. V. Rashmi, and K. Ramchandran, "One extra bit of download ensures perfectly private information retrieval," in *Proc. IEEE ISIT*, Jun. 2014, pp. 856–860.
[29] 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.
[30] S. Kumar, H.-Y. Lin, E. Rosnes, and A. G. I. Amat, "Achieving maximum distance separable private information retrieval capacity with linear codes," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4243–4273, Jul. 2019.
[31] R. Zhou, C. Tian, H. Sun, and T. Liu, "Capacity-achieving private information retrieval codes from MDS-coded databases with minimum message size," *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 4904–4916, Aug. 2020.
[32] Q. Wang and M. Skoglund, "Symmetric private information retrieval from MDS coded distributed storage with non-colluding and colluding servers," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 5160–5175, Aug. 2019.
[33] Q. Wang, H. Sun, and M. Skoglund, "Symmetric private information retrieval with mismatched coded messages and randomness," in *Proc. IEEE ISIT*, Jul. 2019, pp. 365–369.
[34] Q. Wang and M. Skoglund, "On PIR and symmetric PIR from colluding databases with adversaries and eavesdroppers," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 3183–3197, May 2019.
[35] K. Banawan and S. Ulukus, "Private information retrieval through wiretap channel II: Privacy meets security," *IEEE Trans. Inf. Theory*, vol. 66, no. 7, pp. 4129–4149, Jul. 2020.
[36] J. Cheng, N. Liu, and W. Kang, "The capacity of symmetric private information retrieval under arbitrary collusion and eavesdropping patterns," 2020, *arXiv:2010.08249*.
[37] S. Kumar, A. G. I. Amat, E. Rosnes, and L. Senigagliesi, "Private information retrieval from a cellular network with caching at the edge," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 4900–4912, Jul. 2019.
[38] T. Guo, R. Zhou, and C. Tian, "On the information leakage in private information retrieval systems," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2999–3012, 2020.
[39] H.-Y. Lin, S. Kumar, E. Rosnes, A. G. I. Amat, and E. Yaakobi, "Multi-server weakly-private information retrieval," *IEEE Trans. Inf. Theory*, vol. 68, no. 2, pp. 1197–1219, Feb. 2022.
[40] I. Samy, M. Attia, R. Tandon, and L. Lazos, "Asymmetric leaky private information retrieval," *IEEE Trans. Inf. Theory*, vol. 67, no. 8, pp. 5352–5369, Aug. 2021.
[41] C. Tian, H. Sun, and J. Chen, "Capacity-achieving private information retrieval codes with optimal message size and upload cost," *IEEE Trans. Inf. Theory*, vol. 65, no. 11, pp. 7613–7627, Nov. 2019.
[42] Y. Zhou, Q. Wang, H. Sun, and S. Fu, "The minimum upload cost of symmetric private information retrieval," in *Proc. IEEE ISIT*, Jun. 2020, pp. 1030–1034.
[43] H. Yang, W. Shin, and J. Lee, "Private information retrieval for secure distributed storage systems," *IEEE Trans. Inf. Forensics Security*, vol. 13, pp. 2953–2964, 2018.
[44] Z. Jia, H. Sun, and S. A. Jafar, "Cross subspace alignment and the asymptotic capacity of X-secure T-private information retrieval," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5783–5798, Sep. 2019.
[45] Z. Jia, H. Sun, and S. A. Jafar, "Cross subspace alignment and the asymptotic capacity of X-secure T-private information retrieval," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5783–5798, Sep. 2019.
[46] Z. Jia and S. A. Jafar, "X-secure T-private information retrieval from MDS coded storage with Byzantine and unresponsive servers," *IEEE Tran. Inf. Theory*, vol. 66, no. 12, pp. 7427–7438, Dec. 2020.
[47] M. A. Attia, D. Kumar, and R. Tandon, "The capacity of private information retrieval from uncoded storage constrained databases," *IEEE Trans. Inf. Theory*, vol. 66, no. 11, pp. 6617–6634, Nov. 2020.
[48] Y.-P. Wei, B. Arasli, K. Banawan, and S. Ulukus, "The capacity of private information retrieval from decentralized uncoded caching databases," *Information*, vol. 10, p. 372, Dec. 2019.
[49] K. Banawan, B. Arasli, Y.-P. Wei, and S. Ulukus, "The capacity of private information retrieval from heterogeneous uncoded caching databases," *IEEE Trans. Inf. Theory*, vol. 66, no. 6, pp. 3407–3416, Jun. 2020.
[50] K. Banawan, B. Arasli, and S. Ulukus, "Improved storage for efficient private information retrieval," in *Proc. IEEE ITW*, Aug. 2019, pp. 1–5.
[51] C. Tian, "On the storage cost of private information retrieval," *IEEE Trans. Inf. Theory*, vol. 66, no. 12, pp. 7539–7549, Dec. 2020.
[52] K. Banawan and S. Ulukus, "Private information retrieval from non-replicated databases," in *Proc. IEEE ISIT*, Jul. 2019, pp. 1272–1276.
[53] N. Raviv, I. Tamo, and E. Yaakobi, "Private information retrieval in graph-based replication systems," *IEEE Trans. Inf. Theory*, vol. 66, no. 6, pp. 3590–3602, Jun. 2020.
[54] K. Banawan and S. Ulukus, "Asymmetry hurts: Private information retrieval under asymmetric traffic constraints," *IEEE Trans. Inf. Theory*, vol. 65, no. 11, pp. 7628–7645, Nov. 2019.
[55] H. Sun and S. A. Jafar, "The capacity of private computation," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3880–3897, Jun. 2019.
[56] K. Banawan and S. Ulukus, "Noisy private information retrieval: On separability of channel coding and information retrieval," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8232–8249, Dec. 2019.
[57] Z. Chen, Z. Wang, and S. A. Jafar, "The asymptotic capacity of private search," *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 4709–4721, Aug. 2020.
[58] S. Vithana, K. Banawan, and S. Ulukus, "Semantic private information retrieval," *IEEE Trans. Inf. Theory*, vol. 68, no. 4, pp. 2635–2652, Apr. 2022.
[59] M. Naor and B. Pinkas, "Oblivious transfer and polynomial evaluation," in *Proc. 31st Annu. ACM Symp. Theory Comput.*, 1999, pp. 245–254.

Zhusheng Wang (Student Member, IEEE) received the B.Sc. degree in information and communication engineering from Zhejiang University, Hangzhou, China, in 2016, and the M.Sc. degree in electrical and computer engineering from the University of Michigan, Ann Arbor, MI, USA, in 2018. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA. His research interests include information theory, private information retrieval, and machine learning.

Sennur Ulukus (Fellow, IEEE) received the B.S. and M.S. degrees in electrical and electronics engineering from Bilkent University and the Ph.D. degree in electrical and computer engineering from Wireless Information Network Laboratory, Rutgers University.

She is the Anthony Ephremides Professor of Information Sciences and Systems with the Department of Electrical and Computer Engineering, University of Maryland, College Park, where she also holds a joint appointment with the Institute for Systems Research (ISR). Prior to joining UMD, she was a Senior Technical Staff Member with AT&T Labs-Research. She is a Distinguished Scholar-Teacher with the University of Maryland. Her research interests are in information theory, wireless communications, machine learning, signal processing, and networks; with recent focus on private information retrieval, age of information, machine learning for wireless, distributed coded computing, group testing, physical-layer security, energy harvesting communications, and wireless energy and information transfer.

Dr. Ulukus received the 2003 IEEE Marconi Prize Paper Award in Wireless Communications, the 2019 IEEE Communications Society Best Tutorial Paper Award, the 2020 IEEE Communications Society Women in Communications Engineering Outstanding Achievement Award, the 2020 IEEE Communications Society Technical Committee on Green Communications and Computing Distinguished Technical Achievement Recognition Award, the 2005 NSF CAREER Award, the 2011 ISR Outstanding Systems Engineering Faculty Award, and the 2012 ECE George Corcoran Outstanding Teaching Award. She has been an Area Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS since 2019 and a Senior Editor for the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING since 2020. She was an Area Editor for the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING from 2016 to 2020, an Editor for the IEEE Journal on Selected Areas in Communications-Series on Green Communications and Networking from 2015 to 2016, an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2007 to 2010, and an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS from 2003 to 2007. She was a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS in 2008, 2015, and 2021, *Journal of Communications and Networks* in 2012, and the IEEE TRANSACTIONS ON INFORMATION THEORY in 2011. She is the TPC Chair of 2021 IEEE Globecom, and was a TPC Co-Chair of 2019 IEEE ITW, 2017 IEEE ISIT, 2016 IEEE Globecom, 2014 IEEE PIMRC, and 2011 IEEE CTW. She was a Distinguished Lecturer of the IEEE Information Theory Society form 2018 to 2019.