Asymmetric Leaky Private Information Retrieval

Islam Samy   Mohamed Attia   Ravi Tandon   Loukas Lazos

Department of Electrical and Computer Engineering
University of Arizona, Tucson, AZ, USA.
Email: {*islamsamy, madel, tandonr, llazos*}@email.arizona.edu

Abstract—Information-theoretic formulations of the private information retrieval (PIR) problem have been investigated under a variety of scenarios. Symmetric private information retrieval (SPIR) is a variant where a user is able to privately retrieve one out of K messages from N non-colluding replicated databases without learning anything about the remaining $K-1$ messages. However, the goal of perfect privacy can be too taxing for certain applications. In this paper, we investigate if the information-theoretic capacity of SPIR (equivalently, the inverse of the minimum download cost) can be increased by relaxing both user and DB privacy definitions. Such relaxation is relevant in applications where privacy can be traded for communication efficiency.

We introduce and investigate the Asymmetric Leaky PIR (AL-PIR) model with different privacy leakage budgets in each direction. For user privacy leakage, we bound the probability ratios between all possible realizations of DB queries by a function of a non-negative constant ϵ. For DB privacy, we bound the mutual information between the undesired messages, the queries, and the answers, by a function of a non-negative constant δ. We propose a general AL-PIR scheme that achieves an upper bound on the optimal download cost for arbitrary ϵ and δ. We show that the optimal download cost of AL-PIR is upper-bounded as $D^*(\epsilon,\delta) \leq 1 + \frac{1}{N-1} - \frac{\delta e^\epsilon}{N^{K-1}-1}$. Second, we obtain an information-theoretic lower bound on the download cost as $D^*(\epsilon,\delta) \geq 1 + \frac{1}{Ne^\epsilon-1} - \frac{\delta}{(Ne^\epsilon)^{K-1}-1}$. The gap analysis between the two bounds shows that our AL-PIR scheme is optimal when $\epsilon = 0$, i.e., under perfect user privacy and it is optimal within a maximum multiplicative gap of $\frac{N-e^{-\epsilon}}{N-1}$ for any $\epsilon > 0$ and $\delta > 0$.

I. INTRODUCTION

In the era of big data and data analytics, users who access a plethora of online services face serious privacy risks. Their online behavior and data access patterns can be analyzed to reveal sensitive personal information and breach their privacy [1]. One possible solution to such data leakages is to retrieve information privately by executing a private information retrieval (PIR) protocol. In a PIR protocol, the identity of the message retrieved by the user remains secret from the database(s). This is typically achieved at the expense of an increased communication cost to ensure that the desired message remains hidden among others. In the pioneering work by Chor *et al.* [1], the authors considered one-bit long messages. The overhead was calculated as the sum of the queries sent

by the user (upload cost) and the answers provided by the database (download cost). Under arbitrarily large messages, the download cost becomes the dominant factor of the PIR overhead. This allows the PIR rate to be defined as the ratio of the message size to the number of downloaded bits. The maximum of these rates is referred to as the PIR capacity and its reciprocal as the download cost.

Since the introduction of the PIR problem in [1], an extensive body of works have investigated efficient PIR schemes that yield either computational [2]–[5] or information-theoretic privacy guarantees [6]–[42]. The former achieves privacy assuming that the DBs are computationally-bounded. Essentially, privacy is preserved due to the intractability of computationally-hard problems. In information-theoretic PIR, the DBs are assumed to be computationally unbounded, thus achieving a higher level of assurance. Perfect privacy is guaranteed if the queries do not reveal any information about the desired message (privacy) and the answers are sufficient to recover it (decodability). An intuitive PIR solution is to download all K messages from a database. In fact, this is the only way to guarantee perfect privacy in the single database case. However, privacy comes at an impractical communication overhead.

Review of Recent Progress on Information-Theoretic PIR: A practical way to increase the PIR capacity is to consider a distributed storage system (DSS) of N databases. Shah. *et al.* [6] proposed a PIR scheme that achieves a rate of $1 - \frac{1}{N}$ when K messages are replicated across N non-colluding databases. Later, Sun and Jafar [7] characterized the PIR capacity for any N and K as $(1 + 1/N + 1/N^2 + \cdots + 1/N^{K-1})^{-1}$. The original scheme introduced in [7] achieves capacity when the message size L is allowed to grow as a function of N and K. Subsequently, they characterized the PIR capacity for a fixed message size [8]. Since the appearance of the fundamental result of Sun and Jafar [7], numerous important and practically relevant variations of PIR have been considered.

Multi-round PIR allows multiple rounds of communication between the user and databases. While interaction does not increase capacity, it can reduce the storage overhead at each database [10]. Sun and Jafar [11] considered the robust PIR problem where $M - N$ out of a total of $M > N$ databases fail to respond to user queries. Additionally, they characterized the capacity when $T < N$ databases collude and share the received queries. Tajeddine *et al.* [12] considered MDS-PIR for coded databases where each message is separately coded using an (N, M) MDS code. Banawan and Ulukus [13] derived

the coded PIR capacity for arbitrary N, M, and K. Wang and Skogland [14] showed that the PIR capacity remains $1 - \frac{M}{N}$ even if each message is coded. In [15] and [16], the scenario of N MDS-coded databases with T colluding ones was presented. However, the capacity of this case is still an open problem (for other variants of MDS-PIR, see [17]–[20]). In [21] and [22], the case of multi-message PIR, where the user can use one query to request more than one messages, was investigated. Banawan and Ulukus [23] characterized the PIR capacity with Byzantine databases where any subset of databases can be adversarial and respond untruthfully. In [24], Tajeddine *et al.* studied the same model but in the presence of colluding databases. Banawan and Ulukus [25] studied PIR through a wiretap channel, where an eavesdropper tries to decode the content sent through the channel. Other variants of PIR in the presence of eavesdroppers are studied in [26], [27].

The problem of PIR was also studied when the user has a cache or side-information, which can be useful in increasing PIR capacity [28]–[32]. PIR from storage-constrained databases was studied in [33], [43], [44], where capacity was characterized under the assumption of uncoded storage across databases. Recently, Tian *et al.* [34] proposed a new capacity-achieving scheme with an optimal message size of $N - 1$ and a minimum upload cost. Other lines of work considered different privacy requirements from the original PIR model in [7]. The problem of symmetric PIR (SPIR) was studied in [9], where the user must be able to retrieve the message of interest privately (user privacy), while at the same time the databases must avoid any information leakage about the remaining $K-1$ messages (DB privacy). The SPIR optimal download cost was characterized as $\frac{N}{N-1}$ with common randomness at least $\alpha = \frac{1}{N-1}$ bits per desired message bits. Latent-variable PIR was considered recently in [45], where privacy is required for a latent variable describing a predefined user attribute. Additional interesting variants of PIR can be found in [35]–[42].

The novel coding schemes and fundamental ideas developed in the above works have also helped in advancing other problems beyond PIR. For instance, an interesting connection between blind interference alignment (BIA) and PIR was studied in [46] showing that a good BIA scheme translates to a good PIR protocol. Secure and private distributed matrix multiplication has been considered in [47]–[51] addressing the problem of computing a product of two matrices with some constraints on the identity of the product matrices and/or the information content in the matrices. Jia and Jafar [17] showed the connection of the secure and private distributed matrix multiplication to one variation of the MDS-PIR problem. Recently, the problem of private set intersection (PSI) was studied in [52] from a PIR perspective and capacity results were obtained.

Relaxing Privacy Metrics for PIR: The above works have all focused on perfect privacy, either for the user (as in PIR), or for both the user and the DBs (as in SPIR). The perfect privacy requirement usually comes at the expense of high download cost and does not allow tuning the PIR efficiency and privacy according to the application requirements. In scenarios of frequent message retrieval, trading user or DB privacy for communication efficiency could be desirable. Ideally, one would select a desired leakage level and then design a leakage-constrained PIR scheme that guarantees such privacy while maximizing the PIR capacity.

A few previous works have introduced privacy definitions that relax the notion of perfect privacy. Asonov *et al.* replaced privacy with the concept of repudiation [53]. The repudiation property is achieved if some uncertainty remains about the desired message. However, this metric does not provide any information-theoretic privacy guarantees, as repudiation is satisfied even if the retrieved message can be identified with almost certainty. Recently, Toledo *et al.* [54] adopted a game-based differential privacy definition to increase the PIR capacity at the expense of bounded privacy loss. However, their privacy definition only captures the privacy of the submitted queries. The authors propose several schemes that hide the query identity and study their cost. Although the query privacy can be thought of as functional equivalent to information-theoretic PIR in some cases, it does not satisfy the perfect privacy definition.

In our prior work, we introduced the Leaky PIR (L-PIR) where a bounded amount of leakage is allowed about the message identity [55]. We adopted a concept similar to differential privacy to bound the leakage as a function of a non-negative constant ϵ. The leakage in privacy is achieved by constructing multiple biased "retrieval paths" across databases where each path realizes one query per database. Lin *et al.* [56], [57] relaxed user privacy by allowing bounded mutual information between the queries and the corresponding requested message index. Unlike [56], [57], which deal with the average leakage (measured by mutual information), the L-PIR model in [55] satisfies the privacy leakage constraints strictly for all possible query/message index combinations, and thus provides stronger privacy guarantees.

In another recent work, Guo *et al.* [58] considered the problem of SPIR with perfect user privacy and relaxed DB privacy. DB privacy was relaxed by allowing a bounded mutual information (no more than δ) between the undesired messages, the queries, and the answers received by the user. Similar to the original work on SPIR in [9], SPIR with relaxed DB privacy in [58] requires sharing common randomness among DBs and comes at the expense of a loss in the PIR capacity.

Summary of contributions– We investigate a three-way tradeoff between user privacy, DB privacy, and the communication efficiency of PIR. We study the problem of *Asymmetric Leaky PIR* (AL-PIR) where some information about the identity of the desired message is allowed to leak to the DBs, and some information about the undesired messages is allowed to leak to the user. The goal is to trade privacy in both directions for achieving gains in PIR capacity, thus making PIR more communication-efficient. For user privacy, we adopt the metric introduced in our prior work [55], where the privacy bound is determined as a function of a non-negative constant ϵ. For bounding DB privacy, we adopt a mutual information-based leakage metric to be bounded by a non-negative constant δ. We next summarize the main contributions:

- We propose an AL-PIR scheme that satisfies the leakage

budgets in both directions for arbitrary values of (ϵ, δ), an arbitrary number of K messages, and an arbitrary number of N databases. The achievable download cost of this scheme is given by $D(\epsilon, \delta) = 1 + \frac{1}{N-1} - \frac{\delta e^{\epsilon}}{N^{K-1}-1}$. This cost also represents an upper bound on the optimal download cost (lower bound on the capacity) of the AL-PIR. We use an alternate perfect privacy PIR scheme that follows a *path-based approach*, where a user's query is equivalent to selecting one of several possible paths across databases. A path is defined as a set of queries, one per database, that achieves decodability, however different paths incur different download costs. We leverage this cost imbalance to introduce leakage through the idea of biasing the path selection probabilities. A path giving a lower download cost can be used more frequently compared to higher download cost paths. This biasing introduces user privacy leakage. The path selection probabilities are chosen to minimize the download cost while satisfying the privacy budget, measured by ϵ. To achieve DB privacy, our scheme requires sharing common randomness among databases. We combine the path-based approach with the ideas of the scheme presented in [58] to arrive at our general AL-PIR scheme. In particular, achieving a DB privacy leakage of no more than δL bits, requires common randomness given by $\left(\frac{1}{N-1} - \frac{e^{\epsilon}+N^{K-1}-1}{N^{K-1}-1}\delta\right)L$ bits, which represents an upper bound on the optimal common randomness size.

- We present a converse proof to obtain a lower bound on the optimum download cost (upper bound on capacity). This bound is characterized by $D^*(\epsilon, \delta) \geq 1 + \frac{1}{Ne^{\epsilon}-1} - \frac{\delta}{(Ne^{\epsilon})^{K-1}-1}$. The upper and lower bounds are shown to match each other at extreme values of epsilon ($\epsilon = 0$; $\epsilon \to \infty$) and for any δ. Moreover, we show through gap analysis that our upper and lower bounds are within a maximum multiplicative gap of $\frac{N-e^{-\epsilon}}{N-1}$ for any $\epsilon > 0$ and $\delta \geq 0$.
- We derive a lower bound on the optimal required common randomness at the databases. This bound characterizes that achieving a DB privacy leakage of no more than δL bits, requires shared randomness of size no less than $\left(\frac{1}{Ne^{\epsilon}-1} - \frac{(Ne^{\epsilon})^{K-1}}{(Ne^{\epsilon})^{K-1}-1}\delta\right)L$ bits.
- We investigate the tradeoffs variations in both sides of leakage as special cases of our general (ϵ, δ) AL-PIR scheme. In particular, we show a three-way tradeoff between download cost, user privacy, and DB privacy, such that enhancing one of them would be at the expense of the other two. We also show matching results for the following special cases for our derived bounds on the AL-PIR model: a) perfect user privacy (original PIR) [7], b) perfect user and DB privacy (SPIR) [9], c) Leaky user privacy (L-PIR) [55], and d) perfect user privacy and leaky DB privacy [58].

II. SYSTEM MODEL: ASYMMETRIC LEAKY PIR

We study the PIR problem illustrated in Figure 1. We consider N databases $\mathrm{DB}_1, \mathrm{DB}_2, \ldots, \mathrm{DB}_N$ and K independent
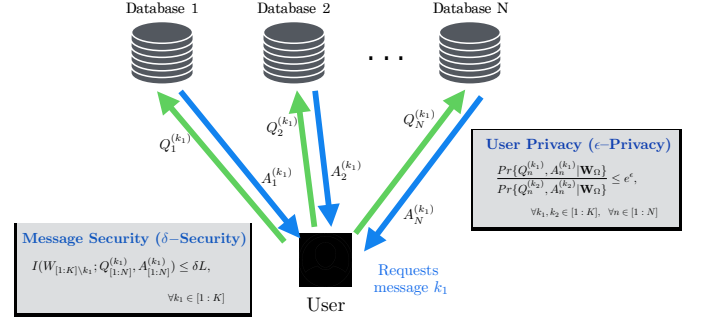

Fig. 1. Asymmetric leaky private information retrieval (AL-PIR) problem.

messages $W_1, W_2, \ldots, W_K$, each of size L bits, such that

$$H(W_1, W_2, \cdots, W_K) = \sum_{k=1}^{K} H(W_k), \qquad (1)$$

$$H(W_1) = H(W_2) = \cdots = H(W_K) = L. \qquad (2)$$

A user interested in privately retrieving W_k, $k \in [1 : K]$[1] sends N separate queries $Q_1^{(k)}, \cdots, Q_N^{(k)}$ to each of the N DBs, where $Q_n^{(k)}$ denotes the query sent to the nth database (DB_n), $n \in [1 : N]$, when retrieving message W_k. The N DBs are assumed to be replicated and non-colluding, i.e., they store all the K messages and they do not share the queries received from the user. We also assume the DBs are interested in achieving privacy, i.e., the user must only decode the requested message subject to a leakage constraint. To achieve DB privacy, the N DBs are allowed to share common randomness denoted by a random variable S of size αL bits, i.e., $H(S) = \alpha L$. Moreover, S is not known to the user.

Upon receiving $Q_n^{(k)}$, the nth database generates the corresponding answer $A_n^{(k)}$ as a deterministic function of the query $Q_n^{(k)}$, the K messages, and the shared common randomness S, i.e.,

$$H\left(A_n^{(k)}|Q_n^{(k)}, W_1, \ldots, W_K, S\right) = 0. \qquad (3)$$

The user must be able to decode the desired message W_k upon receiving the answers from the N databases. Formally, the AL-PIR scheme must satisfy the following correctness, user privacy, and DB privacy constraints.

Correctness: Given queries $Q_{[1:N]}^{(k)} \triangleq \{Q_1^{(k)}, \cdots, Q_N^{(k)}\}$, the user must be able to decode the desired message W_k, with probability of error P_e, by collecting the corresponding answers $A_{[1:N]}^{(k)} \triangleq \{A_1^{(k)}, \cdots, A_N^{(k)}\}$ from the N DBs, i.e.,

$$H\left(W_k|Q_{[1:N]}^{(k)}, A_{[1:N]}^{(k)}\right) = o(L)L, \qquad (4)$$

where $o(L)$ is any function that approaches zero as $L \to \infty$. $o(L)$ is set to zero if P_e is required to be exactly zero.

$\delta-$**DB privacy:** In the original SPIR formulation [9], the authors assume no leakage to the user about the undesired messages. For a desired message W_k, perfect DB privacy is

---

[1]**Notation:** Through this work, we use the notation $[1 : X]$ to represent the set of integers from 1 to X.

satisfied if

$$I\left(\mathbf{W}_{[1:K]\backslash k}; Q_{[1:N]}^{(k)}, A_{[1:N]}^{(k)}\right) = 0, \quad \forall k \in [1:K], \quad (5)$$

where $\mathbf{W}_{[1:K]\backslash k} \triangleq (W_1, \ldots, W_{k-1}, W_{k+1}, \ldots, W_K)$ is the set of all messages except W_k. In this work, we relax this condition by assuming a general leaky DB privacy constraint. The leaked information about the undesired messages must be bounded as,

$$I(\mathbf{W}_{[1:K]\backslash k}; Q_{[1:N]}^{(k)}, A_{[1:N]}^{(k)}) \leq \delta L, \quad \forall k \in [1:K], \quad (6)$$

where $\delta \geq 0$ is a non-negative constant.

$\epsilon-$**user privacy:** Under perfect user privacy, the privacy constraints are expressed as,

$$(A_n^{(k_1)}, Q_n^{(k_1)}, W_1, \cdots, W_K) \sim (A_n^{(k_2)}, Q_n^{(k_2)}, W_1, \cdots, W_K), \quad (7)$$

$\forall k_1, k_2 \in [1:K]$. This guarantees that the submitted queries are always independent of the message index. The previous constraint can be alternatively expressed as,

$$(A_n^{(k_1)}, Q_n^{(k_1)}|\mathbf{W}_\Omega) \sim (A_n^{(k_2)}, Q_n^{(k_2)}|\mathbf{W}_\Omega), \quad \forall k_1, k_2 \in [1:K], \quad (8)$$

where $\mathbf{W}_\Omega$ is any subset of the K messages, i.e., $\mathbf{W}_\Omega \subseteq \{W_1, \ldots, W_K\}$. In this work, the privacy constraint is relaxed such that given any subset $\mathbf{W}_\Omega$ of the K messages, the following likelihood ratio is bounded as follows:

$$\frac{Pr\{Q_n^{(k_1)} = \pi, A_n^{(k_1)} = \gamma|\mathbf{W}_\Omega\}}{Pr\{Q_n^{(k_2)} = \pi, A_n^{(k_2)} = \gamma|\mathbf{W}_\Omega\}} \leq e^\epsilon, \quad \forall k_1, k_2 \in [1:K], \quad (9)$$

$\forall n \in [1:N]$, where π and γ represent any possible realizations for the queries and answers, respectively and ϵ is a non-negative constant. Unlike perfect user privacy constraint which ensures that queries and answers are independent of the message index, the leaky privacy definition allows some queries and answers to be used more frequently when certain messages are retrieved. By setting $\epsilon = 0$, the $\epsilon-$user privacy definition in (9) becomes equivalent to the perfect privacy constraint in (8).

Although there is a similarity between the definition of the user privacy and the definition of differential privacy (DP), we emphasize that this similarity is just in the mathematical formulation. The main conceptual difference between our privacy definition and the DP definition is that in our privacy model, we do not have the concept of neighboring datasets (that differ in one element). Instead, our model considers a fixed input dataset for each query. Actually, we note that our definition bears more similarity to local differential privacy (LDP). In local differential privacy, a user holding private data X wishes to reveal Y through a private mechanism. The mechanism satisfies ϵ-LDP if $P(Y = y|X = x)/P(Y = y|X = x') \leq e^\epsilon$ for all realizations x, x' of the private data. The analogy of the $\epsilon-$user privacy with ϵ-LDP is then clear since here, the private data refers to the index of the desired message, and the output of the mechanism corresponds to the query sent by the user to the DB. One thing that is different for the

$\epsilon-$user privacy is that in a DP setting, the information privacy is guaranteed against an adversary who designs algorithms for data processing. In the user privacy definition, the side that represents the adversary is the database. Whereas, the user (the side that seeks privacy) is the one that can exclusively design the queries. Generally, for the $\epsilon-$user privacy, there are potential challenges related to the accumulation of leakage across multiple queries. An adversary that monitors multiple queries can combine the outputs to obtain extra information. To avoid such a scenario, the parameter ϵ could be adjusted to ensure that the accumulated leakage across multiple queries is always within the allowed limits.

Guarantees and applications of the leaky privacy definitions: The leaky definition in (9) guarantees that the difference between the distributions of query and answers for any two indices and any given realization is always bounded. We highlight that this is guaranteed for every message index pair and not on average. From an application perspective, the leaky definition guarantees that the database has a bounded success probability in differentiating the requested message index from any other index, and for any realization that satisfies the privacy definition. Moreover, this bounded success probability can be controlled by tuning ϵ.

For database privacy, from an application perspective, the leaky SPIR metric in (6) guarantees that the user will not decode more than δL bits, on average, from undesired messages. Note that this definition does not impose any restrictions on how the leaked bits are distributed. Depending on the scheme, the leaked bits could belong to one or multiple undesired messages and could have a different distribution per realization. This guarantee is only on average.

Our privacy definitions could be helpful in applications where privacy is required, but guaranteeing perfect privacy makes PIR impractical from a communication overhead standpoint. To provide a concrete example, consider a multimedia streaming application. From the user perspective, allowing the database to determine with bounded probability, the multimedia content (e.g., movie) that the user retrieves is a significant privacy improvement from the current landscape. This information is not as sensitive as other information types such as patient records. Moreover, the message size is prohibitively large to allow for perfect privacy solutions. Generally, the parameter ϵ should be tuned based on the sensitivity of the retrieved content. It can be even set to zero (this means perfect privacy) for applications that do not permit any leakage. From the database perspective, allowing the user to access some additional multimedia content with each query (which for most schemes would be partial content) does not degrade privacy in a meaningful way.

Other leaky user privacy definitions: To relax DB privacy, we adopt the mutual information metric in [58]. On the other hand, we use the probability metric we introduced in [55] to bound the leakage of user privacy. The latter metric strictly satisfies the privacy constraint for all possible query/message index combinations. We note that there are other weaker metrics one can use for relaxing user privacy. In [56], Lin *et al.* proposed a metric η that gives a bound on the average

privacy leakage over all databases for a desired message index given by a random variable $\theta \in [1 : K]$ such that,

$$\frac{1}{N} \sum_{n=1}^{N} I(\theta; Q_n^{(\theta)}) \leq \eta. \tag{10}$$

Jia *et al.* introduced the following privacy constraint [59],

$$H(A_n^{(k+1)}|W_1, \ldots, W_k) - H(A_n^{(k)}|W_1, \ldots, W_k) = \rho L,$$
$$\forall k \in [1 : K], \quad \forall n \in [1 : N], \tag{11}$$

where parameter ρ controls the leakage budget, with $0 \leq \rho \leq \frac{1}{N}$. In contrast to our ϵ-user privacy definition in (9), both of the metrics provide average privacy guarantees, i.e., they bound the average privacy leakage over all possible retrieval schemes. This means that the privacy leakage is allowed to exceed the required bound in the case of individual message retrievals. In this work, we extend the definition in (7) to investigate the scenario when the distribution of the sent queries and the corresponding answers is allowed to depend on the requested message index within predefined limits. Also, the AL-PIR model satisfies the $\epsilon-$user privacy definition strictly over all possible realizations of answers and queries. This ensures that leakage is always within the allowed budget ϵ for all individual message retrievals.

We highlight that there are more stricter privacy metrics that can be employed, such as min-entropy which is a stronger notion compared to bounded mutual information. However, we believe that this will require new tools to obtain lower bounds on the download cost and constructing schemes with bounded min-entropy.

Communication Cost: To evaluate the performance of the AL-PIR scheme, we adopt the Shannon theoretic formulation where the message size is assumed to be arbitrarily long and therefore, the upload cost is negligible compared to the download cost [7]. In this case, the AL-PIR rate is the reciprocal of the download cost $D(\epsilon, \delta)$, which characterizes the total information bits the user has to download to retrieve one desired message bit. Let $D_{\epsilon,\delta}$ be the total number of downloaded bits to retrieve message W_k, for some ϵ and δ, and L be the size of the desired message. The normalized download cost is given by,

$$D(\epsilon, \delta) = \frac{D_{\epsilon,\delta}}{L} = \frac{\sum_n H(A_n^{(k)})}{H(W_k)}. \tag{12}$$

We say that the pair $(L, D_{\epsilon,\delta})$ is achievable if there exists an AL-PIR scheme that satisfies the correctness, DB privacy, and user privacy conditions in (4), (6), and (9), respectively, and can retrieve a message of size L bits by downloading a total of $D_{\epsilon,\delta}$ bits. Our goal is to find the optimal download cost $D^*(\epsilon, \delta)$ such that

$$D^*(\epsilon, \delta) = \min\{D_{\epsilon,\delta}/L : (L, D_{\epsilon,\delta}) \text{ is achievable}\}. \tag{13}$$

The capacity of the AL-PIR $C^*(\epsilon, \delta)$ is the reciprocal of $D^*(\epsilon, \delta)$,

$$C^*(\epsilon, \delta) = \max\{L/D_{\epsilon,\delta} : (L, D_{\epsilon,\delta}) \text{ is achievable}\}. \tag{14}$$

Optimal common randomness size: We are also interested in characterizing the fundamental limits of common randomness S needed to be stored at the databases. In general, the common randomness size α is a function of the privacy budget parameters (ϵ, δ). Therefore, in the following discussion, we use the notation $H(S) = \alpha(\epsilon, \delta)L$. We define $\alpha^*(\epsilon, \delta)$ as the minimum common randomness size that satisfies the correctness, DB privacy, and user privacy conditions in (4), (6), and (9), respectively, i.e.,

$$\alpha^*(\epsilon, \delta) = \min\{\alpha(\epsilon, \delta) : (4), (6), \text{ and } (9) \text{ are satisfied}\}. \tag{15}$$

III. MAIN RESULTS AND DISCUSSION

In this section, we present our main results on the optimal download cost and the required amount of shared randomness for AL-PIR. Given desired privacy budgets ϵ and δ for the user and DB privacy leakage, respectively, we state our main results in the following Theorems.

Theorem 1: Define $d_1(\epsilon, \delta) := 1 + \frac{1}{N-1} - \frac{\delta e^\epsilon}{N^{K-1}-1}$. For $N \geq 2$ and shared randomness S with size $H(S) \geq \alpha_1(\epsilon, \delta)L$, where

$$\alpha_1(\epsilon, \delta) = \begin{cases} \frac{1}{N-1} - \frac{e^\epsilon + N^{K-1} - 1}{N^{K-1} - 1} \delta, & 0 \leq \delta < \delta_1(\epsilon), \\ 0, & \delta > \delta_1(\epsilon), \end{cases} \tag{16}$$

the optimal download cost of AL-PIR, satisfying both the $\epsilon-$user privacy and $\delta-$DB privacy definitions, is upper-bounded by

$$D^*(\epsilon, \delta) \leq D^{\text{UB}}(\epsilon, \delta) = \begin{cases} d_1(\epsilon, \delta), & 0 \leq \delta < \delta_1(\epsilon), \\ d_1(\epsilon, \delta_1(\epsilon)), & \delta \geq \delta_1(\epsilon). \end{cases} \tag{17}$$

In (16) and (17), $\delta_1(\epsilon)$ is the maximum DB privacy leakage (when no common randomness is required, i.e., $\alpha_1(\epsilon, \delta) = 0$) which is a function of the allowed user privacy leakage ϵ, and is given by,

$$\delta_1(\epsilon) = \frac{N^{K-1} - 1}{(N-1)(e^\epsilon + N^{K-1} - 1)}. \tag{18}$$

The proof of Theorem 1 is presented in Section IV. As a result of Theorem 1, we have the following remark.

Remark 1: The required size of shared randomness for our achievability scheme, as given by $\alpha_1(\epsilon, \delta)$ in (16), yields an upper bound on the optimal size of minimum shared randomness $\alpha^*(\epsilon, \delta)$ as defined in (15), i.e., $\alpha^*(\epsilon, \delta) \leq \alpha_1(\epsilon, \delta)$. Moreover, $\alpha_1(\epsilon, \delta)$ is also sufficient to satisfy (ϵ', δ') privacy constraints, such that $\epsilon' \geq \epsilon$ and $\delta' \geq \delta$. In other words, if a given amount of common randomness is sufficient to satisfy (ϵ, δ) privacy, then it is also sufficient if the privacy budgets are increased.

In Figure 2, we show the effect of ϵ and δ on the download cost for the case when $N = K = 2$. We can observe the following: a) the download cost is a monotonically decreasing function of the privacy budgets ϵ and δ; b) as ϵ approaches infinity, which corresponds to no user privacy, the achieved download cost approaches 1; c) for $\epsilon = 0$ (perfect user privacy) and as δ approaches zero (perfect DB privacy), the achieved download cost is 2 which matches the case of SPIR studied in [9] where the optimal download cost is $\frac{N}{N-1} = 2$; and d) for
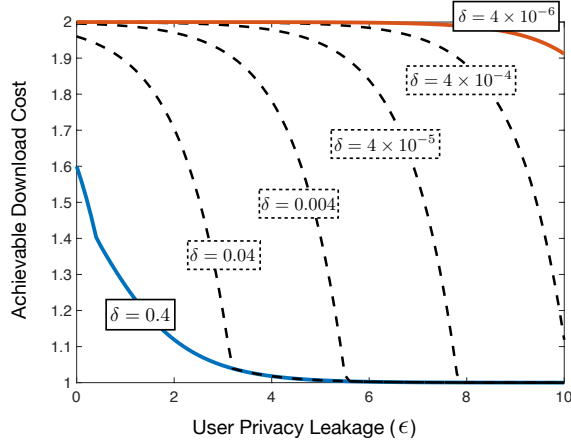
Fig. 2. The achievable download cost for our AL-PIR scheme when $N = K = 2$ as a function of ϵ for different values of δ.

$\delta \geq \delta_1(\epsilon) = 1/(e^\epsilon + 1)$ (or $\epsilon \geq \ln(1/\delta - 1)$), the download cost is only a function of ϵ (the line corresponding to $\delta = 0.4$).

Theorem 2: Define $d_2(\epsilon, \delta) := 1 + \frac{1}{Ne^\epsilon - 1} - \frac{\delta}{(Ne^\epsilon)^{K-1} - 1}$. For $N \geq 2$, and shared randomness S with size $H(S) \geq \alpha_2(\epsilon, \delta)L$, where

$$\alpha_2(\epsilon, \delta) = \begin{cases} \frac{1}{Ne^\epsilon - 1} - \frac{(Ne^\epsilon)^{K-1}}{(Ne^\epsilon)^{K-1} - 1} \delta, & 0 \leq \delta < \delta_2(\epsilon), \\ 0, & \delta > \delta_2(\epsilon), \end{cases}$$
(19)

the optimal download cost of AL-PIR subject to $\epsilon-$user privacy and $\delta-$DB privacy is lower-bounded by

$$D^*(\epsilon, \delta) \geq D^{LB}(\epsilon, \delta) = \begin{cases} d_2(\epsilon, \delta), & 0 \leq \delta < \delta_2(\epsilon), \\ d_2(\epsilon, \delta_2(\epsilon)), & \delta \geq \delta_2(\epsilon), \end{cases}$$
(20)

where

$$\delta_2(\epsilon) = \frac{(Ne^\epsilon)^{K-1} - 1}{(Ne^\epsilon - 1)(Ne^\epsilon)^{K-1}}.$$
(21)

Furthermore, the optimal size of common randomness satisfying $\epsilon-$user privacy and $\delta-$DB privacy is lower-bounded by $\alpha^*(\epsilon, \delta) \geq \alpha_2(\epsilon, \delta)$.

The proof of Theorem 2 is presented in Section V. We note that the results in Theorems 1 and 2 hold for $N \geq 2$ DBs. In the following proposition, we characterize the capacity for the case of one database.

Proposition 3: The optimal download cost $D^*(\epsilon, \delta)$ for $N = 1$ and for any $0 \leq \epsilon < \infty$ is given by:

$$D^*(\epsilon, \delta) = \begin{cases} \infty, & \delta < (K - 1), \\ K, & \delta = (K - 1). \end{cases}$$
(22)

The above result shows that the problem of AL-PIR for one database is degenerate. In particular, to satisfy the ϵ-user privacy constraint, any query/answer pair has to be requested to retrieve each of the K messages with non-zero probability. Since $N = 1$, the only solution is to download all messages, i.e., a download cost of K. However, upon downloading all K messages, the leakage about the remaining $(K - 1)$ messages

is fixed and given by $\delta = K - 1$. Hence, if the DB privacy budget is $\delta < (K - 1)$, the AL-PIR problem is infeasible and the capacity is 0, i.e., $D^*(\epsilon, \delta < K - 1) = \infty$. We prove Proposition 3 in Appendix D.

In the next Corollary, we show that our proposed scheme in Theorem 1 is information-theoretically optimal for perfect user privacy, i.e., $\epsilon = 0$, and is optimal within a maximum multiplicative gap ratio of $\frac{N - e^{-\epsilon}}{N - 1}$ for any (ϵ, δ). The proof of the corollary is presented in Appendix A.

Corollary 1: The multiplicative gap ratio between the upper and lower bounds on the download cost of the AL-PIR, given by Theorems 1 and 2, respectively, is bounded as follows:

$$\frac{D^{UB}(\epsilon, \delta)}{D^{LB}(\epsilon, \delta)} \leq \frac{N - e^{-\epsilon}}{N - 1}.$$
(23)

In Figure 3, we show the upper and lower bounds on the download cost of the AL-PIR and the numerical multiplicative gap ratio, as a function of system parameters (N, K, ϵ, δ). Specifically, in Figure 3a, we set the allowed DB privacy leakage to the maximum leakage, i.e., $\delta > \max(\delta_1(\epsilon), \delta_2(\epsilon))$ as defined in (18) (no shared randomness required for this case). This gives the results of the L-PIR model considered in [55]. As the number of messages increases, both upper and lower bounds increase, whereas both decrease with N. This happens as increasing N increases the number of bits that can be utilized as a side information to retrieve the desired message. On the other hand, increasing K adds an overhead on any retrieval scheme to satisfy the privacy by considering the symmetry among downloaded bits from different messages. We observe a similar trend for the multiplicative gap ratio as well. In Figure 3b, we fix the value of the DB privacy leakage to $\delta = 4 \times 10^{-5}$. This choice insures that $\delta < \min(\delta_1(\epsilon), \delta_2(\epsilon))$ for all $\epsilon \in [0 : 10]$ considered in the plots. We note that while increasing K does not have significant impact on the bounds, both the download cost and multiplicative gap ratio decrease with N. Moreover, we observe that the bounds match when $\epsilon = 0$, i.e., when perfect user privacy is required, and when $\epsilon \to \infty$, i.e., no user privacy is required.

The generality of the AL-PIR problem formulation allows us to recover several existing results on PIR as special cases of Theorems 1 and 2. These cases are discussed in the following remark.

*Remark 2 (**Connections to state-of-the-art results**):* From Theorems 1 and 2, the lower and upper bounds on the optimal download cost $D^*(\epsilon, \delta)$ for any (ϵ, δ) can be used to derive the following prior results.

• **No user privacy and perfect DB privacy** ($\epsilon \to \infty, \delta = \delta_1(\epsilon \to \infty) = \delta_2(\epsilon \to \infty) = 0$). From the shared randomness bounds (16) and (18), when $\epsilon \to \infty$ and $\delta = 0$, we get that $\alpha^*(\epsilon \to \infty, 0) = \alpha_1(\epsilon \to \infty, 0) = \alpha_2(\epsilon \to \infty, 0) = 0$, i.e., no shared randomness is needed. Substituting the ϵ and δ values in the download cost bounds (17) and (20), we get $D^{UB}(\epsilon \to \infty, \delta = 0) = D^{LB}(\epsilon \to \infty, \delta = 0) = 1$, meaning that the upper and lower bounds are matching and give an optimal download cost of $D^*(\epsilon \to \infty, \delta = 0) = 1$. That is, AL-PIR is achieved by only downloading the requested file from any of the databases.
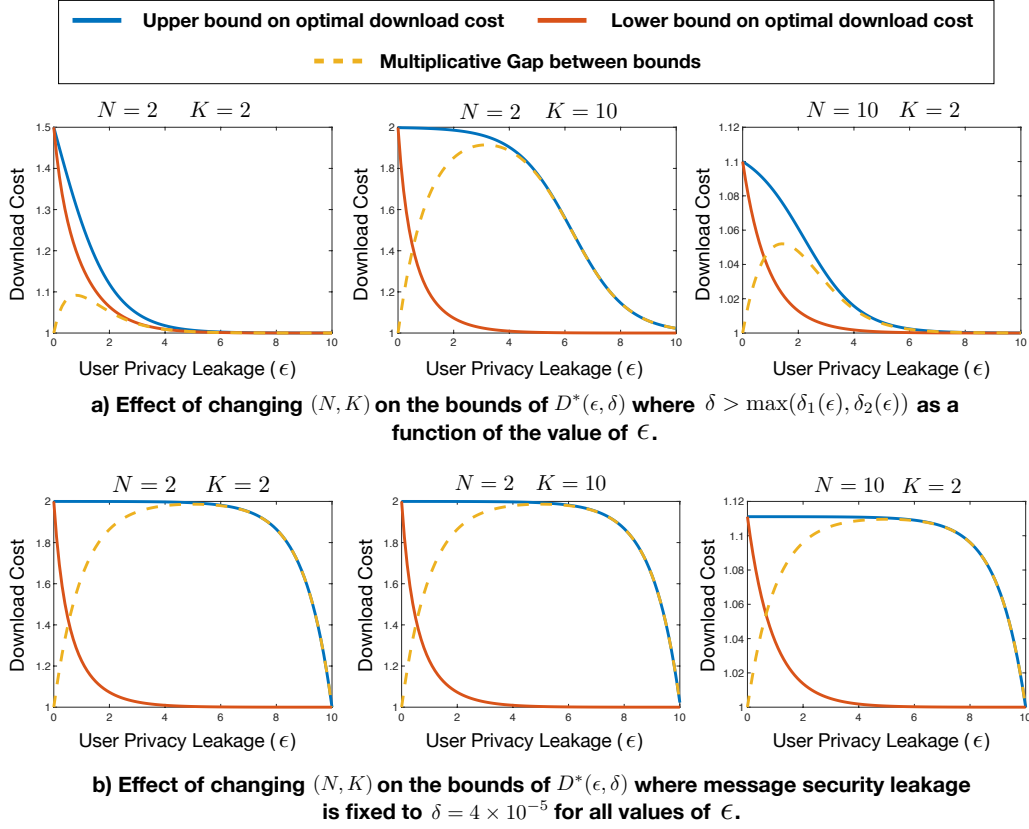
• **Perfect user privacy and maximum leakage on DB**

a) Effect of changing (N, K) **on the bounds of** $D^*(\epsilon, \delta)$ **where** $\delta > \max(\delta_1(\epsilon), \delta_2(\epsilon))$ **as a function of the value of** ϵ.

b) Effect of changing (N, K) **on the bounds of** $D^*(\epsilon, \delta)$ **where message security leakage is fixed to** $\delta = 4 \times 10^{-5}$ **for all values of** ϵ.

Fig. 3. Lower and upper bounds on the download cost of AL-PIR for different values of N, K and δ as ϵ increases.

privacy [7] ($\epsilon = 0, \delta = \delta_1(\epsilon = 0) = \delta_2(\epsilon = 0) = \frac{N^{K-1}-1}{N^{K-1}(N-1)}$). We obtain the original PIR result in [7] for perfect user privacy leakage $\epsilon = 0$. For this special case, we get the optimal required shared randomness characterized by $\alpha^*(\epsilon, \delta) = \alpha_1(\epsilon, \delta) = \alpha_2(\epsilon, \delta) = 0$, i.e., no shared randomness is needed. Using the bounds in (17) and (20), we obtain matching upper and lower bounds, giving an optimal download cost of

$$D^*(\epsilon = 0, \delta = \frac{N^{K-1}-1}{N^{K-1}(N-1)}) = D^{\text{LB}}(\epsilon = 0, \delta =$$

$$\frac{N^{K-1}-1}{N^{K-1}(N-1)}) = D^{\text{UB}}(\epsilon = 0, \delta = \frac{N^{K-1}-1}{N^{K-1}(N-1)})$$

$$= 1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}. \tag{24}$$

- **Perfect user privacy and DB privacy [9]** ($\epsilon = 0, \delta = 0$). By setting $\epsilon = 0$, $\delta = 0$ in Theorems 1 and 2, we obtain the SPIR results in [9] where the optimal required shared randomness is given by $\alpha^*(0,0) = \alpha_1(0,0) = \alpha_2(0,0) = \frac{1}{N-1}$ and the optimal download cost is obtained using the bounds in (17) and (20) as

$$D^*(\epsilon = 0, \delta = 0) = D^{\text{LB}}(\epsilon = 0, \delta = 0) = D^{\text{UB}}(\epsilon = 0, \delta = 0)$$

$$= 1 + \frac{1}{N-1}. \tag{25}$$

- **Leaky user privacy and maximum leakage on DB privacy [55]**. We obtain the L-PIR results in [55] for any

level of user privacy leakage ϵ and a DB privacy leakage $\delta \geq \max(\delta_1(\epsilon), \delta_2(\epsilon))$, where the optimal required shared randomness is given by $\alpha^*(\epsilon, \delta) = \alpha_1(\epsilon, \delta) = \alpha_2(\epsilon, \delta) = 0$ and the bounds on the optimal download cost are obtained using (17) and (20) as

$$D^*(\epsilon, \delta) \geq D^{\text{LB}}(\epsilon, \delta_2(\epsilon)) = 1 + \frac{1}{Ne^\epsilon} + \cdots + \frac{1}{(Ne^\epsilon)^{K-1}},$$

$$D^*(\epsilon, \delta) \leq D^{\text{UB}}(\epsilon, \delta_1(\epsilon)) = 1 + \frac{N^{K-1}-1}{(N-1)(e^\epsilon + N^{K-1} - 1)}. \tag{26}$$

- **Perfect user privacy and Leaky DB privacy [58]** ($\epsilon = 0, \delta$). For perfect user privacy $\epsilon = 0$ and DB privacy leakage characterized by δ, we obtain the results in [58], where the optimal required shared randomness is characterized by $\alpha^*(0, \delta) = \alpha_1(0, \delta) = \alpha_2(0, \delta) = \frac{1}{N-1} - \frac{N^{K-1}}{N^{K-1}-1}\delta$ and the optimal download cost is obtained using the bounds in (17) and (20) as

$$D^*(\epsilon = 0, \delta) = D^{\text{LB}}(\epsilon = 0, \delta) = D^{\text{UB}}(\epsilon = 0, \delta)$$

$$= \frac{N}{N-1} - \frac{\delta}{N^{K-1}-1}. \tag{27}$$

IV. PROOF OF THEOREM 1 : UPPER BOUND ON $D^*(\epsilon, \delta)$ FOR THE AL-PIR

The leakage in user privacy is achieved using the path-based approach introduced in our previous work [55]. A retrieval path is equivalent to a set of queries across databases that

guarantee decodability. Possible retrieval paths have different download costs. The probability of selecting each path is chosen to minimize the download cost while satisfying the privacy budget, measured by ϵ, which is a process referred to as *path biasing*. First, we give the following example for $N = K = 2$ to describe the idea of path biasing to achieve $\epsilon-$user privacy leakage with DB Privacy leakage ($\delta \geq \delta_1(\epsilon)$).

A. AL-PIR Example for $N = 2$, $K = 2$, and privacy leakage $(\epsilon, \delta \geq \delta_1(\epsilon))$

Consider the simplest non-trivial PIR setting with $N = 2$ DBs and $K = 2$ messages denoted by W_1 and W_2. To motivate the construction of AL-PIR, we first recall the perfect PIR scheme proposed by Sun and Jafar in [7]. Assume that the messages $W_1 = \{a_1, \ldots, a_4\}$ and $W_2 = \{b_1, \ldots, b_4\}$, are each $L = 4$ bits long. Figure 4 shows a retrieval structure for W_1 using the scheme in [7]. The main idea is that one can use coding and leverage side information from the other database to reduce the download cost to $3/2$. We highlight that the shown bit indices represent one possible permutation of the real indices. Thus, W_1 retrieval can be obtained through multiple bit structures that are selected uniformly and have an equal download cost of $3/2$.

In Figure 5, we show an alternative PIR scheme in which the requested message can be downloaded via sequences of structures that give unequal download cost. In particular, when the user wants to retrieve message W_1, it picks one of the four possible queries/paths:

- Path $\mathcal{P}_1$:$(\emptyset, W_1)$: Send no request to DB_1 and request W_1 from DB_2. This path/query has a download cost of L bits.
- Path $\mathcal{P}_2$:$(W_1, \emptyset)$: Request W_1 from DB_1 and send no request to DB_2. This path has a download cost of L bits.
- Path $\mathcal{P}_3$:$(W_2, W_1 \oplus W_2)$: Request W_2 from DB_1 and $W_1 \oplus W_2$ from DB_2. This path has a download cost of $2L$ bits.
- Path $\mathcal{P}_4$:$(W_1 \oplus W_2, W_2)$: Request $W_1 \oplus W_2$ from DB_1 and W_2 from DB_2. This path has a download cost of $2L$ bits.

Paths $\mathcal{P}_1$ and $\mathcal{P}_2$, which have lower download cost, are selected with probability p, whereas higher download cost paths $\mathcal{P}_3$ and $\mathcal{P}_4$ are selected with probability q. From the total probability theorem, we have

$$2p + 2q = 1. \tag{28}$$

The answer of DB_n can take four different structures, $\pi_{n,1}, \ldots, \pi_{n,4}$. These structures represent the element addition of all possible subsets of $\{W_1, W_2\}$. Note that the selection probability of any structure $\pi_{n,j}$, $j \in [1 : 4]$ equals the selection probability of all paths containing that structure. Also, there is one path per message that contains each structure $\pi_{n,j}$. For example, $\pi_{1,2} = \{W_1\}$ is paired with $\pi_{2,2} = \{\emptyset\}$ to retrieve W_1, or it can be paired with $\pi_{2,3} = \{W_1 \oplus W_2\}$ for W_2 retrieval. Let the path selection probabilities be uniform, i.e., $p = q = \frac{1}{4}$. Thus, each structure is selected with probability $\frac{1}{4}$, irrespective of the requested message index. It is straightforward to show that this probability assignment satisfies the perfect privacy definition in (7). Moreover, although the cost

Fig. 4. The original PIR scheme in [7] for $N = 2, K = 2$, and $L = 4$.
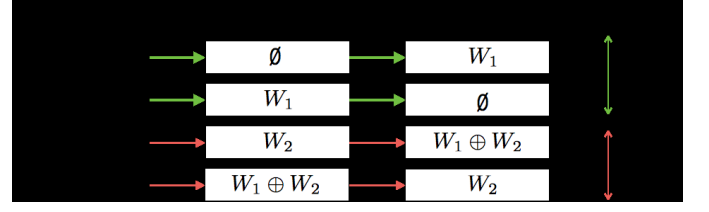

Fig. 5. AL-PIR scheme for $N = 2, K = 2$, general ϵ, and $\delta \geq \delta_1(\epsilon)$.

varies per path, the uniform path selection yields an optimal average download cost of $3/2$. Therefore, this path-based PIR scheme is also optimal and matches the result of Sun and Jafar [7] for perfect privacy.

Improving the download cost via path biasing (achieving $\epsilon-$user privacy). The leaky privacy definition in (9) together with the path-based scheme described above, lead us to consider schemes that bias the path selection process for retrieving desired messages. We next show that this helps reduce the average download cost for any non-zero ϵ. Intuitively, if we assign higher selection probability to paths with lower download cost than the average (for example L), an overall lower cost can be achieved at the expense of some bounded loss of privacy due to the biasing. The question we pose is whether there are values $p \neq q$ that yield an average download cost less than $\frac{3}{2}$ and simultaneously satisfy the $\epsilon-$user privacy definition in (9). The probability $Pr\{Q_n^{(i)} = \pi, A_n^{(i)} = \gamma | \mathbf{W}_\Omega\}$ can be expressed as

$$Pr\{Q_n^{(i)} = \pi, A_n^{(i)} = \gamma | \mathbf{W}_\Omega\} =$$
$$Pr\{Q_n^{(i)} = \pi | \mathbf{W}_\Omega\} Pr\{A_n^{(i)} = \gamma | Q_n^{(i)} = \pi, \mathbf{W}_\Omega\}. \tag{29}$$

The term $Pr\{Q_n^{(i)} = \pi | \mathbf{W}_\Omega\}$ depends on the path selection probability. To provide privacy, for any answer to a specific structure π, the term $Pr\{A_n^{(i)} = \gamma | Q_n^{(i)} = \pi, \mathbf{W}_\Omega\}$ should be constant independently of the requested message. To meet the privacy definition in (9), it is sufficient to show that the possible structures to each query satisfy:

$$\frac{\Pr(\pi_{n,j}|i=1)}{\Pr(\pi_{n,j}|i=2)} < e^\epsilon, \quad \forall n \in \{1,2\}, \; j \in [1:4], \tag{30}$$

where $\Pr(\pi_{n,j}|i=k)$, is the probability of retrieving structure $\pi_{n,j}$ when the desired message is k. Based on the scheme in Figure 5, there are two cases for each structure $\pi_{n,j}$:

(i) $\pi_{n,j}$ is used to recover W_1 and W_2 with the same probability either p or q, then

$$\frac{P(\pi_{n,j}|i=1)}{P(\pi_{n,j}|i=2)} = 1, \tag{31}$$

which clearly satisfies (9).

(ii) $\pi_{n,j}$ is selected with different probabilities p and q to

retrieve W_1 and W_2, respectively, and vice versa. Then, p and q must satisfy

$$e^{-\epsilon} \leq \frac{\Pr(\pi_{n,j}|i=1)}{\Pr(\pi_{n,j}|i=2)} = \frac{p}{q} \leq e^{\epsilon}. \quad (32)$$

Invoking the fact that the sum of path probabilities must equal one, we use (28) to substitute by $q = 0.5 - p$ and rewrite (32) as

$$\frac{p}{0.5 - p} \leq e^{\epsilon}. \quad (33)$$

This gives us the following inequality,

$$p \leq \frac{e^{\epsilon}}{2(1 + e^{\epsilon})}. \quad (34)$$

Therefore, we can pick p that satisfies (34) with equality, and then select $q = 0.5 - p$, as a valid choice of path selection probabilities which satisfy the $\epsilon-$user privacy constraint.

Computing the download cost $D(\epsilon, \delta \geq \delta(\epsilon))$. Since our scheme is symmetric with respect to messages, the same download cost is obtained for the retrieval of message W_1 or message W_2. Then, the average download cost can be written as

$$D(\epsilon, \delta \geq \delta(\epsilon)) = \frac{\sum_{j=1}^{4} \Pr\{\mathcal{P} = \mathcal{P}_j\} \cdot D_{\mathcal{P}_j}}{L}, \quad (35)$$

where $\Pr\{\mathcal{P} = \mathcal{P}_j\} \in \{p, q\}$ is the probability that path $\mathcal{P}_j$ is chosen and $D_{\mathcal{P}_j}$ is the cost of path $\mathcal{P}_j$. From Figure 5, we know that $D_{\mathcal{P}_1} = D_{\mathcal{P}_2} = L$, and $D_{\mathcal{P}_3} = D_{\mathcal{P}_4} = 2L$. Hence, $D(\epsilon, \delta \geq \delta(\epsilon))$ equals

$$\begin{aligned}
D(\epsilon, \delta \geq \delta(\epsilon)) &= \frac{2 \times p \times L + 2 \times q \times (2L)}{L} \\
&= 2p + 4q \\
&\stackrel{(a)}{=} 2 - 2p \\
&\stackrel{(b)}{\geq} 2 - \frac{e^{\epsilon}}{(1 + e^{\epsilon})}, \quad (36)
\end{aligned}$$

where (a) follows from (28), and (b) follows from (34). Hence, the download cost of this scheme (when $p = e^{\epsilon}/2(1+e^{\epsilon})$), can be rewritten as

$$D^*(\epsilon, \delta \geq \delta_1(\epsilon)) = \frac{3}{2} - \frac{e^{\epsilon} - 1}{2(e^{\epsilon} + 1)}, \quad (37)$$

which is lower than $\frac{3}{2}$, the optimal download cost under perfect privacy. Note that a lower cost cab be achieved for any ϵ.

Computing DB privacy leakage δ. We have shown in the above example that the biased selection probability of the path-based scheme can trade user privacy for lower download cost. We now calculate the DB privacy leakage. From the above leaky construction, we can show that

$$H(A_{[1:2]}^{(1)}) = D(\epsilon, \delta \geq \delta_1(\epsilon)) \times L \geq \frac{3}{2}L - \frac{e^{\epsilon} - 1}{2(e^{\epsilon} + 1)}L. \quad (38)$$

Similarly, the average size $H(A_{[1:2]}^{(1)}|W_2)$ of answers given W_2 is known and can be expressed as

$$\begin{aligned}
H(A_{[1:2]}^{(1)}|W_2) &= \sum_{j=1}^{4} \Pr\{\mathcal{P} = \mathcal{P}_j\} \cdot D_{\mathcal{P}_j|W_2} \\
&= 2 \times p \times L + 2 \times q \times L = L, \quad (39)
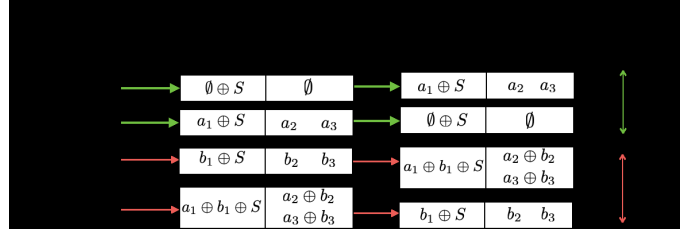\end{aligned}$$

Fig. 6. AL-PIR scheme for $N = 2$, $K = 2$, $\epsilon = \ln(1.5)$, and $\delta = 4/15$.

where $D_{\mathcal{P}_j|W_2}$ is the cost of path $\mathcal{P}_j$ when W_2 is given. This makes the DB privacy leakage, or the information revealed about W_2, equal to

$$\begin{aligned}
I(W_2; A_{[1:2]}^{(1)}) &= H(A_{[1:2]}^{(1)}) - H(A_{[1:2]}^{(1)}|W_2) \\
&\geq \frac{1}{2}L - \frac{e^{\epsilon} - 1}{2(e^{\epsilon} + 1)}L = \delta_1(\epsilon)L. \quad (40)
\end{aligned}$$

We highlight that this construction can achieve a lower DB privacy leakage compared to the perfect privacy scenario in [7] where $I(W_2; A_{[1:2]}^{(1)}) = L/2$, without the need for any shared randomness. However, this construction cannot fulfill the DB privacy constraint if $\delta < \delta_1(\epsilon)$. In the following example, we introduce a construction that can satisfy any DB privacy requirement with the utilization of the common randomness.

B. AL-PIR example with $N = 2$, $K = 2$, $\epsilon = \ln(1.5)$, and $\delta = 4/15$

Figure 6 shows an example of a possible AL-PIR scheme with $N = 2$, $K = 2$, $\epsilon = \ln(1.5)$, and $\delta = 4/15$. We observe that the allowed DB privacy leakage δ is less than $\delta_1(\epsilon)$,

$$\frac{4}{15} = \delta < \delta_1(\epsilon) = \frac{1}{2} - \frac{e^{\epsilon} - 1}{2(e^{\epsilon} + 1)} = 0.4. \quad (41)$$

Now, assume that each of the two messages is of size $L = 3$ bits, $W_1 = \{a_1, a_2, a_3\}$ and $W_2 = \{b_1, b_2, b_3\}$. To satisfy the $\delta-$DB privacy condition, we include the least required amount of shared randomness S that has a size of $\alpha_1(\epsilon, \delta)L$, where $\alpha_1(\epsilon, \delta)$ is computed from (16):

$$\alpha_1(\epsilon, \delta)L = \left(\frac{1}{N-1} - \frac{e^{\epsilon} + N^{K-1} - 1}{N^{K-1} - 1}\delta\right)L = \frac{L}{3} = 1 \text{ bit.} \quad (42)$$

Each message is divided into two parts as follows: W_1 is divided into $W_1^{(1)} = \{a_1\}$ (size of $S = L/3$), and $W_1^{(2)} = \{a_2, a_3\}$; and W_2 is divided into $W_2^{(1)} = \{b_1\}$, and $W_2^{(2)} = \{b_2, b_3\}$.

Suppose that the user wants to retrieve W_1. The user can use any of the four possible paths shown in Figure 6, where a path is defined as a query set $Q_{[1:N]}^{(k)}$ which satisfies, together with its corresponding answer, the correctness and privacy constraints. However, these paths have different download costs. The first two paths have a cost of $4L/3$ bits, whereas the other two paths have a cost of $2L$ bits. The correctness of the scheme is straightforward, the XOR addition of the two structures forming each path results in getting a_1, a_2, and a_3. To reduce the download cost by trading user privacy, similar

to the previous example, we select the lower cost paths with probability $p = 0.3$, whereas the higher cost paths are assigned a probability $q = 0.2$. These selection probabilities are chosen such that both $\epsilon-$user privacy and $\delta-$DB privacy conditions are satisfied. As we will discuss later in more details, the ratio describing $\epsilon-$user privacy leakage in (9) is given by the maximum ratio between the probabilities of selecting different paths, represented here as $p/q = 1.5 = e^\epsilon$.

When W_1 is requested by the user, the DB privacy leakage is described as the information user can decode about W_2. We notice that the first two paths in Figure 6 do not reveal any information about W_2, while using the other two paths, the user can decode the two bits $W_2^{(2)} = \{b_2, b_3\}$. This gives the average DB privacy leakage as

$$I(W_2; A_{[1:2]}^{(1)}) = 2 \times 0.3 \times 0 + 2 \times 0.2 \times 2 \times \frac{L}{3} =$$
$$0.8 \times \frac{L}{3} = \frac{4}{15} L = \delta L. \quad (43)$$

Hence, this achieves the $\delta-$DB privacy condition. The average number of downloaded bits for this scheme is

$$D_{\epsilon,\delta} = 2 \times 0.3 \times 4 + 2 \times 0.2 \times 6 = 4.8 \text{ bits}, \quad (44)$$

which yields a download cost of

$$D\left(\epsilon = \ln(1.5), \delta = \frac{4}{15}\right) = 4.8/3 = 1.6. \quad (45)$$

We highlight that this scheme clearly improves the download cost in comparison to the perfect SPIR, which has a download cost of $N/N-1 = 2$, at the expense of some loss in user and DB privacy.

C. General (ϵ, δ) AL-PIR Construction

In this section, we generalize the AL-PIR scheme in the previous examples for arbitrary values of N, K, and asymmetric privacy leakage characterized by the pair (ϵ, δ). Assume there are $K \geq 2$ messages, $W_1, \ldots, W_K$. Consider a random permutation of the databases indices. Let each message W_k be divided into two parts $W_k = \{W_k^{(1)}, W_k^{(2)}\}$ such that

$$H(W_k^{(1)}) = (N-1)\alpha_1(\epsilon, \delta)L, \quad (46)$$
$$H(W_k^{(2)}) = L - (N-1)\alpha_1(\epsilon, \delta)L, \quad (47)$$

where $\alpha_1(\epsilon, \delta)$ is the minimum required amount of shared randomness for the AL-PIR scheme to ensure the $\delta-$DB privacy and computed as

$$\alpha_1(\epsilon, \delta) = \max\left(0, \frac{1}{N-1} - \frac{e^\epsilon + N^{K-1} - 1}{N^{K-1} - 1}\delta\right)$$
$$= \begin{cases} \frac{1}{N-1} - \frac{e^\epsilon + N^{K-1} - 1}{N^{K-1} - 1}\delta, & 0 \leq \delta < \delta_1(\epsilon), \\ 0, & \delta > \delta_1(\epsilon). \end{cases} \quad (48)$$

Furthermore, let each $W_k^{(1)}$ and $W_k^{(2)}$ be divided into $N-1$ equal sub-packets,

$$W_k^{(1)} = \{W_{k,1}^{(1)}, \ldots, W_{k,N-1}^{(1)}\}, \quad (49)$$
$$W_k^{(2)} = \{W_{k,1}^{(2)}, \ldots, W_{k,N-1}^{(2)}\}, \quad (50)$$

such that for all $\ell \in [1 : N-1]$, $H(W_{k,\ell}^{(1)}) = \alpha_1(\epsilon, \delta)L$, $H(W_{k,\ell}^{(2)}) = (\frac{1}{N-1} - \alpha_1(\epsilon, \delta))L$. For instance, in the example of Figure 6 where $\alpha_1(\epsilon, \delta) = 1/3$ and $L = 3$ bits, W_1 is divided into $W_1^{(1)} = \{a_1\}$ of size 1 bit, and $W_1^{(2)} = \{a_2, a_3\}$ of size 2 bits.

For a requested message W_i, the DBs mask $W_k^{(1)}$'s, $k \in [1 : K] \setminus i$, with the secret key S. The content of $W_k^{(2)}$'s may be allowed to leak to the user. To retrieve a required message W_i, the user first selects one of the possible retrieval paths across the N DBs. Any path is formed by a set of N queries, $Q_{[1:N]}^{(k)}$, which are submitted to the respective DBs. The selected path has to fulfill two requirements: (i) the path correctly recovers W_i; (ii) the N submitted queries satisfy both the $\epsilon-$user privacy and $\delta-$DB privacy conditions. The user sends the following query vector to DB$_n$

$$\pi_{n,i} = (x_1, \ldots, x_{i-1}, (x_i + n)_N, x_{i+1}, \ldots, x_K),$$
$$x_k \in [0 : N-1], \ k \in [1 : K], \quad (51)$$

where $(x_i + n)_N$ denotes $(x_i + n) \pmod N$. This $K \times 1$ vector gives the indices of the K message bits, one bit for each message, that should be included in the answers. The design of $\pi_{n,i}$ makes sure that all submitted queries $\pi_{n,i}$'s include the same indices of all undesired messages, and different indices of the required W_i. Then, the identical undesired bits, within the N collected answers, can be utilized to decode the desired bits. After DB$_n$ receives the query $\pi_{n,i}$, it responds with answer $\gamma_n(\pi_{n,i})$,

$$\gamma_n(\pi_{n,i}) = \{ \bigoplus_{k \in [1:K] \setminus i} W_{k,x_k}^{(1)} \oplus S \oplus W_{i,(x_i+n)_N}^{(1)},$$
$$\bigoplus_{k \in [1:K] \setminus i} W_{k,x_k}^{(2)} \oplus W_{i,(x_i+n)_N}^{(2)} \}, \quad (52)$$

where $\bigoplus$ represents the summation via XOR operation. We denote by $W_{k,0}^{(1)}$ and $W_{k,0}^{(2)}$ the null or the empty set $\emptyset$. This response ensures protecting all $W_{k,\ell}^{(1)}$'s by encoding them with S. We observe that once one of the N queries is designed, i.e., the indices x_k's are chosen, the remaining $N-1$ queries are deterministic functions of these chosen indices. As $x_k \in [0 : N-1]$ for any $k \in [1 : K]$, each $\pi_{n,i}$ can be represented by N^K different vectors. Each of these vectors creates one possible path to retrieve W_i. Thus, for a specific permutation of DBs indices, we have N^K possible paths in general. For the example in Figure 6, there are $N^K = 4$ paths for the retrieval of W_1. The first retrieval path is created from the queries $\pi_{1,1} = (0,0)$, and $\pi_{2,1} = (1,0)$ with corresponding answers $\gamma_1(\pi_{1,1}) = \{S, \emptyset\}$, and $\gamma_2(\pi_{2,1}) = \{a_1 \oplus S, \{a_2, a_3\}\}$. A general form for one possible path is shown in Figure 7.

Analysis of Correctness: The user can decode the sub-messages $W_{i,\ell}^{(1)}$ and $W_{i,\ell}^{(2)}$, $\forall \ell \in [1 : N-1]$, of the requested message $(W_{i,\ell})$ using the information retrieved from DBs $N - x_i$ and $(N + \ell - x_i)_N$ as follows

$$\{W_{i,\ell}^{(1)}, W_{i,\ell}^{(2)}\} = \gamma_{N-x_i}(\pi_{N-x_i,i}) \oplus \gamma_{N+\ell-x_i}(\pi_{(N+\ell-x_i)_N,i})$$
$$= \{ \bigoplus_{k \in [1:K] \setminus i} W_{k,x_k}^{(1)} \oplus S \oplus W_{i,0}^{(1)}, \bigoplus_{k \in [1:K] \setminus i} W_{k,x_k}^{(2)} \oplus W_{i,0}^{(2)} \}$$
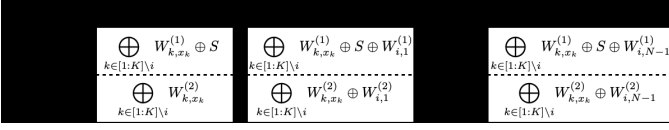
Fig. 7. One path of the AL-PIR scheme that retrieves W_i with $x_i = N-1$.

$$\oplus \Big\{ \bigoplus_{k \in [1:K] \setminus i} W_{k,x_k}^{(1)} \oplus S \oplus W_{i,\ell}^{(1)}, \ \bigoplus_{k \in [1:K] \setminus i} W_{k,x_k}^{(2)} \oplus W_{i,\ell}^{(2)} \Big\}$$

$$= \big\{ W_{i,0}^{(1)} \oplus W_{i,\ell}^{(1)}, \ W_{i,0}^{(2)} \oplus W_{i,\ell}^{(2)} \big\}$$

$$= \big\{ W_{i,\ell}^{(1)}, \ W_{i,\ell}^{(2)} \big\}. \tag{53}$$

Proof of $\epsilon-$user privacy: We note that the total download cost for each path is not fixed, but it depends on the choice of x_k's, $k \in [1:K]$. Then, we have two types of paths:

- **Lower cost paths** ($\forall \mathbf{k} \in [\mathbf{1}:\mathbf{K}] \setminus \mathbf{i}, \ \mathbf{x_k} = \mathbf{0}$):
  Generally, N possible paths belong to this case, those created from queries $\pi_{n,i}$ where $x_i = 0, 1, \ldots, N-1$. In this case, we have

$$\gamma_{N-x_i}(\pi_{N-x_i,i}) = \Big\{ \bigoplus_{k \in [1:K] \setminus i} W_{k,0}^{(1)} \oplus S \oplus W_{i,0}^{(1)},$$

$$\bigoplus_{k \in [1:K] \setminus i} W_{k,0}^{(2)} \oplus W_{i,0}^{(2)} \Big\} = \{S, \emptyset\}, \tag{54}$$

i.e., we only download the secret key S of size $\alpha_1(\epsilon, \delta)L$ from DB_{N-x_i}. Whereas, for other databases, all structures download data of the form $\big\{ W_{i,\ell}^{(1)} \oplus S, W_{i,\ell}^{(2)} \big\}$, each of size $\frac{L}{N-1}$ bits. In total for these type of paths, the user needs to download $(N-1)\frac{L}{N-1} + \alpha_1(\epsilon, \delta)L = (1 + \alpha_1(\epsilon, \delta))L$ bits from all DBs.

- **Higher cost paths** ($\exists \ \mathbf{k} \in [\mathbf{1}:\mathbf{K}] \setminus \mathbf{i}, \mathbf{x_k} \neq \mathbf{0}$):
  For this case, there are $N^K - N$ possible paths. Here, all requested query structures are of size $\frac{L}{N-1}$ bits, and the user needs to download $\frac{N}{N-1}L$ bits in total from all DBs.

Without loss of generality, we assign probabilities p and q, where $p \geq q$, to lower cost and higher cost paths, respectively[2], such that

$$N \times p + (N^K - N) \times q = 1. \tag{55}$$

We can see that any query $\pi_{n,i}$ with certain x_k's can be used to recover any desired message. This is obtained by requesting that query with other $N-1$ queries that share the same x_k's of the $K-1$ remaining messages. This is crucial to satisfy the $\epsilon-$user privacy requirements because accessing a structure does not eliminate any of the message possibilities. Furthermore, each structure is selected to retrieve W_i with the same probability of selecting the path coming through it, either p or q. Similar to (29), $\Pr\{Q_n^{(i)} = \pi, A_n^{(i)} = \gamma | \mathbf{W}_\Omega\}$ can be expressed as

$$\Pr\{Q_n^{(i)} = \pi, A_n^{(i)} = \gamma | \mathbf{W}_\Omega\} =$$

$$\Pr\{Q_n^{(i)} = \pi | \mathbf{W}_\Omega\} \Pr\{A_n^{(i)} = \gamma | Q_n^{(i)} = \pi, \mathbf{W}_\Omega\}. \tag{56}$$

[2]Due to symmetry, paths belonging to the same type are assigned the same probability. Assigning different probabilities does not improve the download cost or the privacy.

The term $\Pr\{A_n^{(i)} = \gamma | Q_n^{(i)} = \pi, \mathbf{W}_\Omega\}$ is also a constant, independent of the requested message. Thus, to meet the definition in (9), we show that possible structures of each query satisfy:

$$\frac{\Pr(\pi_{n,i}|i = k_1)}{\Pr(\pi_{n,i}|i = k_2)} \leq e^\epsilon, \forall n, \ k_1, k_2 \in [1:K], \tag{57}$$

where $\Pr(\pi_{n,i}|i = k_1)$, is the probability of selecting structure $\pi_{n,i}$ when the desired message is W_{k_1}. The following lemma generalizes the condition in (34) to satisfy $\epsilon-$user privacy for any K. It states the upper bound on the path biasing that does not violate the $\epsilon-$user privacy.

Lemma 1: To preserve $\epsilon-$user privacy definition of the AL-PIR, the biased probability p has to satisfy the following inequality

$$p \leq \frac{e^\epsilon}{Ne^\epsilon + N^K - N}. \tag{58}$$

Proof: Based on the proposed scheme, each structure $\pi_{n,i}$ can be selected with probability p or q, then

$$\frac{\Pr(\pi_{n,i}|i = k_1)}{\Pr(\pi_{n,i}|i = k_2)} \in \left\{ \frac{p}{p}, \frac{q}{q}, \frac{p}{q}, \frac{q}{p} \right\} \leq e^\epsilon. \tag{59}$$

As $p \geq q$, we only need to guarantee that

$$\frac{p}{q} \leq e^\epsilon. \tag{60}$$

Substituting (55) in the inequality, we get

$$e^\epsilon \geq \frac{p}{q} = \frac{(N^K - N)p}{(N^K - N)q} = \frac{(N^K - N)p}{1 - Np}. \tag{61}$$

By rearranging the above inequality, we get the following:

$$p \leq \frac{e^\epsilon}{Ne^\epsilon + N^K - N}. \tag{62}$$

Equation (55) can be used to find the following equivalent condition:

$$q \geq \frac{1}{Ne^\epsilon + N^K - N}. \tag{63}$$

∎

Analysis of $\delta-$DB privacy: We show that the proposed AL-PIR scheme satisfies the DB privacy leakage constraint in (6). From the previous construction, we categorized the paths into two groups: (a) N paths of size $(1 + \alpha_1(\epsilon, \delta))L$ bits; and (b) $N^K - N$ paths of size $\frac{N}{N-1}L$ bits. Then, the expected size of the answers, $H(A_{[1:N]}^{(i)})$ can be expressed as follows:

$$H(A_{[1:N]}^{(i)}) = pN(1 + \alpha_1(\epsilon, \delta))L + q(N^K - N)\frac{N}{N-1}L$$

$$\overset{(a)}{=} pN\alpha_1(\epsilon, \delta)L + L + \frac{N^K - N}{N-1}qL$$

$$\overset{(b)}{=} L + pN\alpha_1(\epsilon, \delta)L + \frac{1 - pN}{N-1}L$$

$$= L + \frac{L}{N-1} - pN\left(\frac{1}{N-1} - \alpha_1(\epsilon, \delta)\right)L, \tag{64}$$

where (a) and (b) follow from (55). We then calculate

$H(A^{(i)}_{[1:N]}|\mathbf{W}_{[1:K]\setminus i})$ as follows:

$$
\begin{aligned}
H(A^{(i)}_{[1:N]}|\mathbf{W}_{[1:K]\setminus i}) &\overset{(a)}{=} H(W_i, A^{(i)}_{[1:N]}|\mathbf{W}_{[1:K]\setminus i})\\
&\overset{(b)}{=} H(W_i) + H(S, A^{(i)}_{[1:N]}|\mathbf{W}_{[1:K]\setminus i}, W_i)\\
&= H(W_i) + H(S) + H(A^{(i)}_{[1:N]}|\mathbf{W}_{[1:K]\setminus i}, W_i, S)\\
&\overset{(c)}{=} H(W_i) + H(S) = (1 + \alpha_1(\epsilon,\delta))L,
\end{aligned}
\tag{65}
$$

where (a) follows the correctness property in (4) whereas (b) and (c) hold from the fact that answers are function of messages and the shared randomness.

Lemma 2: To preserve $\delta-$DB privacy for $\delta < \delta_1(\epsilon)$, the biased probability p has to satisfy the following inequality

$$
p \geq \frac{e^\epsilon}{Ne^\epsilon + N^K - N}.
\tag{66}
$$

Proof: According to (64) and (65), we can express the DB privacy leakage as

$$
\begin{aligned}
\delta L &\geq I(\mathbf{W}_{[1:K]\setminus i}; A^{(i)}_{[1:N]})\\
&= H(A^{(i)}_{[1:N]}) - H(A^{(i)}_{[1:N]}|\mathbf{W}_{[1:K]\setminus i})\\
&= L + \frac{L}{N-1} - pN\left(\frac{1}{N-1} - \alpha_1(\epsilon,\delta)\right)L\\
&\quad - (1 + \alpha_1(\epsilon,\delta))L\\
&\overset{(a)}{=} (pN-1)\alpha_1(\epsilon,\delta)L + \frac{1-pN}{N-1}L\\
&= (1-pN)(\frac{1}{N-1} - \alpha_1(\epsilon,\delta))L\\
&\overset{(b)}{=} (1-pN)\min\left(\frac{1}{N-1}, \frac{e^\epsilon + N^{K-1} - 1}{N^{K-1}-1}\delta\right)L,
\end{aligned}
\tag{67}
$$

where (a) follows from (55) and (b) follows from (48). For the commonly shared randomness S, we have one of the following two cases:

- **No shared randomness is needed ($\alpha_1(\epsilon,\delta)=0$):**
  In this case, the condition in (67) can be written as follows,

$$
\delta L \geq \frac{1-pN}{N-1}L \overset{(a)}{\geq} \frac{N^{K-1}-1}{(N-1)(e^\epsilon + N^{K-1}-1)}L = \delta_1(\epsilon)L,
\tag{68}
$$

  where step (a) follows by applying the $\epsilon-$user privacy condition obtained in Lemma 1. Therefore, we obtain the bound on the DB privacy leakage $\delta \geq \delta_1(\epsilon)$, i.e., DB privacy leakage is maximized which covers the L-PIR model previously considered in [55]. This case requires no condition on the biased probability p as the inequality in (68) is achieved for any p. We highlight that this scheme obtains a better DB privacy compared to the perfect PIR scheme proposed in [7], without the need to any shared amount of randomness. The latter scheme causes a leakage of $\frac{N^{K-1}-1}{(N-1)N^{K-1}}L$ bits.
- **Shared randomness is needed ($\alpha_1(\epsilon,\delta)>0$):**
  For any $\alpha_1(\epsilon,\delta)>0$, we always have

$$
\frac{1}{N-1} > \frac{e^\epsilon + N^{K-1} - 1}{N^{K-1}-1}\delta.
\tag{69}
$$

From (67), we get the following relation on p:

$$
(1-pN) \leq \frac{N^{K-1}-1}{e^\epsilon + N^{K-1} - 1},
\tag{70}
$$

which leads to the proof of Lemma 2.

∎

Lemmas 1 and 2 lead to the following necessary condition on p to simultaneously satisfy the $\epsilon-$user privacy and $\delta-$DB privacy definitions.

Lemma 3: To preserve $\epsilon-$user privacy and $\delta-$DB privacy, the biased probability p has to satisfy the following condition with equality

$$
p = \frac{e^\epsilon}{Ne^\epsilon + N^K - N}.
\tag{71}
$$

Proof: For $\delta < \delta_1(\epsilon)$, the proof follows directly by applying Lemmas 1 and 2. For $\delta \geq \delta_1(\epsilon)$, the proof follows from Lemma 1 where we pick the maximum value of the biasing probability p in order to minimize the download cost, i.e., maximize the probability of picking the paths of lower cost.

∎

Analysis of download cost: Given that the scheme is symmetric across all messages, the download cost can be written as,

$$
\begin{aligned}
D(\epsilon,\delta) &= \frac{H(A^{(i)}_{[1:N]})}{L}\\
&\overset{(a)}{=} 1 + \frac{1}{N-1} - pN\left(\frac{1}{N-1} - \alpha_1(\epsilon,\delta)\right)\\
&\overset{(b)}{=} 1 + \frac{1}{N-1} - pN\min\left(\frac{1}{N-1}, \frac{e^\epsilon + N^{K-1}-1}{N^{K-1}-1}\delta\right)\\
&\overset{(c)}{=} 1 + \frac{1}{N-1} - \frac{e^\epsilon}{e^\epsilon + N^{K-1}-1}\\
&\quad \times \min\left(\frac{1}{N-1}, \frac{e^\epsilon + N^{K-1}-1}{N^{K-1}-1}\delta\right),
\end{aligned}
\tag{72}
$$

where (a) follows from (64), (b) follows (48), and (c) is due to Lemma 3. According to the size of the available randomness S, we have one of the following two cases:

- **No shared randomness is needed ($\alpha_1(\epsilon,\delta)=0$):**
  This case corresponds to $\delta \geq \delta_1(\epsilon)$. The download cost in (72) can be written as follows,

$$
\begin{aligned}
D(\epsilon,\delta) &= 1 + \frac{1}{N-1} - \frac{e^\epsilon}{e^\epsilon + N^{K-1}-1} \times \frac{1}{N-1}\\
&= 1 + \frac{N^{K-1}-1}{(N-1)(e^\epsilon + N^{K-1}-1)}\\
&= 1 + \frac{N^{K-1}}{e^\epsilon + N^{K-1}-1}\left(\frac{1}{N} + \cdots + \frac{1}{N^{K-1}}\right)\\
&= d_1(\epsilon, \delta_1(\epsilon)).
\end{aligned}
\tag{73}
$$

- **Shared randomness is needed ($\alpha_1(\epsilon,\delta)>0$):**
  For any $\alpha_1(\epsilon,\delta)>0$, we have

$$
\frac{1}{N-1} > \frac{e^\epsilon + N^{K-1}-1}{N^{K-1}-1}\delta.
\tag{74}
$$

Then, the download cost in (72) can be re-expressed as

$$
D(\epsilon,\delta) = 1 + \frac{1}{N-1}
$$

$$-\frac{e^\epsilon}{e^\epsilon+N^{K-1}-1}\times\frac{e^\epsilon+N^{K-1}-1}{N^{K-1}-1}\delta$$

$$=1+\frac{1}{N-1}-\frac{\delta e^\epsilon}{N^{K-1}-1}=d_1(\epsilon,\delta).\quad(75)$$

Both cases in (73) and (75) yield the upper bound in (17) for the download cost of AL-PIR and prove Theorem 1.

V. PROOF OF THEOREM 2 : LOWER BOUND ON $D^*(\epsilon,\delta)$

Without loss of generality, assume the requested message is W_1. We can bound $D^*(\epsilon,\delta)$ as follows

$$D^*(\epsilon,\delta)=\frac{\sum_{n=1}^N H(A_n^{(1)})}{L}$$
$$\geq\frac{H(A_{[1:N]}^{(1)})}{L}$$
$$\geq\frac{H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})}{L}.\quad(76)$$

To further bound $D^*(\epsilon,\delta)$, we first state the following two lemmas. Proofs of both lemmas can be found in the appendices. In Lemma 4, we introduce the relation between the entropy of answers downloaded to retrieve different messages given a certain message. We emphasize that, under perfect privacy definitions, the entropy should be exactly the same regardless of the requested message,

$$H(A_n^{(k_1)}|W_{k_1},Q_n^{(k_1)})=H(A_n^{(k_2)}|W_{k_1},Q_n^{(k_2)}),$$
$$\forall k_1\neq k_2,\ n\in[1:N].\quad(77)$$

However, this does not hold under the $\epsilon-$user privacy definition.

Lemma 4: Under the $\epsilon-$user privacy definition, for any k_1 and $k_2\in[1:K]$ and a non-negative constant ϵ, we have the following inequality

$$H(A_n^{(k_1)}|W_{k_1},Q_n^{(k_1)})\geq\frac{1}{e^\epsilon}H(A_n^{(k_2)}|W_{k_1},Q_n^{(k_2)})$$
$$\forall k_1\neq k_2,\ n\in[1:N].\quad(78)$$

Using Lemma 4, we get the following recursion lemma.
Lemma 5: For $k\in[2,K]$, we have

$$H(A_{[1:N]}^{(k)}|W_{[1:k-1]},Q_{[1:N]}^{(k)})$$
$$\geq(1-o(L))L+\frac{1}{Ne^\epsilon}H(A_{[1:N]}^{(k+1)}|W_{[1:k]},Q_{[1:N]}^{(k+1)}).$$

Using Lemmas 4, and 5, we bound $H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})$ as follows

$$H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})$$
$$=H(W_1,A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})-H(W_1|A_{[1:N]}^{(1)},Q_{[1:N]}^{(1)})$$
$$\overset{(a)}{=}H(W_1,A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})-o(L)L$$
$$=H(W_1|Q_{[1:N]}^{(1)})+H(A_{[1:N]}^{(1)}|W_1,Q_{[1:N]}^{(1)})-o(L)L$$
$$\overset{(b)}{=}(1-o(L))L+H(A_{[1:N]}^{(1)}|W_1,Q_{[1:N]}^{(1)})$$
$$\geq(1-o(L))L+H(A_n^{(1)}|W_1,Q_{[1:N]}^{(1)})$$
$$\overset{(c)}{=}(1-o(L))L+H(A_n^{(1)}|W_1,Q_n^{(1)})$$

$$\overset{(d)}{\geq}(1-o(L))L+\frac{1}{e^\epsilon}H(A_n^{(2)}|W_1,Q_n^{(2)}),\quad(79)$$

where (a) is due to the correctness property in (4), (b) follows from the fact that the message content is independent of queries, (c) comes from the fact that the answer $A_n^{(1)}$ is conditionally independent of the queries submitted to other DBs given the query $Q_n^{(1)}$, whereas (d) comes from Lemma 4. The addition of the previous relation over all possible n's gives us the following

$$NH(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})$$
$$\geq N(1-o(L))L+\frac{1}{e^\epsilon}\sum_{n=1}^N H(A_n^{(2)}|W_1,Q_n^{(2)}).\quad(80)$$

Dividing by N,

$$H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})$$
$$\geq(1-o(L))L+\frac{1}{Ne^\epsilon}\sum_{n=1}^N H(A_n^{(2)}|W_1,Q_n^{(2)})$$
$$\geq(1-o(L))L+\frac{1}{Ne^\epsilon}\sum_{n=1}^N H(A_n^{(2)}|W_1,Q_{[1:N]}^{(2)})$$
$$\geq(1-o(L))L+\frac{1}{Ne^\epsilon}H(A_{[1:N]}^{(2)}|W_1,Q_{[1:N]}^{(2)})$$
$$\overset{(a)}{=}(1-o(L))L+\frac{1}{Ne^\epsilon}H(A_{[1:N]}^{(2)}|W_1,Q_{[1:N]}^{(2)})$$
$$+\frac{1}{Ne^\epsilon}H(W_2|A_{[1:N]}^{(2)},W_1,Q_{[1:N]}^{(2)})-\frac{o(L)L}{Ne^\epsilon}$$
$$=(1-o(L))L+\frac{1}{Ne^\epsilon}H(W_2,A_{[1:N]}^{(2)}|W_1,Q_{[1:N]}^{(2)})-\frac{o(L)L}{Ne^\epsilon}$$
$$=(1-o(L))L+\frac{1}{Ne^\epsilon}H(W_2|W_1,Q_{[1:N]}^{(2)})$$
$$+\frac{1}{Ne^\epsilon}H(A_{[1:N]}^{(2)}|W_1,W_2,Q_{[1:N]}^{(2)})-\frac{o(L)L}{Ne^\epsilon}$$
$$\overset{(b)}{=}(1-o(L))L+\frac{1}{Ne^\epsilon}L+\frac{1}{Ne^\epsilon}H(A_{[1:N]}^{(2)}|W_1,W_2,Q_{[1:N]}^{(2)})$$
$$-\frac{o(L)L}{Ne^\epsilon}$$
$$=(1-o(L))L+\frac{1}{Ne^\epsilon}(1-o(L))L$$
$$+\frac{1}{Ne^\epsilon}H(A_{[1:N]}^{(2)}|W_1,W_2,Q_{[1:N]}^{(2)}),\quad(81)$$

where (a) comes from the correctness property in (4), and (b) is due to the message independence. Following the same iterative process used in [7], and invoking the recursion property in Lemma 5, we get

$$H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})$$
$$\geq(1+\frac{1}{Ne^\epsilon}+\cdots+\frac{1}{(Ne^\epsilon)^{K-1}})(1-o(L))L$$
$$+\frac{1}{(Ne^\epsilon)^{K-1}}H(A_{[1:N]}^{(K)}|W_{[1:K]},Q_{[1:N]}^{(K)})$$
$$\overset{(a)}{=}(1-o(L))L+\frac{(Ne^\epsilon)^{K-1}-1}{(Ne^\epsilon)^{K-1}(Ne^\epsilon-1)}(1-o(L))L$$
$$+\frac{1}{(Ne^\epsilon)^{K-1}}H(A_{[1:N]}^{(K)}|W_{[1:K]},Q_{[1:N]}^{(K)}),\quad(82)$$

where (a) follows from the rule of finite sum of geometric series. Under the L-PIR model presented in [55], the term $H(A_{[1:N]}^{(K)}|W_{[1:K]}, Q_{[1:N]}^{(K)})$ is replaced by zero as answers are functions of only the K messages. However, this does not hold in the presence of common randomness. From the $\delta-$DB privacy definition in (6), we get the following:

$$H(A_{[1:N]}^{(K)}|W_{[1:K]}, Q_{[1:N]}^{(K)})$$
$$= H(A_{[1:N]}^{(K)}|W_{[1:K-1]}, Q_{[1:N]}^{(K)})$$
$$\quad - I(W_K; A_{[1:N]}^{(K)}|W_{[1:K-1]}, Q_{[1:N]}^{(K)})$$
$$\overset{(a)}{=} H(A_{[1:N]}^{(K)}|W_{[1:K-1]}, Q_{[1:N]}^{(K)}) - H(W_K)$$
$$= H(A_{[1:N]}^{(K)}|Q_{[1:N]}^{(K)}) - I(A_{[1:N]}^{(K)}; W_{[1:K-1]}|Q_{[1:N]}^{(K)}) - H(W_K)$$
$$= H(A_{[1:N]}^{(K)}|Q_{[1:N]}^{(K)}) - I(A_{[1:N]}^{(K)}, Q_{[1:N]}^{(K)}; W_{[1:K-1]}) - H(W_K)$$
$$\geq H(A_{[1:N]}^{(K)}|Q_{[1:N]}^{(K)}) - \delta L - L, \tag{83}$$

where (a) follows since all messages are independent and W_K is a deterministic function of $A_{[1:N]}^{(K)}$. By symmetry, we can assume that $H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)}) = H(A_{[1:N]}^{(K)}|Q_{[1:N]}^{(K)})$. Then,

$$H(A_{[1:N]}^{(K)}|W_{[1:K]}, Q_{[1:N]}^{(K)}) \geq H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)}) - \delta L - L. \tag{84}$$

Since $H(A_{[1:N]}^{(K)}|W_{[1:K]}, Q_{[1:N]}^{(K)}) \geq 0$, we obtain

$$H(A_{[1:N]}^{(K)}|W_{[1:K]}, Q_{[1:N]}^{(K)})$$
$$\geq \max\left(0, \ H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)}) - \delta L - L\right).$$

Next, we can express (82) using (85) as

$$H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)}) \geq$$
$$(1 + \frac{(Ne^\epsilon)^{K-1} - 1}{(Ne^\epsilon)^{K-1}(Ne^\epsilon - 1)})(1 - o(L))L$$
$$+ \frac{1}{(Ne^\epsilon)^{K-1}} \max\left(0, H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)}) - \delta L - L\right). \tag{85}$$

Dividing by L and allowing it to approach ∞, we get

$$\frac{H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})}{L} \geq 1 + \frac{(Ne^\epsilon)^{K-1} - 1}{(Ne^\epsilon)^{K-1}(Ne^\epsilon - 1)}$$
$$+ \frac{1}{(Ne^\epsilon)^{K-1}} \max\left(0, \frac{H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})}{L} - \delta - 1\right). \tag{86}$$

Following (86), the following two inequalities are true

$$\frac{H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})}{L} \geq 1 + \frac{(Ne^\epsilon)^{K-1} - 1}{(Ne^\epsilon)^{K-1}(Ne^\epsilon - 1)}, \tag{87}$$

$$\frac{H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})}{L} \geq 1 + \frac{(Ne^\epsilon)^{K-1} - 1}{(Ne^\epsilon)^{K-1}(Ne^\epsilon - 1)}$$
$$+ \frac{1}{(Ne^\epsilon)^{K-1}}(\frac{H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})}{L} - \delta - 1). \tag{88}$$

The inequality in (88) can be rearranged as

$$\frac{(Ne^\epsilon)^{K-1} - 1}{(Ne^\epsilon)^{K-1}} \frac{H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})}{L}$$
$$\geq \ 1 + \frac{(Ne^\epsilon)^{K-1} - 1}{(Ne^\epsilon)^{K-1}(Ne^\epsilon - 1)} - \frac{1}{(Ne^\epsilon)^{K-1}}(\delta + 1)$$
$$= \frac{(Ne^\epsilon)^{K-1} - 1}{(Ne^\epsilon)^{K-1}} + \frac{(Ne^\epsilon)^{K-1} - 1}{(Ne^\epsilon)^{K-1}(Ne^\epsilon - 1)} - \frac{\delta}{(Ne^\epsilon)^{K-1}}, \tag{89}$$

$$\frac{H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})}{L} \geq \ 1 + \frac{1}{Ne^\epsilon - 1} - \frac{\delta}{(Ne^\epsilon)^{K-1} - 1}. \tag{90}$$

From (87) and (90), we get

$$\frac{H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})}{L} \geq \max\big(1 + \frac{(Ne^\epsilon)^{K-1} - 1}{(Ne^\epsilon)^{K-1}(Ne^\epsilon - 1)},$$
$$1 + \frac{1}{Ne^\epsilon - 1} - \frac{\delta}{(Ne^\epsilon)^{K-1} - 1}\big).$$

Substituting by (91) in (76), we can lower bound $D^*(\epsilon, \delta)$ as

$$D^*(\epsilon, \delta) \geq \frac{H(A_{[1:N]}^{(1)}|Q_{[1:N]}^{(1)})}{L} \geq D^{\text{LB}}(\epsilon, \delta)$$
$$= \max\big(1 + \frac{(Ne^\epsilon)^{K-1} - 1}{(Ne^\epsilon)^{K-1}(Ne^\epsilon - 1)}$$
$$, \ 1 + \frac{1}{Ne^\epsilon - 1} - \frac{\delta}{(Ne^\epsilon)^{K-1} - 1}\big). \tag{91}$$

For a fixed ϵ, $D^{\text{LB}}(\epsilon, \delta)$ is monotonically decreasing in δ until we reach $\delta = \delta_2(\epsilon) = \frac{(Ne^\epsilon)^{K-1} - 1}{(Ne^\epsilon - 1)(Ne^\epsilon)^{K-1}}$ at which

$$1 + \frac{(Ne^\epsilon)^{K-1} - 1}{(Ne^\epsilon)^{K-1}(Ne^\epsilon - 1)} = 1 + \frac{1}{Ne^\epsilon - 1} - \frac{\delta}{(Ne^\epsilon)^{K-1} - 1}. \tag{92}$$

After this point, $D^{\text{LB}}(\epsilon, \delta)$ is fixed at the value $1 + \frac{(Ne^\epsilon)^{K-1} - 1}{(Ne^\epsilon)^{K-1}(Ne^\epsilon - 1)}$. Then, we can alternatively represent $D^{\text{LB}}(\epsilon, \delta)$ as

$$D^*(\epsilon, \delta) \geq D^{\text{LB}}(\epsilon, \delta)$$
$$= \begin{cases} 1 + \frac{1}{Ne^\epsilon - 1} - \frac{\delta}{(Ne^\epsilon)^{K-1} - 1} = d_2(\epsilon, \delta), & 0 \leq \delta < \delta_2(\epsilon), \\ 1 + \frac{(Ne^\epsilon)^{K-1} - 1}{(Ne^\epsilon)^{K-1}(Ne^\epsilon - 1)} = d_2(\epsilon, \delta_2(\epsilon)), & \delta \geq \delta_2(\epsilon). \end{cases} \tag{93}$$

This proves the lower bound on $D^*(\epsilon, \delta)$ in Theorem 2.

A. Required amount of shared randomness

In this section, we prove the lower bound in Theorem 2 on the required amount of shared randomness to achieve the minimum download cost derived in (91). From the $\delta-$DB privacy in (6), given a requested message W_k, we get

$$\delta L \geq I(A_{[1:N]}^{(k)}, Q_{[1:N]}^{(k)}; W_{[1:K]\backslash k})$$
$$= I(A_{[1:N]}^{(k)}; W_{[1:K]\backslash k}|Q_{[1:N]}^{(k)})$$
$$= H(A_{[1:N]}^{(k)}|Q_{[1:N]}^{(k)}) - H(A_{[1:N]}^{(k)}|W_{[1:K]\backslash k}, Q_{[1:N]}^{(k)})$$
$$= H(A_{[1:N]}^{(k)}|Q_{[1:N]}^{(k)}) - H(W_k, A_{[1:N]}^{(k)}|W_{[1:K]\backslash k}, Q_{[1:N]}^{(k)})$$

$$+ H(W_k|A^{(k)}_{[1:N]}, W_{[1:K]\backslash k}, Q^{(k)}_{[1:N]})$$

$$\overset{(a)}{=} H(A^{(k)}_{[1:N]}|Q^{(k)}_{[1:N]}) - H(W_k, A^{(k)}_{[1:N]}|W_{[1:K]\backslash k}, Q^{(k)}_{[1:N]})$$
$$+ o(L)L$$

$$= H(A^{(k)}_{[1:N]}|Q^{(k)}_{[1:N]}) - H(W_k|W_{[1:K]\backslash k}, Q^{(k)}_{[1:N]})$$
$$- H(A^{(k)}_{[1:N]}|W_{[1:K]}, Q^{(k)}_{[1:N]}) + o(L)L$$

$$= H(A^{(k)}_{[1:N]}|Q^{(k)}_{[1:N]}) - L - H(A^{(k)}_{[1:N]}|W_{[1:K]}, Q^{(k)}_{[1:N]})$$
$$+ o(L)L$$

$$= H(A^{(k)}_{[1:N]}|Q^{(k)}_{[1:N]}) - (1 - o(L))L$$
$$- H(A^{(k)}_{[1:N]}|W_{[1:K]}, S, Q^{(k)}_{[1:N]}) - I(S; A^{(k)}_{[1:N]}|W_{[1:K]}, Q^{(k)}_{[1:N]})$$

$$\overset{(b)}{=} H(A^{(k)}_{[1:N]}|Q^{(k)}_{[1:N]}) - (1 - o(L))L$$
$$- I(S; A^{(k)}_{[1:N]}|W_{[1:K]}, Q^{(k)}_{[1:N]})$$

$$= H(A^{(k)}_{[1:N]}|Q^{(k)}_{[1:N]}) - (1 - o(L))L - H(S|W_{[1:K]}, Q^{(k)}_{[1:N]})$$
$$+ H(S|A^{(k)}_{[1:N]}, W_{[1:K]}, Q^{(k)}_{[1:N]})$$

$$\overset{(c)}{=} H(A^{(k)}_{[1:N]}|Q^{(k)}_{[1:N]}) - (1 - o(L))L - H(S)$$
$$+ H(S|A^{(k)}_{[1:N]}, W_{[1:K]}, Q^{(k)}_{[1:N]})$$

$$\geq H(A^{(k)}_{[1:N]}|Q^{(k)}_{[1:N]}) - (1 - o(L))L - H(S), \quad (94)$$

where (a) follows from the correctness property, (b) comes from the fact that answers are function of the K messages and the common randomness S, and (c) is because the common randomness S is independent of the K messages. Dividing by L, allowing it to approach ∞, and substituting by (90) in (94), we get

$$\delta \geq 1 + \frac{1}{Ne^\epsilon - 1} - \frac{\delta}{(Ne^\epsilon)^{K-1} - 1} - 1 - \frac{H(S)}{L}$$
$$= \frac{1}{Ne^\epsilon - 1} - \frac{\delta}{(Ne^\epsilon)^{K-1} - 1} - \frac{H(S)}{L}. \quad (95)$$

Rearranging the inequality, we get the following bound on $\alpha(\epsilon, \delta)$,

$$\frac{H(S)}{L} = \alpha(\epsilon, \delta) \geq \frac{1}{Ne^\epsilon - 1} - \frac{\delta}{(Ne^\epsilon)^{K-1} - 1} - \delta, \quad (96)$$

which is also a valid bound on the optimal common randomness $\alpha^*(\epsilon, \delta)$. Then, following that $\alpha^*(\epsilon, \delta) \geq 0$, we obtain the following bound,

$$\alpha^*(\epsilon, \delta) \geq \alpha_2(\epsilon, \delta) = \max\left(0, \frac{1}{Ne^\epsilon - 1} - \frac{(Ne^\epsilon)^{K-1}}{(Ne^\epsilon)^{K-1} - 1}\delta\right)$$
$$= \begin{cases} \frac{1}{Ne^\epsilon - 1} - \frac{(Ne^\epsilon)^{K-1}}{(Ne^\epsilon)^{K-1} - 1}\delta, & 0 \leq \delta < \delta_2(\epsilon), \\ 0, & \delta > \delta_2(\epsilon), \end{cases}$$
$$(97)$$

which completes the proof of the lower bound on the optimal common randomness size in Theorem 2.

VI. CONCLUSIONS

We studied the AL-PIR problem that relaxes the perfect privacy requirements for both user and DB privacy. The allowed leakage is asymmetric allowing for different privacy leakage in each direction. We showed that allowing privacy leakage provides an opportunity to improve the optimal download cost. We introduced an AL-PIR scheme that gives an upper bound on the optimal download cost for arbitrary leakage budgets. We investigated possible tradeoffs that stem by adjusting the level of privacy at both user and DB sides. We further obtained a lower bound on the download cost and showed that the multiplicative gap between the upper and lower bounds is bounded by $\frac{N - e^{-\epsilon}}{N-1}$, i.e., our AL-PIR scheme is optimal for perfect user privacy, $\epsilon = 0$, and is optimal within a gap of at most $\frac{N}{N-1}$ for any ϵ.

REFERENCES

[1] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," in *Proceedings of the IEEE Annual Symposium on Foundations of Computer Science*, 1995, pp. 41–50.

[2] S. Yekhanin *et al.*, "Locally decodable codes," *Foundations and Trends® in Theoretical Computer Science*, vol. 6, no. 3, pp. 139–255, 2012.

[3] W. Gasarch, "A survey on private information retrieval," *Bulletin of the EATCS*, vol. 82, no. 72-107, p. 113, 2004.

[4] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *Proceeding of the IEEE Symposium on Security and Privacy. S&P*, 2000, pp. 44–55.

[5] R. Ostrovsky and W. E. Skeith, "A survey of single-database private information retrieval: Techniques and applications," in *International Workshop on Public Key Cryptography*. Springer, 2007, pp. 393–411.

[6] N. B. Shah, K. Rashmi, and K. Ramchandran, "One extra bit of download ensures perfectly private information retrieval," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2014, pp. 856–860.

[7] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4075–4088, 2017.

[8] ——, "Optimal download cost of private information retrieval for arbitrary message length," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 12, pp. 2920–2932, 2017.

[9] ——, "The capacity of symmetric private information retrieval," *IEEE Transactions on Information Theory*, vol. 65, no. 1, pp. 322–329, 2018.

[10] ——, "Multiround private information retrieval: Capacity and storage overhead," *IEEE Transactions on Information Theory*, vol. 64, no. 8, pp. 5743–5754, 2018.

[11] ——, "The capacity of robust private information retrieval with colluding databases," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2361–2370, 2017.

[12] R. Tajeddine, O. W. Gnilke, and S. El Rouayheb, "Private information retrieval from MDS coded data in distributed storage systems," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 7081–7093, 2018.

[13] K. Banawan and S. Ulukus, "The capacity of private information retrieval from coded databases," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1945–1956, 2018.

[14] Q. Wang and M. Skoglund, "Symmetric private information retrieval for MDS coded distributed storage," in *Proceedings of the IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6.

[15] R. Freij-Hollanti, O. W. Gnilke, C. Hollanti, and D. A. Karpuk, "Private information retrieval from coded databases with colluding servers," *SIAM Journal on Applied Algebra and Geometry*, vol. 1, no. 1, pp. 647–664, 2017.

[16] H. Sun and S. A. Jafar, "Private information retrieval from MDS coded data with colluding servers: Settling a conjecture by Freij-Hollanti et al." *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 1000–1022, 2018.

[17] Z. Jia and S. A. Jafar, "X-secure T-private information retrieval from mds coded storage with byzantine and unresponsive servers," *IEEE Transactions on Information Theory*, vol. 66, no. 12, pp. 7427–7438, 2020.

[18] H.-Y. Lin, S. Kumar, E. Rosnes, and A. G. i Amat, "An MDS-PIR capacity-achieving protocol for distributed storage using non-MDS linear codes," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 966–970.

[19] H. Sun and C. Tian, "Breaking the MDS-PIR capacity barrier via joint storage coding," *Information*, vol. 10, no. 9, p. 265, 2019.

[20] R. Zhou, C. Tian, T. Liu, and H. Sun, "Capacity-achieving private information retrieval codes from MDS-coded databases with minimum message size," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 370–374.

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

[22] Y. Zhang and G. Ge, "Private information retrieval from MDS coded databases with colluding servers under several variant models," *arXiv preprint arXiv:1705.03186*, 2017.

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

[24] R. Tajeddine, O. W. Gnilke, D. Karpuk, R. Freij-Hollanti, and C. Hollanti, "Robust private information retrieval from coded systems with byzantine and colluding servers," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 2451–2455.

[25] K. Banawan and S. Ulukus, "Private information retrieval through wiretap channel II: Privacy meets security," *IEEE Transactions on Information Theory*, vol. 66, no. 7, pp. 4129–4149, 2020.

[26] Q. Wang and M. Skoglund, "Secure private information retrieval from colluding databases with eavesdroppers," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 2456–2460.

[27] Q. Wang, H. Sun, and M. Skoglund, "The capacity of private information retrieval with eavesdroppers," *IEEE Transactions on Information Theory*, vol. 65, no. 5, pp. 3198–3214, 2018.

[28] R. Tandon, "The capacity of cache aided private information retrieval," in *Proceedings of the IEEE Annual Allerton Conference on Communication, Control, and Computing*, 2017, pp. 1078–1082.

[29] Y.-P. Wei, K. Banawan, and S. Ulukus, "Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching," *IEEE Transactions on Information Theory*, vol. 65, no. 5, pp. 3215–3232, 2018.

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

[31] A. Heidarzadeh, B. Garcia, S. Kadhe, S. El Rouayheb, and A. Sprintson, "On the capacity of single-server multi-message private information retrieval with side information," in *Proceedings of the IEEE 56th Annual Allerton Conference on Communication, Control, and Computing*, 2018, pp. 180–187.

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

[33] M. A. Attia, D. Kumar, and R. Tandon, "The capacity of private information retrieval from uncoded storage constrained databases," *IEEE Transactions on Information Theory*, vol. 66, no. 11, pp. 6617–6634, 2020.

[34] C. Tian, H. Sun, and J. Chen, "Capacity-achieving private information retrieval codes with optimal message size and upload cost," *IEEE Transactions on Information Theory*, vol. 65, no. 11, pp. 7613–7627, 2019.

[35] Q. Wang and M. Skoglund, "Symmetric private information retrieval from mds coded distributed storage with non-colluding and colluding servers," *IEEE Transactions on Information Theory*, vol. 65, no. 8, pp. 5160–5175, 2019.

[36] R. Tajeddine, O. W. Gnilke, D. Karpuk, R. Freij-Hollanti, and C. Hollanti, "Private information retrieval from coded storage systems with colluding, byzantine, and unresponsive servers," *IEEE Transactions on information theory*, vol. 65, no. 6, pp. 3898–3906, 2019.

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

[38] Z. Jia, H. Sun, and S. A. Jafar, "Cross subspace alignment and the asymptotic capacity of X-secure T-private information retrieval," *IEEE Transactions on Information Theory*, vol. 65, no. 9, pp. 5783–5798, 2019.

[39] S. Kumar, H.-Y. Lin, E. Rosnes, and A. G. i Amat, "Achieving maximum distance separable private information retrieval capacity with linear codes," *IEEE Transactions on Information Theory*, vol. 65, no. 7, pp. 4243–4273, 2019.

[40] N. Raviv and I. Tamot, "Private information retrieval is graph based replication systems," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 1739–1743.

[41] K. Banawan and S. Ulukus, "Private information retrieval from non-replicated databases," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 1272–1276.

[42] K. Banawan, B. Arasli, Y.-P. Wei, and S. Ulukus, "The capacity of private information retrieval from heterogeneous uncoded caching databases," *IEEE Transactions on Information Theory*, vol. 66, no. 6, pp. 3407–3416, 2020.

[43] Y.-P. Wei, B. Arasli, K. Banawan, and S. Ulukus, "Private information retrieval from decentralized uncoded caching databases," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 2114–2118.

[44] K. Banawan, B. Arasli, Y.-P. Wei, and S. Ulukus, "Private information retrieval from heterogeneous uncoded caching databases," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 1267–1271.

[45] I. Samy, M. A. Attia, R. Tandon, and L. Lazos, "Latent-variable private information retrieval," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 1071–1076.

[46] H. Sun and S. A. Jafar, "Blind interference alignment for private information retrieval," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2016, pp. 560–564.

[47] W.-T. Chang and R. Tandon, "On the capacity of secure distributed matrix multiplication," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–6.

[48] Z. Jia and S. A. Jafar, "On the capacity of secure distributed matrix multiplication," *arXiv preprint arXiv:1908.06957*, 2019.

[49] R. G. D'Oliveira, S. El Rouayheb, and D. Karpuk, "GASP codes for secure distributed matrix multiplication," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 1107–1111.

[50] W.-T. Chang and R. Tandon, "On the upload versus download cost for secure and private matrix multiplication," in *2019 IEEE Information Theory Workshop (ITW)*. IEEE, 2019, pp. 1–5.

[51] M. Aliasgari, O. Simeone, and J. Kliewer, "Distributed and private coded matrix computation with flexible communication load," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 1092–1096.

[52] Z. Wang, K. Banawan, and S. Ulukus, "Private set intersection: A multi-message symmetric private information retrieval perspective," *arXiv preprint arXiv:1912.13501*, 2019.

[53] D. Asonov and J.-C. Freytag, "Repudiative information retrieval," in *Proceedings of the ACM workshop on Privacy in the Electronic Society*, 2002, pp. 32–40.

[54] R. R. Toledo, G. Danezis, and I. Goldberg, "Lower-cost-private information retrieval," *Proceedings on Privacy Enhancing Technologies*, no. 4, pp. 184–201, 2016.

[55] I. Samy, R. Tandon, and L. Lazos, "On the capacity of leaky private information retrieval," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 1262–1266.

[56] H.-Y. Lin, S. Kumar, E. Rosnes, A. G. i Amat, and E. Yaakobi, "Weakly-private information retrieval," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 1257–1261.

[57] ——, "The capacity of single-server weakly-private information retrieval," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 415–427, 2021.

[58] T. Guo, R. Zhou, and C. Tian, "On the information leakage in private information retrieval systems," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2999–3012, 2020.

[59] Z. Jia, "On the capacity of weakly-private information retrieval," Ph.D. dissertation, UC Irvine, 2019.

APPENDIX A
PROOF OF COROLLARY 1

We first notice that for any $\epsilon > 0$, we have $\delta_1(\epsilon) \geq \delta_2(\epsilon)$. This follows as we can express $d_1(\epsilon, \delta_1(\epsilon)) = 1 + \delta_1(\epsilon)$ and $d_2(\epsilon, \delta_2(\epsilon)) = 1 + \delta_2(\epsilon)$. Then, from Theorems 1 and 2 and for any $\delta \geq \max(\delta_1(\epsilon), \delta_2(\epsilon))$, $D^*(\epsilon, \delta)$ can be bounded as follows

$$1 + \delta_2(\epsilon) = d_2(\epsilon, \delta_2(\epsilon)) = D^{LB}(\epsilon, \delta) \leq D^*(\epsilon, \delta)$$
$$\leq D^{UB}(\epsilon, \delta) = d_1(\epsilon, \delta_1(\epsilon)) = 1 + \delta_1(\epsilon), \quad (98)$$

which proves that $\delta_1(\epsilon)$ must be greater than or equal $\delta_2(\epsilon)$ for any value of $\epsilon \geq 0$.

Following that, we can write the multiplicative gap ratio between the upper and lower bounds on $D^*(\epsilon, \delta)$ given in (17) and (20) as follows:

$$\frac{D^{\text{UB}}(\epsilon, \delta)}{D^{\text{LB}}(\epsilon, \delta)} = \begin{cases} \frac{\gamma_1(\epsilon) - \gamma_2(\epsilon)\delta}{\gamma_3(\epsilon) - \gamma_4(\epsilon)\delta}, & \delta < \delta_2(\epsilon), \\[2ex] \frac{\gamma_1(\epsilon) - \gamma_2(\epsilon)\delta}{\gamma_3(\epsilon) - \gamma_4(\epsilon)\delta_2(\epsilon)}, & \delta_2(\epsilon) \leq \delta < \delta_1(\epsilon), \\[2ex] \frac{\gamma_1(\epsilon) - \gamma_2(\epsilon)\delta_1(\epsilon)}{\gamma_3(\epsilon) - \gamma_4(\epsilon)\delta_2(\epsilon)}, & \delta \geq \delta_1(\epsilon), \end{cases} \tag{99}$$

where we have $\gamma_1(\epsilon) = 1 + \frac{1}{N-1}$, $\gamma_2(\epsilon) = \frac{e^\epsilon}{N^{K-1}-1}$, $\gamma_3(\epsilon) = 1 + \frac{1}{Ne^\epsilon - 1}$, and $\gamma_4(\epsilon) = \frac{1}{(Ne^\epsilon)^{K-1}-1}$. Then, we can upper bound (99) as follows,

$$\frac{D^{\text{UB}}(\epsilon, \delta)}{D^{\text{LB}}(\epsilon, \delta)} \leq \begin{cases} \frac{\gamma_1(\epsilon) - \gamma_2(\epsilon)\delta}{\gamma_3(\epsilon) - \gamma_4(\epsilon)\delta}, & \delta < \delta_2(\epsilon), \\[2ex] \frac{\gamma_1(\epsilon) - \gamma_2(\epsilon)\delta_2(\epsilon)}{\gamma_3(\epsilon) - \gamma_4(\epsilon)\delta_2(\epsilon)}, & \delta \geq \delta_2(\epsilon). \end{cases} \tag{100}$$

For any δ, we have the bound $\frac{\gamma_1(\epsilon) - \gamma_2(\epsilon)\delta}{\gamma_3(\epsilon) - \gamma_4(\epsilon)\delta} \leq \frac{\gamma_1(\epsilon)}{\gamma_3(\epsilon)}$ valid when $\frac{\gamma_1(\epsilon)\gamma_4(\epsilon)}{\gamma_2(\epsilon)\gamma_3(\epsilon)} \leq 1$. We can prove that $\frac{\gamma_1(\epsilon)\gamma_4(\epsilon)}{\gamma_2(\epsilon)\gamma_3(\epsilon)} \leq 1$ in the following:

$$\frac{\gamma_1(\epsilon)\gamma_4(\epsilon)}{\gamma_2(\epsilon)\gamma_3(\epsilon)} = \frac{\left(1 + \frac{1}{N-1}\right)\frac{1}{(Ne^\epsilon)^{K-1}-1}}{\left(1 + \frac{1}{Ne^\epsilon - 1}\right)\frac{e^\epsilon}{N^{K-1}-1}} = \frac{\frac{N^{K-1}-1}{N-1}}{\frac{(Ne^\epsilon)^{K-1}-1}{Ne^\epsilon-1}} \cdot e^{-2\epsilon}$$

$$= \frac{\sum_{k=0}^{K-2} N^k}{\sum_{k=0}^{K-2}(Ne^\epsilon)^k} \cdot e^{-2\epsilon} \leq e^{-2\epsilon} \leq 1, \tag{101}$$

for any $\epsilon \geq 0$ with equality when $\epsilon = 0$. Eventually, we can bound the multiplicative gap ratio for any value of δ as

$$\frac{D^{\text{UB}}(\epsilon, \delta)}{D^{\text{LB}}(\epsilon, \delta)} \leq \frac{\gamma_1(\epsilon)}{\gamma_3(\epsilon)} = \frac{1 + \frac{1}{N-1}}{1 + \frac{1}{Ne^\epsilon - 1}} = \frac{N - e^{-\epsilon}}{N - 1}. \tag{102}$$

APPENDIX B
PROOF OF LEMMA 4

Assume that $A_n^{(k_1)}$ the answer of any DB_n, given any requested message $k_1 \in [1 : K]$, can take one of T different structures. Each of them is requested by a certain query, i.e., $Q_n^{(k_1)}$ also takes T different forms. Let π_t and $\gamma(\pi_t)$ be the tth form that $Q_n^{(k_1)}$ and $A_n^{(k_1)}$ can take, respectively. Then, $H(A_n^{(k_1)}|W_{k_1}, Q_n^{(k_1)})$ can be written as

$$H(A_n^{(k_1)}|W_{k_1}, Q_n^{(k_1)})$$

$$= \sum_{t=1}^{T} \Pr(Q_n^{(k_1)} = \pi_t) H(A_n^{(k_1)}|W_{k_1}, Q_n^{(k_1)} = \pi_t)$$

$$= \sum_{t=1}^{T} \Pr(Q_n^{(k_1)} = \pi_t) H(A_n^{(k_1)} = \gamma(\pi_t)|W_{k_1})$$

$$\overset{(a)}{=} \sum_{t=1}^{T} \Pr(Q_n^{(k_1)} = \pi_t) H(A_n^{(k_2)} = \gamma(\pi_t)|W_{k_1})$$

$$\overset{(b)}{\geq} \sum_{t=1}^{T} e^{-\epsilon} \Pr(Q_n^{(k_2)} = \pi_t) H(A_n^{(k_2)} = \gamma(\pi_t)|W_{k_1})$$

$$= \frac{1}{e^\epsilon} \sum_{t=1}^{T} \Pr(Q_n^{(k_2)} = \pi_t) H(A_n^{(k_2)} = \gamma(\pi_t)|W_{k_1})$$

$$= \frac{1}{e^\epsilon} \sum_{t=1}^{T} \Pr(Q_n^{(k_2)} = \pi_t) H(A_n^{(k_2)} = \gamma(\pi_t)|W_{k_1}, Q_n^{(k_2)} = \pi_t)$$

$$= \frac{1}{e^\epsilon} H(A_n^{(k_2)}|W_{k_1}, Q_n^{(k_2)}), \tag{103}$$

where (a) follows from the fact that the entropy of certain answer structure $\gamma(\pi_t)$ is independent of the requested message, it only depends on the query form π_t. Whereas, (b) comes from the definition in (9) and the corresponding interpretation in (57).

APPENDIX C
PROOF OF LEMMA 5

We can bound $H(A_{[1:N]}^{(k)}|W_{[1:k-1]}, Q_{[1:N]}^{(k)})$ as follows:

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

$$\overset{(a)}{=} H(A_{[1:N]}^{(k)}|W_{[1:k-1]}, Q_{[1:N]}^{(k)})$$
$$\quad + H(W_k|A_{[1:N]}^{(k)}, W_{[1:k-1]}, Q_{[1:N]}^{(k)}) - o(L)L$$

$$= H(W_k, A_{[1:N]}^{(k)}|W_{[1:k-1]}, Q_{[1:N]}^{(k)}) - o(L)L$$

$$= H(W_k|W_{[1:k-1]}, Q_{[1:N]}^{(k)})$$
$$\quad + H(A_{[1:N]}^{(k)}|W_{[1:k]}, Q_{[1:N]}^{(k)}) - o(L)L$$

$$= L + H(A_{[1:N]}^{(k)}|W_{[1:k]}, Q_{[1:N]}^{(k)}) - o(L)L$$

$$= (1 - o(L))L + H(A_{[1:N]}^{(k)}|W_{[1:k]}, Q_{[1:N]}^{(k)})$$

$$\geq (1 - o(L))L + H(A_n^{(k)}|W_{[1:k]}, Q_{[1:N]}^{(k)})$$

$$\overset{(b)}{=} (1 - o(L))L + H(A_n^{(k)}|W_{[1:k]}, Q_n^{(k)}), \quad \forall n \in [1 : N], \tag{104}$$

where (a) is due to the correctness property in (4), and (b) comes from the fact that the answer $A_n^{(k)}$ is conditionally independent of the queries submitted to other DBs given the query $Q_n^{(k)}$. By adding the relation in (104) over all possible n's and dividing by N, we get the following:

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

$$\geq (1 - o(L))L + \frac{1}{N} \sum_{n=1}^{N} H(A_n^{(k)}|W_{[1:k]}, Q_n^{(k)})$$

$$\overset{(a)}{\geq} (1 - o(L))L + \frac{1}{Ne^\epsilon} \sum_{n=1}^{N} H(A_n^{(k+1)}|W_{[1:k]}, Q_n^{(k+1)})$$

$$= (1 - o(L))L + \frac{1}{Ne^\epsilon} \sum_{n=1}^{N} H(A_n^{(k+1)}|W_{[1:k]}, Q_{[1:N]}^{(k+1)})$$

$$\geq (1 - o(L))L + \frac{1}{Ne^\epsilon} H(A_{[1:N]}^{(k+1)}|W_{[1:k]}, Q_{[1:N]}^{(k+1)}), \tag{105}$$

where (a) follows using similar steps as in the proof of Lemma 4.

APPENDIX D
PROOF OF PROPOSITION 3

Here, we prove the bound in proposition 3 for $N = 1$. We show that the relaxed privacy conditions have no benefits when there is only one database even if we ignore the DB privacy leakage constraint ($\delta = K - 1$). Assuming that the requested message is W_1, we lower bound $D_{\epsilon,\delta}$ as follows:

$$
\begin{aligned}
D_{\epsilon,\delta} &= H(A_1^{(1)}) \geq H(A_1^{(1)}|Q_1^{(1)}) \\
&= H(W_1, A_1^{(1)}|Q_1^{(1)}) - H(W_1|A_1^{(1)}, Q_1^{(1)}) \\
&\overset{(a)}{=} H(W_1|Q_1^{(1)}) + H(A_1^{(1)}|W_1, Q_1^{(1)}) - o(L)L \\
&= (1 - o(L))L + H(A_1^{(1)}|W_1, Q_1^{(1)}), \qquad (106)
\end{aligned}
$$

where (a) follows the correctness property in (4). Let there be T different structures, $\pi_1, \ldots, \pi_T$, the query sent to the databases can take. For each structure π_t, $t \in [1 : T]$, the answer is on the form of $\gamma(\pi_t)$ then we get, for $j \in [2 : K]$,

$$
\begin{aligned}
&H(A_1^{(1)}|W_1, Q_1^{(1)}) - H(A_1^{(j)}|W_1, Q_1^{(j)}) \\
&= \left(H(A_1^{(1)}, W_1|Q_1^{(1)}) - H(W_1|Q_1^{(1)}) \right) \\
&\quad - \left(H(A_1^{(j)}, W_1|Q_1^{(j)}) - H(W_1|Q_1^{(j)}) \right) \\
&= H(A_1^{(1)}, W_1|Q_1^{(1)}) - H(A_1^{(j)}, W_1|Q_1^{(j)}) \\
&= H(A_1^{(1)}|Q_1^{(1)}) + H(W_1|A_1^{(1)}, Q_1^{(1)}) \\
&\quad - H(A_1^{(j)}|Q_1^{(j)}) - H(W_1|A_1^{(j)}, Q_1^{(j)}) \\
&\overset{(a)}{=} H(A_1^{(1)}|Q_1^{(1)}) + o(L) - H(A_1^{(j)}|Q_1^{(j)}) \\
&\quad - \sum_{t=1}^{T} \Pr(Q_1^{(j)} = \pi_t) H(W_1|A_1^{(j)}, Q_1^{(j)} = \pi_t) \\
&= H(A_1^{(1)}|Q_1^{(1)}) + o(L) - H(A_1^{(j)}|Q_1^{(j)}) \\
&\quad - \sum_{t=1}^{T} \Pr(Q_1^{(j)} = \pi_t) H(W_1|A_1^{(j)} = \gamma(\pi_t)), \qquad (107)
\end{aligned}
$$

where (a) also comes from (4). We emphasize from the user privacy constraint in (9) that all queries or structures must be requested with non-zero probability, otherwise the constraint in (9) can not be met. This dictates that $\gamma(\pi_t)$, the answer of any structure π_t, has to fulfill the decodability conditions, i.e.,

$$
H(W_1|A_1^{(1)} = \gamma(\pi_t)) = o(L). \qquad (108)
$$

As the form of the answer $\gamma(\pi_t)$ is the same regardless of the requested message, this implies that

$$
H(W_1|A_1^{(j)} = \gamma(\pi_t)) = H(W_1|A_1^{(1)} = \gamma(\pi_t)) = o(L). \qquad (109)
$$

From (107) and (109), we get the following

$$
\begin{aligned}
&H(A_1^{(1)}|W_1, Q_1^{(1)}) \\
&= H(A_1^{(j)}|W_1, Q_1^{(j)}) + H(A_1^{(1)}|Q_1^{(1)}) - H(A_1^{(j)}|Q_1^{(j)}). \qquad (110)
\end{aligned}
$$

Assuming the symmetry across all messages, we have

$$
H(A_1^{(1)}|Q_1^{(1)}) = H(A_1^{(j)}|Q_1^{(j)}), \quad \forall j \in [2 : K]. \qquad (111)
$$

Using this fact, we have

$$
H(A_1^{(1)}|W_1, Q_1^{(1)}) = H(A_1^{(j)}|W_1, Q_1^{(j)}), \quad \forall j \in [2 : K]. \qquad (112)
$$

This allows us to write $D_{\epsilon,\delta}$ as follows

$$
\begin{aligned}
D_{\epsilon,\delta} &\geq (1 - o(L))L + H(A_1^{(2)}|W_1, Q_1^{(2)}) \\
&= (1 - o(L))L + H(W_2, A_1^{(2)}|W_1, Q_1^{(2)}) \\
&\quad - H(W_2|A_1^{(2)}, W_1, Q_1^{(2)}) \\
&= (1 - o(L))L + H(W_2, A_1^{(2)}|W_1, Q_1^{(2)}) - o(L)L \\
&= (1 - 2o(L))L + H(W_2|Q_1^{(2)}) + H(A_1^{(2)}|W_1, W_2, Q_1^{(2)}) \\
&= 2(1 - o(L))L + H(A_1^{(2)}|W_1, W_2, Q_1^{(2)}). \qquad (113)
\end{aligned}
$$

Completing the proof inductively using equations (107) to (112), we get

$$
\begin{aligned}
D_{\epsilon,\delta} &\geq (K - 1)(1 - o(L))L + H(A_1^{(K)}|W_{[1:K-1]}, Q_1^{(K)}) \\
&= (K - 1)(1 - o(L))L + H(W_K, A_1^{(K)}|W_{[1:K-1]}, Q_1^{(K)}) \\
&\quad - H(W_K|A_1^{(K)}, W_{[1:K-1]}, Q_1^{(K)}) \\
&= (K - 1)(1 - o(L))L + H(W_K, A_1^{(K)}|W_{[1:K-1]}, Q_1^{(K)}) \\
&\quad - o(L) \\
&\overset{(a)}{=} (K - 1)(1 - o(L))L + H(W_K|W_{[1:K-1]}, Q_1^{(K)}) \\
&\quad + H(A_1^{(K)}|W_{[1:K-1]}, Q_1^{(K)}) - o(L) \\
&\geq K(1 - o(L))L, \qquad (114)
\end{aligned}
$$

where (a) comes from the fact that the answer must be a function of the K messages. Dividing by L and taking the limit $L \to \infty$, we arrive at the desired lower bound:

$$
D^*(\epsilon, \delta) \geq K. \qquad (115)
$$

Islam Samy received the B.S. degree in electrical engineering from Alexandria University in 2011 and the M.Sc. degree in wireless communication from Nile University in 2014. He was awarded the Ph.D. degree in the Electrical and Computer Engineering from the The University of Arizona in 2021. His main research interests include secure and fair resource allocation for heterogeneous coexisting systems, wireless communication, and information theory.

Mohamed Attia received his B.Sc. degree in Electrical Engineering from Alexandria University, Alexandria, Egypt, in 2012, his M.Sc. degree in Electronics and Communications Engineering from the American University in Cairo, Egypt in 2015, and his PhD degree in Electrical and Computer Engineering from University of Arizona, Tucson, AZ, USA in 2020. He is currently working as a wireless systems engineer at Enterprise Connectivity, Advanced Engineering, Ford Motor Company in Dearborn, MI, USA. His research interests include Information Theory, Machine Learning, Wireless Communications, private information retrieval, and Distributed Data Analysis.

Ravi Tandon is an Associate Professor in the Department of ECE at the University of Arizona. Prior to joining the University of Arizona in Fall 2015, he was a Research Assistant Professor at Virginia Tech with positions in the Bradley Department of ECE, Hume Center for National Security and Technology and at the Discovery Analytics Center in the Department of Computer Science. He received the B.Tech. degree in Electrical Engineering from the Indian Institute of Technology, Kanpur (IIT Kanpur) in 2004 and the Ph.D. degree in Electrical and Computer Engineering from the University of Maryland, College Park (UMCP) in 2010. From 2010 to 2012, he was a post-doctoral research associate at Princeton University. He is a recipient of the 2018 Keysight Early Career Professor Award, NSF CAREER Award in 2017, and a Best Paper Award at IEEE GLOBECOM 2011. He is a Senior Member of IEEE and currently serves as an Editor for IEEE Transactions on Wireless Communications and IEEE Transactions on Communications. His current research interests include information theory and its applications to wireless networks, communications, security and privacy, machine learning and data mining.

Loukas Lazos is a Professor of Electrical and Computer Engineering at the University of Arizona. Dr. Lazos received his Ph.D. degree in Electrical Engineering from the University of Washington in 2006. His research interests are in the areas of network security, privacy, and wireless communications. His current research focuses on secret-free authentication methods for IoT devices, integrity verification of physical properties, secure and fair channel access for emerging wireless technologies, and fair resource allocation for heterogeneous coexisting technologies, and private information retrieval. Dr. Lazos is the recipient of the US National Science Foundation (NSF) Faculty Early CAREER Development Award (2009) for his research in security of multi-channel wireless networks. He has served as a technical program chair for the IEEE CNS conference, the IEEE GLOBECOM symposium on communications and information systems security and the IEEE DSPAN workshop. He is an associate editor for the IEEE Transactions on Information and Forensics Security journal and the IEEE Transactions on Mobile Computing journal. He has also served and continues to serve on the organization and technical program committees of many international conferences and on expert panels of several government agencies.