The Asymptotic Capacity of Private Search

Zhen Chen, *Student Member, IEEE*, Zhiying Wang, *Member, IEEE*, and Syed Ali Jafar, *Fellow, IEEE*

***Abstract*— The private search problem is introduced, where a dataset comprised of L i.i.d. records is replicated across N non-colluding servers, and a user wishes to search for all records that match a privately chosen value, without revealing any information about the chosen value to any individual server. Each record contains P symbols, and each symbol takes values uniformly and independently from an alphabet of size K. Considering the large number of records in modern datasets, it is assumed that L is much larger than the alphabet size K. The capacity of private search is the maximum number of bits of desired information that can be retrieved per bit of download. The asymptotic (large K) capacity of private search is shown to be $1 - 1/N$, even when the scope of private search is further generalized to allow OR search, AND search, NOT search and sequence search. The results are based on the asymptotic behavior of a new converse bound for private information retrieval with arbitrarily dependent messages. The asymptotic behavior is also applicable to T-colluding servers or (N, T)-MDS coded servers.**

***Index Terms*— Private search, asymptotic capacity, private information retrieval.**

I. Introduction

SEARCH is among the most frequent operations performed on large online datasets. With privacy concerns increasingly taking center stage in online interactions, a private search functionality is highly desirable. As a basic formulation of the information-theoretically private search problem, consider a *dataset* that is replicated across N non-colluding servers. There are L i.i.d. records in the dataset, each record is comprised of P symbols, and each symbol is from an alphabet of size K. A basic form of private search, called *exact* private search, allows a user to privately choose one symbol from the alphabet, and then search for all records that contain this symbol, without revealing any information about the queried symbol. Suppose the record length P is a constant, and $L \gg K \gg 1$, i.e., the alphabet size K is large, but the number of records in the dataset is much larger. This is not an uncommon scenario. For example, consider datasets of DNA sequences. When searching for a DNA pattern of length ℓ (e.g., $\ell = 10$), the alphabet size is $K = 4^\ell$, while current DNA sequencing machines produce millions of records (called reads) per run. Since the upload cost of private search can be made independent of L while the download cost scales linearly with L, the communication cost of private search for large L is dominated by the download cost. The *capacity* of private search is therefore defined as the maximum number of bits of desired information that can be retrieved per bit of download. Furthermore, since $K \gg 1$, the *asymptotic capacity* of private search, i.e., the capacity for large K is of particular interest. Characterizing the asymptotic capacity of private search is our main goal in this work.

Private search (PS) has been studied in computer science for decades. One branch focuses on designing searchable encryption schemes, which enable users to store encrypted data at the servers and execute search over ciphertext domain [2], [3]. Encryption preserves the user privacy computationally. Various models of search functionality have been explored in this framework, such as keyword search [2], [3], similarity search [4], [5], OR and AND search [6], [7] and ranked search [8]. Another branch allows servers to store unencrypted data, and relies on private information retrieval (PIR) [9] schemes to guarantee the privacy of the user's query. Problems investigated in this framework include keyword search [10], streaming data search [11]–[13], and media search [14], [15]. Our work is along the latter line and tries to characterize the asymptotic capacity of private search. Recall that in its original form as introduced by Chor et al. in [9], the goal of PIR is to allow a user to retrieve an arbitrary desired message out of μ *independent* messages that are replicated across N distributed and non-colluding servers, without revealing any information about the identity of the desired message to any individual server. The capacity of PIR is the maximum number of bits of desired information that can be retrieved per bit of download, and was shown in [16] to be $\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{\mu-1}}\right)^{-1}$. The capacity of many variants of PIR has since been characterized, such as PIR with colluding servers [17], PIR with coded servers [18]–[20], symmetric PIR [21], [22], PIR with side information [23]–[26] and multi-message PIR [27].

Particularly relevant to this paper is the generalized form of PIR introduced in [28], [29], known as the private computation problem [28] or the private function retrieval problem [29]–[31]. As its main result, [28] establishes the capacity of PIR when the messages have arbitrary linear dependencies. A supplementary result of [28] shows that even if non-linear dependencies are allowed, the asymptotic capacity of private computation approaches $1 - 1/N$ provided that the message set includes an unbounded number of independent messages. Some other types of private computations are also

Manuscript received February 15, 2019; revised October 30, 2019; accepted February 11, 2020. Date of publication March 2, 2020; date of current version July 14, 2020. This work was supported in part by NSF under Grant CNS-1731384 and Grant CCF-1907053, and in part by the Office of Naval Research (ONR) under Grant N00014-18-1-2057. This article was presented in part at the 2018 IEEE International Symposium on Information Theory (ISIT). *(Corresponding author: Zhen Chen.)*

The authors are with the Center for Pervasive Communications and Computing (CPCC), University of California Irvine, Irvine, CA 92697 USA (e-mail: zhenc4@uci.edu; zhiying@uci.edu; syed@uci.edu).

Communicated by J. Kliewer, Associate Editor for Coding Techniques.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2020.2977082

investigated, i.e., private polynomial computation [32] which allows polynomial relationships among messages, and private inner product retrieval [33] which considers the inner product of messages. Private search is a form of PIR with a specific form of dependency among messages, that is not covered by these prior works. This is because in private search the dependencies among messages are neither linear nor of a polynomial form, and no two messages are independent. To see this clearly, consider exact search with alphabet set $\{A, B, C\}$ (which implies $K = 3$). Assume there are $L = 4$ records, A, A, B, C, each of size $P = 1$. We search for all records that match a queried symbol. Denote the retrieved message for a query by W_θ, for some $\theta \in [3]$, which is comprised of 4 i.i.d. bits, i.e., $W_\theta = (W_\theta(1), W_\theta(2), W_\theta(3), W_\theta(4))$, such that $W_\theta(l) = 1$ if the l-th record matches the queried symbol, and $W_\theta(l) = 0$ otherwise. For example, if A is queried, the corresponding message $W_1 = (1, 1, 0, 0)$. If B is queried, the corresponding message $W_2 = (0, 0, 1, 0)$. If C is queried, the corresponding message $W_3 = (0, 0, 0, 1)$. It is easily seen that any two messages, W_i, W_j, $i \neq j$, are identically distributed but not independent, e.g., $W_i(l) = 1$ implies $W_j(l) = 0$. This dependency is neither linear nor in a polynomial form. To approach the private search problem, we first consider a broader generalization of PIR to include messages with arbitrary dependencies (DPIR in short). Then we consider private search as a special case of DPIR.

Since simple keyword search often yields far too coarse results, almost all the search engines such as Google, Bing, Yahoo, Linkedin and Facebook, and large database management systems like MySQL and PostgreSQL support OR search, AND search and NOT search. These searches allow a broader range of search operations by connecting various pieces of information with OR, AND or NOT operators to make the search more precise. For example, instead of retrieving all emails from "Alice", a user might only want those emails from Alice that are marked urgent and pertain to finance, in which case what is needed is the ability to search on the conjunction of the keywords, "Alice" and "urgent" and "finance" [6]. In other cases, it is desirable to search for symbols that appear in consecutive positions in a record, e.g., to search for a phrase. Therefore as natural generalizations of exact private search, we also consider OR private search, AND private search, NOT private search and sequence private search. OR private search looks for all records which contain any of M symbols, AND private search looks for all records which contain all of M symbols, and NOT search looks for all records which do not contain the chosen symbol. Finally, sequence private search allows the user to search for all records that contain an M-symbol long sequence.

Our main contributions are as follows.

- We start with a general non-asymptotic converse for dependent private information retrieval or DPIR (Theorem 1). Converse here denotes a lower bound on the download cost, or equivalently, an upper bound on the capacity.
- The converse combined with a general achievability result for DPIR that was established in [28], leads us to a sufficient condition under which the asymptotic capacity of DPIR is characterized to be $1 - 1/N$ (Theorem 3).
- The sufficient condition of Theorem 3 is shown to hold for exact private search, thus establishing the asymptotic capacity of private search as $1 - 1/N$ (Theorem 4).
- We show that the sufficient condition of Theorem 3 also holds for OR search, AND search, NOT search and sequence search, so that the asymptotic capacity for these generalizations is also equal to $1 - 1/N$ (Theorem 4). Remarkably, for OR search, the asymptotic capacity characterization holds even when M itself grows with K.
- Finally, to illustrate the difficulty of finding general asymptotic capacity results for DPIR, we consider an example of OR private search with special restricted search patterns. For this example, we show that either the new converse bound is not tight, or the asymptotic capacity is not $1 - 1/N$ (Proposition 1). The asymptotic capacity for this example remains open.

The paper is organized as follows. Section II presents the problem statement. The download lower bound of DPIR and the asymptotic capacity of various forms of private search are characterized in Section III. Section IV presents the proofs of the results. Section V concludes the paper with a discussion of generalization of our settings, including extending Theorem 3 to T-colluding DPIR and MDS-coded DPIR, and the non-asymptotic capacity of private search.

Notation: We use parentheses $(a_1, a_2, \ldots, a_n)$ to represent a vector or a tuple (sequence) and braces $\{s_1, s_2, \ldots, s_N\}$ to represent a set. $[z_1 : z_2]$ represents the set $\{z_1, z_1+1, \cdots, z_2\}$, for $z_1, z_2 \in \mathbb{N}$, $z_1 < z_2$, $[z]$ represents $[1 : z]$ for $z \in \mathbb{N}$. Let $W_1, W_2, \ldots$ be random variables, and $S = \{s_1, s_2, \ldots, s_N\}$ be a subset of indices where $s_1 < s_2 < \cdots < s_N$. The random vector $(W_{s_1}, W_{s_2}, \ldots, W_{s_N})$ is represented by W_S. $A \sim B$ means that random vectors A and B are identically distributed. A function $f(L) = o(L)$ means that $\lim_{L\to\infty} f(L)/L = 0$. $o(L)$ can be positive or negative. A function $f(L) = O(L)$ means that $\lim_{L\to\infty} |f(L)| / L \leq c$, for some constant $c > 0$.

II. Problem Statement

A. Dependent Private Information Retrieval (DPIR)

Consider $\mu \in \mathbb{N}$ messages, $W_m, m \in [\mu]$, each comprised of L symbols, $W_m = (W_m(1), W_m(2), \cdots, W_m(L))$. The random vectors $(W_1(l), W_2(l), \cdots, W_\mu(l))$, for all $l \in [L]$, are i.i.d., and have a distribution that is identical to the distribution of the random vector $(w_1, w_2, \cdots, w_\mu)$. Namely, for different l, the vectors $(W_1(l), W_2(l), \cdots, W_\mu(l))$ are independent; but for any particular l, the variables $W_m(l), m \in [\mu]$ have dependencies defined by the joint distribution of $w_m, m \in [\mu]$.

Example 1: For $L = 2, \mu = 3$, let (w_1, w_2, w_3) be a binary random vector and the distribution be $p(w_1, w_2, w_3) = 1/3$ for $(w_1, w_2, w_3) \in \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$, and $p(w_1, w_2, w_3) = 0$ for all other cases. One realization of the messages can be:

W_1	1	0
W_2	0	0
W_3	0	1

The first bit (column) and the second bit (column) of the messages are independent. Within the first bit (column), only one entry can be 1.

The amount of information carried by the m-th message, $m \in [\mu]$, is

$$H(W_m) = LH(w_m). \tag{1}$$

We say that the DPIR problem is *balanced* if all messages $W_m, \forall m \in [\mu]$ carry the same amount of information,

$$H(W_1) = H(W_2) = \cdots = H(W_\mu) \triangleq LH(w), \tag{2}$$

i.e. $\forall m \in [\mu], H(w_m) = H(w)$.

We note here that the random variable w may depend on the number of messages μ, especially in the context of private search, i.e., $H(w) = H(w(\mu))$. For compact notation, we will not explicitly show the dependence on μ.

The problem of DPIR is as follows. There are N servers and each server stores all μ messages. A user privately generates $\theta \in [\mu]$ and wishes to retrieve W_θ while keeping θ private from each server. Depending on θ, the user employs N queries $Q_1^{[\theta]}, \cdots, Q_N^{[\theta]}$ and sends $Q_n^{[\theta]}$ to the n-th server. The n-th server returns a response string $A_n^{[\theta]}$ which is a function of $Q_n^{[\theta]}$ and $W_{[\mu]}$, i.e.,

$$\forall \theta \in [\mu], \forall n \in [N],\ H(A_n^{[\theta]} \mid Q_n^{[\theta]}, W_{[\mu]}) = 0. \tag{3}$$

From all the information that is now available to the user, he must be able to decode the desired message W_θ, with probability of error $P_e \to 0$ as $L \to \infty$. This is called the "correctness" constraint. From Fano's inequality, we have

$$\text{[Correctness]}\ H\left(W_\theta \mid A_{[N]}^{[\theta]}, Q_{[N]}^{[\theta]}\right) = o(L). \tag{4}$$

To protect the user's privacy, θ must be indistinguishable from θ', from the perspective of each server, $\forall \theta, \theta' \in [\mu]$, i.e.,

$$\text{[Privacy]}\ (Q_n^{[\theta]}, A_n^{[\theta]}, W_{[\mu]}) \sim (Q_n^{[\theta']}, A_n^{[\theta']}, W_{[\mu]}). \tag{5}$$

The DPIR *rate* characterizes how many bits of desired information are retrieved per downloaded bit, and is limited by the worst case as,

$$R \triangleq \frac{\min_{m\in[\mu]} LH(w_m)}{D}, \tag{6}$$

where D is the expected total number of bits downloaded by the user from all the servers. If the DPIR problem is balanced, then the minimum over m may be ignored. The supremum of achievable rates R is the *capacity* $C_{DPIR}(\mu, N)$.

B. Private Search

We first define exact search and OR search. Later we define AND search, NOT search and sequence search. Examples of different kinds of search are given in Table I.

TABLE I

Example of Different Types of Private Search. The Dataset Contains $L = 3$ Records (A, B, C), (A, C, B) and (B, B, B). Each Record Contains $P = 3$ Symbols

Type	Pattern	(A, B, C)	(A, C, B)	(B, B, B)	W_m
exact	A	1	1	0	110
OR	A or B	1	1	1	111
AND	A and B	1	1	0	110
NOT	not A	0	0	1	001
sequence	(A, B)	1	0	0	100

1) Exact Search and OR Search: Consider a dataset Δ comprised of L i.i.d. records: $\Delta = (\Delta_1, \Delta_2, \cdots, \Delta_L)$. Each record Δ_l, $l \in [L]$, is a sequence of length P, where P is constant, denoted by $(\delta_{l1}, \delta_{l2}, \cdots, \delta_{lP})$, and each symbol δ_{li} takes values uniformly and independently from the alphabet set $\mathcal{U} = \{U_1, U_2, \cdots, U_K\}$. The dataset is replicated across N non-colluding servers.

For all $l \in [L]$, $\delta_{l_i} \in \mathcal{U}$, $i \in [P]$,

$$P(\Delta_l = (\delta_{l1}, \delta_{l2}, \cdots, \delta_{lP})) = \frac{1}{K^P}, \tag{7}$$

$$H(\Delta) = LH(\Delta_l) = L\log_2(K^P) = LP\log_2 K \text{ bits}. \tag{8}$$

A user privately chooses a set (search pattern), $S = \{U_{\theta_1}, U_{\theta_2}, \cdots, U_{\theta_M}\}$, $S \subset \mathcal{U}$, $M < K$, and searches for all records in Δ that contain at least one element of S. Note that even though each record is of length P, given a search pattern S the search result for a record is only a single bit, indicating whether the P symbols in the record contain an element in S or not. The overall search result for the dataset is L (independent) bits. We refer to the $M = 1$ setting as *exact* private search, and to the $M > 1$ setting as *OR* private search, because the output of the search reveals the exact value of a matching record if $M = 1$, but not if $M > 1$. In general, for OR search we allow M to grow with K (either $o(K)$ or $\Omega(K)$) in the asymptotic regime $K \to \infty$.

To view private search as a special case of DPIR, we treat the result of each possible search pattern as one message. A similar technique has been also used in the work of Fanti [15]. There are a total of $\mu = \binom{K}{M}$ search patterns. Let us arbitrarily label them $S_m, m \in [\mu]$. Correspondingly, there are a total of μ messages. Label these messages W_m, so that $\forall m \in [\mu]$,

$$W_m = (W_m(1), W_m(2), \cdots, W_m(L)), \tag{9}$$

and

$$W_m(l) = \begin{cases} 1, & \text{if } \exists i \in [M],\ U_{\theta_i} \in \{\delta_{l1}, \cdots, \delta_{lP}\}, \\ 0, & \text{otherwise.} \end{cases}$$

Note that each message is comprised of L i.i.d. bits. $\forall l \in [L]$,

$$H(w) = H(W_m(l)) = H_2\left(\frac{(K-M)^P}{K^P}\right), \forall m \in [\mu], \tag{10}$$

where the binary entropy function is defined as follows.

$$H_2(p) = -p\log_2(p) - (1-p)\log_2(1-p), \tag{11}$$

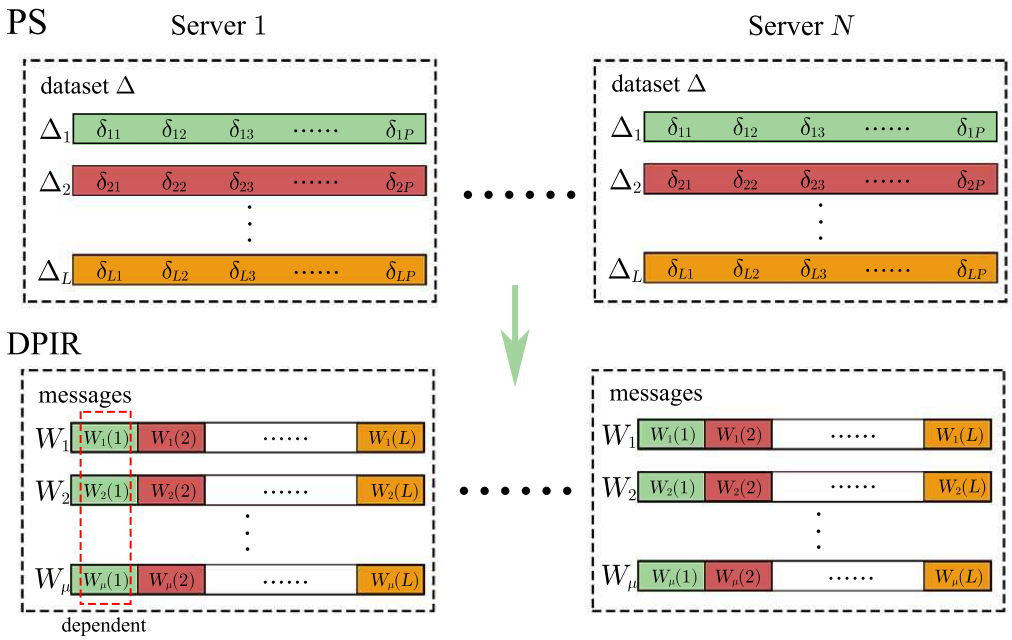

Fig. 1. Relationship between PS and DPIR. The i-th record (row) Δ_i in PS corresponds to the i-th symbol (column) of each message in DPIR. Different messages (rows) in DPIR correspond to different search patterns S.

$H_2(0) = H_2(1) = 0$. The second equation in (10) is based on the facts that for each record there are a total of K^P possible realizations and $(K-M)^P$ of those do not match. Fig. 1 shows the relationship between private search and DPIR. For example, suppose there are $L = 2$ records of length $P = 1$, the alphabet is $\mathcal{U} = \{A, B, C\}$ of size $K = 3$, and we do exact search ($M = 1$). Let the records be $\Delta_1 = A$, $\Delta_2 = C$. Then the $\mu = 3$ messages are shown as in Example 1. See Table I for additional examples.

2) AND Search: For AND private search, a set $S_m = \{U_{\theta_1}, U_{\theta_2}, \cdots, U_{\theta_M}\}$ is chosen out of a total of $\mu = \binom{K}{M}$ possibilities. In general, M can be arbitrary. However, if $M > P$ the problem degenerates into a trivial case where no record can contain all of the chosen symbols. Therefore, we only consider the non-trivial case where $1 \leq M \leq P$. For all $m \in [\mu], l \in [L]$, the l^{th} bit of the corresponding message W_m is defined as,

$$W_m(l) = \begin{cases} 1, & \text{if } \forall i \in [M],\ U_{\theta_i} \in \{\delta_{l1}, \cdots, \delta_{lP}\}, \\ 0, & \text{otherwise.} \end{cases}$$

The L-bits of each message are i.i.d., and $\forall l \in [L], \forall m \in [\mu], H(w) = H(W_m(l))$. See Table I for an example.

3) NOT Search: For NOT private search, a user privately chooses a value $S_m = \{U_\theta\}$ out of $\mu = K$ possibilities. The l^{th} symbol of the corresponding message W_m is defined as

$$W_m(l) = \begin{cases} 1, & \text{if } U_\theta \notin \{\delta_{l1}, \cdots, \delta_{lP}\}, \\ 0, & \text{otherwise.} \end{cases}$$

The L bits of each message are i.i.d. and $\forall l \in [L], \forall m \in [\mu], H(w) = H(W_m(l))$. Essentially NOT search is the complement of exact search. For example, in terms of the same chosen symbol, if $W_m(l) = 1$ in exact search, then $W_m(l) = 0$ in NOT search, and vice versa. See Table I for an example.

4) Sequence Search: Sequence private search is similar to AND search, the difference is that the order of symbols matters in sequence search. Specifically, a sequence $S_m = (U_{\theta_1}, U_{\theta_2}, \cdots, U_{\theta_M})$ is chosen, out of $\mu = K^M$ possibilities. For the same reason as AND search, we only consider the non-trivial scenario where $1 \leq M \leq P$. For all $m \in [\mu], l \in [L]$, the l^{th} symbol of the corresponding message W_m,

$$W_m(l) = \begin{cases} 1, & \text{if } S_m \in \{(\delta_{l_{i+1}}, \cdots, \delta_{l_{i+M}}), i \in [0 : P-M]\}, \\ 0, & \text{otherwise.} \end{cases}$$

The L-bits of each message are i.i.d., and $\forall l \in [L], \forall m \in [\mu], H(w) = H(W_m(l))$. See Table I for an example.

Even though in our definitions of private search, we assume that all search sets S (or search sequence S) of size M are allowed, one can generalize the definition to restricted search patterns. One example of such a setting is discussed in Section III-E.

The queries and answers, privacy and correctness constraints, rate and capacity definitions for private search are inherited from DPIR. The *capacity of private search* is denoted $C_{PS}(K, M, P, N)$, and the asymptotic capacity of private search is denoted $\lim_{K\to\infty} C_{PS}(K, M, P, N)$.

III. Results

We present the main results in this section. All proofs appear in Section IV.

A. A General Converse for DPIR

The download cost (expected number of bits of download) for DPIR is bounded as follows.

Theorem 1: For DPIR, denote by $W_1, W_2, \ldots, W_\mu$ an arbitrary permutation of the μ messages. Then

$$D \geq H(W_1) + \frac{H(W_2|W_1)}{N} + \frac{H(W_3|W_{[1:2]})}{N^2} + \cdots + \frac{H(W_\mu|W_{[1:\mu-1]})}{N^{\mu-1}}. \quad (12)$$

Note that the bound depends on the chosen permutation of message indices, so finding the best bound from Theorem 1 requires a further optimization of the permutation. Substituting (12) into (6), we obtain an equivalent bound on capacity. If the messages are independent, we recover the converse bound of [16]. However, Theorem 1 is more broadly useful since it allows arbitrary dependencies. Also note that Theorem 1 is not limited to balanced DPIR.

B. General Achievable Rate for DPIR [28]

A PIR achievable scheme for independent messages is also a DPIR achievable scheme. We use the achievable PIR scheme with $\mu \to \infty$ messages from [28, Theorem 2] (also see [9], [21], [34]), from which we obtain the following lower bound on the capacity of DPIR.

$$C_{\text{DPIR}}(\mu, N) \geq \lim_{\mu\to\infty} C_{\text{DPIR}}(\mu, N) \geq \lim_{\mu\to\infty} C_{\text{PIR}}(\mu, N) = \left(1 - \frac{1}{N}\right) \frac{\min_{m\in[\mu]} H(w_m)}{\max_{m\in[\mu]} H(w_m)}. \quad (13)$$

Theorem 2: The capacity of DPIR satisfies

$$C_{\text{DPIR}}(\mu, N) \geq \left(1 - \frac{1}{N}\right) \frac{H_{\min}}{H_{\max}}, \quad (14)$$

where $H_{\min} = \min_{m\in[\mu]} H(w_m)$ and $H_{\max} = \max_{m\in[\mu]} H(w_m)$.

For balanced DPIR, this gives us $1-1/N$ as a lower bound on capacity. As a simple example of the achievable scheme, assume there are $N = 2$ servers and $\mu = 2$ messages with the size of 1 bit. A user wishes to retrieve W_1. He generates two binary random variables α and β independently, and sends (α, β) to server 1 and $(\alpha+1, \beta)$ to server 2. Here "+" denotes XOR. He downloads $\alpha W_1 + \beta W_2$ and $(\alpha + 1)W_1 + \beta W_2$ from server 1 and server 2, respectively, which allows him to retrieve W_1 privately. Therefore the total download is 2 bits, and $R = 1/2 = 1-1/N$. This achievable scheme requires one multiplication for each symbol of each message. In general, for each server, the computation complexity is $O(\mu L)$.

C. Asymptotic Optimality of Rate $1 - 1/N$ for Balanced DPIR

For balanced DPIR, as the number of messages $\mu \to \infty$, the asymptotic behavior of (12) gives us the following sufficient condition. Here we define $W_k = 0$ if $k > \mu$, and define a sequence function to be a sequence of functions $k_1(\mu), k_2(\mu), \cdots$, where every $k_i, i \geq 1$, is a mapping $\mathbb{N} \to \mathbb{N}$. When it is clear from the context, we drop the variable μ and simply use k_i to denote $k_i(\mu)$. But one should keep in mind that k_i depends on the number of messages μ.

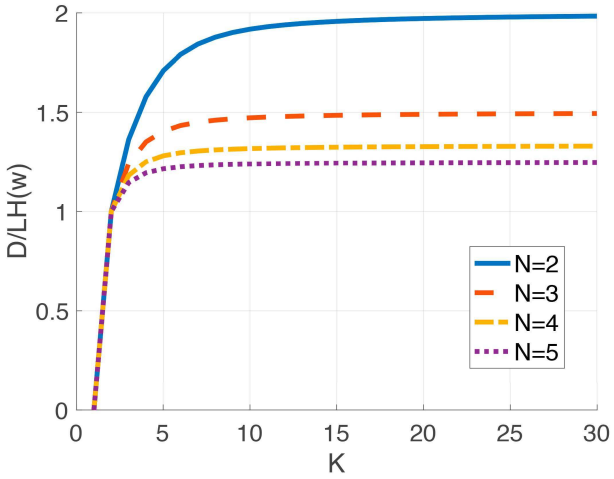

Fig. 2. Normalized download lower bound of exact search ($P = 1$) based on Theorem 1 versus alphabet size K. The asymptotic value $(1 - 1/N)^{-1}$ is the upper bound.

Theorem 3: For balanced DPIR, if there exists an increasing sequence $k_i \in \mathbb{N}, \forall i \in \mathbb{N}$, such that $\forall l \in \mathbb{N}$,

$$\lim_{\mu\to\infty} \frac{I\left(W_{k_{l+1}}; W_{k_{[1:l]}}\right)}{LH(w)} = 0, \quad (15)$$

then the asymptotic capacity is

$$\lim_{\mu\to\infty} C_{\text{DPIR}}(\mu, N) = 1 - \frac{1}{N}. \quad (16)$$

Note since $H(w)$ may depend on μ, the sufficient condition is in general not equivalent to $\lim_{\mu\to\infty} I\left(W_{k_{l+1}}; W_{k_{[1:l]}}\right) = 0$. In particular, (15) provides a measure of "weak" dependency among the messages in the asymptotic regime, such that the capacity of DPIR is $1-1/N$. Intuitively, if we find an infinite sequence of messages that have this weak dependency in DPIR, we know the asymptotic capacity is $1 - 1/N$.

D. Asymptotic Capacity of Private Search

Theorem 4: The asymptotic capacity of private search is

$$\lim_{K\to\infty} C_{PS}(K, M, P, N) = 1 - \frac{1}{N}, \quad (17)$$

for exact search ($M = 1$), NOT search ($M = 1$), OR search ($M > 1$), AND search ($1 \leq M \leq P$) and sequence private search ($1 \leq M \leq P$). For OR search, M can even grow with K, satisfying either $M = o(K)$ or $M = \Omega(K)$.

Theorem 4 is proved by showing that the sufficient condition (15) is satisfied for private search. Note that condition (15) is explicitly proven to be true for balanced DPIR, and private search indeed has balanced messages. Notably, for exact private search, as $K \to \infty$, both $I(W_{k_{l+1}}; W_{k_{[1:l]}})$ and $H(w)$ approach zero. The key to the asymptotic capacity result is that $I(W_{k_{l+1}}; W_{k_{[1:l]}})$ approaches zero much faster than $H(w)$. Furthermore, as shown in Fig. 2, convergence of capacity to its asymptotic value is quite fast, and the larger the value of N, the faster the convergence. For example, when the record size $P = 1$ and the number of servers $N = 5$, the bound (12) for $K = 10$ messages is already within 1% gap from the asymptotic value.

E. Difficulty of Private Search Over Restricted Search Patterns

Finding the capacity of DPIR with arbitrary dependency structures is in general a difficult problem. The difficulty remains even when the problem is limited to asymptotic capacity. To highlight this aspect, we present an example of approximate private search over restricted search patterns where the asymptotic capacity remains an open problem.

Proposition 1: Consider OR private search, with $P = 1$ and $M = \lfloor \frac{K}{2} \rfloor$, where the only search sets allowed are

$$S_k = \{U_{\langle k+1 \rangle}, U_{\langle k+2 \rangle}, \cdots, U_{\langle k+M \rangle}\}, \quad \forall k \in [K], \tag{18}$$

and $\langle m \rangle \triangleq (m \bmod K) + 1$. As $K \to \infty$, either the bound (12) is not tight, or $\lim_{K\to\infty} C_{PS}(K, M, P, N) \neq 1 - \frac{1}{N}$.

Here privacy is required only within the $\mu = K$ choices of search sets.

IV. Proofs

A. Proof of Theorem 1

For the DPIR problem, the total download is bounded as,

$$D \geq H\left(A_{[N]}^{[1]} \mid Q_{[N]}^{[1]}\right) \tag{19}$$

$$\overset{(4)}{=} H\left(A_{[N]}^{[1]}, W_1 \mid Q_{[N]}^{[1]}\right) + o(L) \tag{20}$$

$$= H\left(W_1 \mid Q_{[N]}^{[1]}\right) + H\left(A_{[N]}^{[1]} \mid Q_{[N]}^{[1]}, W_1\right) + o(L) \tag{21}$$

$$\geq H\left(W_1\right) + H\left(A_1^{[1]} \mid Q_{[N]}^{[1]}, W_1\right) + o(L) \tag{22}$$

$$= H\left(W_1\right) + H\left(A_1^{[1]} \mid Q_1^{[1]}, W_1\right) + o(L) \tag{23}$$

$$\overset{(5)}{=} H(W_1) + H\left(A_1^{[2]} \mid Q_1^{[2]}, W_1\right) + o(L) \tag{24}$$

$$= H(W_1) + H\left(A_1^{[2]} \mid Q_{[N]}^{[2]}, W_1\right) + o(L), \tag{25}$$

where (22) follows because the queries are independent of W_1, and $A_1^{[1]}$ is an element of $A_{[N]}^{[1]}$, and (23) follows from the fact $A_1^{[1]}$ and $Q_{[2:N]}^{[1]}$ are independent given $Q_1^{[1]}$. Similarly, for all $n \in [2 : N]$ we have,

$$D \geq H(W_1) + H\left(A_n^{[2]} \mid Q_{[N]}^{[2]}, W_1\right). \tag{26}$$

Adding all of these N inequalities we obtain,

$$ND \geq NH(W_1) + H\left(A_{[N]}^{[2]} \mid Q_{[N]}^{[2]}, W_1\right) \tag{27}$$

$$\Rightarrow D \geq H(W_1) + \frac{H\left(A_{[N]}^{[2]} \mid Q_{[N]}^{[2]}, W_1\right)}{N}. \tag{28}$$

Proceeding recursively in a similar manner as (28), $\forall m \in [2 : \mu - 1]$, we have

$$H\left(A_{[N]}^{[m]} \mid Q_{[N]}^{[m]}, W_1, \cdots, W_{m-1}\right) \geq H(W_m \mid W_1, \cdots, W_{m-1}) + \frac{H\left(A_{[N]}^{[m+1]} \mid Q_{[N]}^{[m+1]}, W_1, \cdots, W_m\right)}{N} \tag{29}$$

and when $m = \mu$,

$$H\left(A_{[N]}^{[\mu]} \mid Q_{[N]}^{[\mu]}, W_1, \cdots, W_{\mu-1}\right) \geq H(W_\mu \mid W_1, \cdots, W_{\mu-1}). \tag{30}$$

Therefore,

$$D \geq H(W_1) + \frac{H\left(A_{[N]}^{[2]} \mid Q_{[N]}^{[2]}, W_1\right)}{N} \tag{31}$$

$$\geq H(W_1) + \frac{H(W_2 \mid W_1)}{N} + \frac{H\left(A_{[N]}^{[3]} \mid Q_{[N]}^{[3]}, W_{[1:2]}\right)}{N^2} \tag{32}$$

$$\geq \cdots$$

$$\geq H(W_1) + \frac{H(W_2 \mid W_1)}{N} + \frac{H(W_3 \mid W_{[1:2]})}{N^2} + \cdots + \frac{H(W_\mu \mid W_{[1:\mu-1]})}{N^{\mu-1}}. \tag{33}$$

B. Proof for Theorem 3

Define m such that $k_m \leq \mu < k_{m+1}$. Note that m is a function of μ and as $\mu \to \infty$, $m \to \infty$. Based on Theorem 1 and equations (1), (2),

$$\begin{aligned} D &\geq H(W_{k_1}) + \frac{H(W_{k_2} \mid W_{k_1})}{N} + \\ &\quad + \frac{H(W_{k_3} \mid W_{k_1}, W_{k_2})}{N^2} + \cdots + \frac{H\left(W_{k_m} \mid W_{k_{[1:m-1]}}\right)}{N^{m-1}} \\ &= H(W_{k_1}) + \frac{H(W_{k_2})}{N} + \cdots + \frac{H(W_{k_m})}{N^{m-1}} \\ &\quad - \frac{I(W_{k_2}; W_{k_1})}{N} - \cdots - \frac{I\left(W_{k_m}; W_{k_{[1:m-1]}}\right)}{N^{m-1}} \\ &= (1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{m-1}}) LH(w) \\ &\quad - \frac{I(W_{k_2}; W_{k_1})}{N} - \cdots - \frac{I\left(W_{k_m}; W_{k_{[1:m-1]}}\right)}{N^{m-1}}. \end{aligned} \tag{34}$$

Normalizing both sides by $LH(w)$ we have

$$\frac{D}{LH(w)} = \left(1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{m-1}}\right) - \frac{I(W_{k_2}; W_{k_1})}{NLH(w)} - \cdots - \frac{I\left(W_{k_m}; W_{k_{[1:m-1]}}\right)}{N^{m-1}LH(w)}. \tag{35}$$

Applying limit $\mu \to \infty$, the reciprocal of rate is bounded as

$$\lim_{\mu\to\infty} \frac{D}{LH(w)} \geq \left(1 - \frac{1}{N}\right)^{-1} - \lim_{\mu\to\infty} \sum_{l=1}^{m-1} \frac{I\left(W_{k_{l+1}}; W_{k_{[1:l]}}\right)}{LH(w)N^l}.$$

Now, we need to show that

$$\lim_{\mu\to\infty} \sum_{l=1}^{m-1} \frac{I\left(W_{k_{l+1}}; W_{k_{[1:l]}}\right)}{LH(w)N^l} = 0. \tag{36}$$

Equivalently, for every $\epsilon > 0$ we will show that

$$\lim_{\mu\to\infty} \sum_{l=1}^{m-1} \frac{I\left(W_{k_{l+1}}; W_{k_{[1:l]}}\right)}{LH(w)N^l} \leq \epsilon. \tag{37}$$

Choose a finite l^* such that

$$\frac{1}{N^{l^*}} \left(1 - \frac{1}{N}\right)^{-1} \leq \epsilon. \tag{38}$$

Note that l^* depends only on N and ϵ. More importantly, it is not a function of μ. Now partition the sum as follows

$$\lim_{\mu\to\infty}\sum_{l=1}^{m-1}\frac{I\left(W_{k_{l+1}};W_{k_{[1:l]}}\right)}{LH(w)N^l} = \lim_{\mu\to\infty}\sum_{l=1}^{l^*-1}\frac{I\left(W_{k_{l+1}};W_{k_{[1:l]}}\right)}{LH(w)N^l} + \lim_{\mu\to\infty}\sum_{l=l^*}^{m-1}\frac{I\left(W_{k_{l+1}};W_{k_{[1:l]}}\right)}{LH(w)N^l}. \tag{39}$$

The first term on the RHS of (39) is zero because it is a sum of finitely many terms (l^* is finite), each of which is zero because (15) holds by assumption. For the second term in (39),

$$\lim_{\mu\to\infty}\sum_{l=l^*}^{m-1}\frac{I\left(W_{k_{l+1}};W_{k_{[1:l]}}\right)}{LH(w)N^l} \le \lim_{\mu\to\infty}\sum_{l=l^*}^{m-1}\frac{1}{N^l} \tag{40}$$

$$\le \frac{1}{N^{l^*}}\lim_{\mu\to\infty}\sum_{l=0}^{m-1-l^*}\frac{1}{N^l} \tag{41}$$

$$\le \frac{1}{N^{l^*}}\left(1-\frac{1}{N}\right)^{-1} \le \epsilon. \tag{42}$$

Thus, the reciprocal of rate is bounded as $1/R \ge (1-1/N)^{-1}$, i.e., the rate is bounded as $R \le 1-1/N$. By Theorem 2 this rate is achievable. Hence proved.

C. Proof of Theorem 4

We treat private search as a balanced DPIR problem. As an application of Theorem 3, we show that (15) is satisfied. Therefore the asymptotic capacity must be $1-1/N$. Note that for all the private search variations, the number of messages $\mu \to \infty$ if and only if the alphabet size $K \to \infty$. So in our proofs we let K grow to infinity.

In the following proofs, we consider a subset of the possible messages, $W_1, W_2, \ldots, W_{f(\mu)}$, for some $f(\mu) \le \mu$ that grows with μ. We use the identity sequence functions $k_i = i$ for $1 \le i \le f(\mu)$, and map i to some $k_i > \mu$ for $i > f(\mu)$. In other words, we only use the first $f(\mu)$ messages in our proofs. Then (15) becomes

$$\lim_{\mu\to\infty}\frac{I\left(W_{l+1};W_{[1:l]}\right)}{LH(w)} = 0 \tag{43}$$

for all $1 \le l \le f(\mu)$.

1) Exact Private Search: We start with the exact private search problem ($M = 1$).

Firstly, consider the case where the record length $P = 1$, note that

$$\lim_{K\to\infty} H(w) = \lim_{K\to\infty} H_2\left(\frac{1}{K}\right) = 0. \tag{44}$$

According to L′Hôpital′s rule,

$$\lim_{K\to\infty}\frac{H_2\left(\frac{1}{K-t}\right)}{H_2\left(\frac{1}{K}\right)} = 1, \tag{45}$$

where t is a constant. The detailed proof of (45) is shown in Appendix A.

Since $\forall l \in [K], W_l(1), \cdots, W_l(L)$ are i.i.d., $H(W_l) = LH\left(W_l(\eta)\right)$, η can be any integer between 1 to L. Consider the dependence among the messages,

$$\frac{H\left(W_{l+1} \mid W_{[1:l]}\right)}{L} = H\left(W_{l+1}(\eta) \mid W_{[1:l]}(\eta)\right) \tag{46}$$

$$= Pr\left(W_{[1:l]}(\eta) = \mathbf{0}\right)\cdot H\left(W_{l+1}(\eta) \mid W_{[1:l]}(\eta) = \mathbf{0}\right) + \sum_{i=1}^{l} Pr\left(W_i(\eta) = 1, W_{[1:l]\setminus\{i\}}(\eta) = \mathbf{0}\right) \cdot H\left(W_{l+1}(\eta) \mid W_i(\eta) = 1, W_{[1:l]\setminus\{i\}}(\eta) = \mathbf{0}\right) \tag{47}$$

$$= \left(1-\frac{l}{K}\right)H_2\left(\frac{1}{K-l}\right) + \sum_{i=1}^{l}\frac{1}{K}\cdot 0 \tag{48}$$

$$= \left(1-\frac{l}{K}\right)H_2\left(\frac{1}{K-l}\right), \tag{49}$$

where $Pr(e)$ is the probability of event e and bold $\mathbf{0}$ is the zero vector. The above equalities are explained as below. The only possible values for $W_{[1:l]}(\eta)$ are either all zeros or 1 one and $l-1$ zeros. Note that the probability $Pr\left(W_{[1:l]}(\eta) = \mathbf{0}\right) = 1 - l/K$. If $W_i(\eta) = 0, \forall i \in [l]$, then $\Delta_\eta \ne U_1, \cdots, U_l$ and Δ_η can only take values from $\{U_{l+1}, U_{l+2}, \cdots, U_K\}$, each with probability $1/(K-l)$. Therefore, conditioning on $W_{[1:l]}(\eta) = \mathbf{0}$, we have $\Delta_\eta = U_{l+1}$, i.e., $W_{l+1}(\eta) = 1$, with probability $1/(K-l)$, and $\Delta_\eta \ne U_{l+1}$, i.e., $W_{l+1}(\eta) = 0$, with probability $1 - 1/(K-l)$. If $W_1(\eta) = 1$, then $\Delta_\eta = U_1$ and $W_2(\eta), \cdots, W_K(\eta)$ must be equal to zero. Thus there is at most one $W_i(\eta) = 1$ and each $W_i(\eta) = 1$ with probability $1/K$. If any $W_i(\eta) = 1$, then $H\left(W_j(\eta) \mid W_i(\eta)\right) = 0, \forall j \ne i$.

Substituting $\mu = K$ into the LHS of (43), we have for any fixed $l \in \mathbb{N}$,

$$\lim_{K\to\infty}\frac{I(W_{l+1};W_1,W_2,\cdots,W_l)}{LH_2\left(\frac{1}{K}\right)} \tag{50}$$

$$= \lim_{K\to\infty}\frac{H(W_{l+1}) - H(W_{l+1} \mid W_{[1:l]})}{LH_2(\frac{1}{K})} \tag{51}$$

$$= \lim_{K\to\infty}\frac{H_2\left(\frac{1}{K}\right) - \left(1-\frac{l}{K}\right)H_2\left(\frac{1}{K-l}\right)}{H_2\left(\frac{1}{K}\right)} \tag{52}$$

$$= 1 - \lim_{K\to\infty}\frac{\left(1-\frac{l}{K}\right)H_2\left(\frac{1}{K-l}\right)}{H_2\left(\frac{1}{K}\right)} = 1-1 = 0. \tag{53}$$

Here (53) follows from (45). Therefore, (43) is satisfied, and based on Theorem 3, the asymptotic capacity is $1-1/N$.

Then consider the case where $P > 1$ is a constant, note that

$$\lim_{K\to\infty} H(w) = \lim_{K\to\infty} H_2\left(\frac{(K-1)^P}{K^P}\right) = 0. \tag{54}$$

According to L′Hôpital′s rule,

$$\lim_{K\to\infty}\frac{H_2\left(\frac{(K-t-1)^P}{(K-t)^P}\right)}{H_2\left(\frac{(K-1)^P}{K^P}\right)} = 1, \tag{55}$$

where t is a constant. The detailed proof of (55) is shown in Appendix B. For any fixed $l \in \mathbb{N}$, we denote by E_i the event that i entries out of $W_1(\eta), \cdots, W_l(\eta)$ are 1, and the remaining $l-i$ entries are 0. Let its probability be

$\tau_i = Pr(E_i)$. Since each record size is P, there are at most P of $W_i(\eta)$ equal to 1, hence $0 \le i \le \min(P, l)$.

$$H(W_{l+1} \mid W_l, \cdots, W_1) \tag{56}$$
$$= LH(W_{l+1}(\eta) \mid W_l(\eta), \cdots, W_1(\eta)) \tag{57}$$
$$= \sum_{i=0}^{\min(P,l)} \binom{l}{i} \tau_i LH(W_{l+1}(\eta) \mid E_i) \tag{58}$$
$$\ge \binom{l}{0} \tau_0 LH(W_{l+1}(\eta) \mid E_0) \tag{59}$$
$$= \tau_0 LH(W_{l+1}(\eta) \mid E_0). \tag{60}$$

Note that

$$\tau_0 = Pr(E_0) = \frac{(K-l)^P}{K^P}, \tag{61}$$

and conditioned on E_0, $W_{l+1}(\eta)$ is 0 with probability $(K - l - 1)^P/(K - l)^P$, thus

$$H(W_{l+1}(\eta) \mid E_0) = H_2\left(\frac{(K-l-1)^P}{(K-l)^P}\right). \tag{62}$$

Therefore,

$$H(W_{l+1} \mid W_l, \cdots, W_1) \ge \tau_0 LH(W_{l+1}(\eta) \mid E_0) \tag{63}$$
$$= \frac{L(K-l)^P}{K^P} H_2\left(\frac{(K-l-1)^P}{(K-l)^P}\right). \tag{64}$$

Substituting $\mu = K$ into the LHS of (43), we have

$$\lim_{K\to\infty} \frac{I(W_{l+1}; W_1, W_2, \cdots, W_l)}{LH_2\left(\frac{(K-1)^P}{K^P}\right)} \tag{65}$$
$$= \lim_{K\to\infty} \frac{H(W_{l+1}) - H(W_{l+1} \mid W_{[1:l]})}{LH_2(\frac{(K-1)^P}{K^P})} \tag{66}$$
$$\le \lim_{K\to\infty} \frac{H_2(\frac{(K-1)^P}{K^P}) - \frac{(K-l)^P}{K^P} H_2\left(\frac{(K-l-1)^P}{(K-l)^P}\right)}{H_2(\frac{(K-1)^P}{K^P})} \tag{67}$$
$$= 1 - \lim_{K\to\infty} \frac{\frac{(K-l)^P}{K^P} H_2\left(\frac{(K-l-1)^P}{(K-l)^P}\right)}{H_2(\frac{(K-1)^P}{K^P})} = 1 - 1 = 0. \tag{68}$$

Here (68) follows from (55). Therefore, (43) is satisfied, and based on Theorem 3, the asymptotic capacity is $1 - 1/N$.

In summary, for arbitrary constant P the asymptotic capacity of exact private search is $1 - 1/N$.

2) NOT Search: Since NOT search essentially is the complement of exact search, the asymptotic capacity of NOT search is $1 - 1/N$.

3) OR Search: For OR search, first we show that any $P > 1$ OR search problems can be viewed as $P = 1$ OR search problem. For example, suppose the alphabet set is $\{A, B, C\}$, i.e., $K = 3$ and record length $P = 2$. A user wishes to search for A or B. It is equivalent to the problem that the alphabet set is all ordered tuples consisting of $\{A, B, C\}$, i.e. $\{AA, AB, AC, BA, BB, BC, CA, CB, CC\}$ and the record length is $P' = 1$. Notice that every character in the new alphabet set is searched for with the same probability. The user wishes to search for AA or AB or AC or BA or BB or BC or CA or CB. In general, for OR search problem with alphabet size K, record length $P > 1$ and search set size M, it is equivalent to the OR search problem with alphabet size $K' = K^P$, record length $P' = 1$ and search set size $M' = K' - (K - M)^P$. Note that if $M = o(K)$,

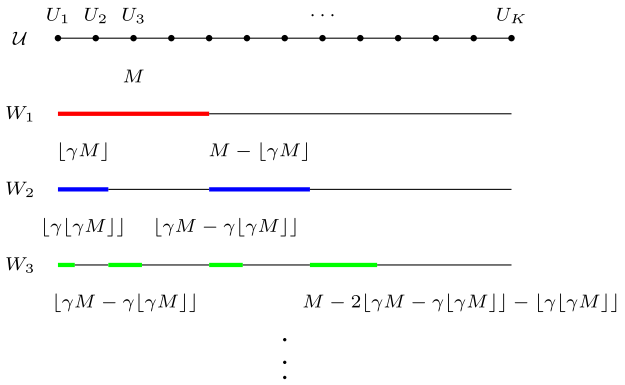

Fig. 3. Partition of the alphabet to obtain a sequence of dependent messages for OR search, $M = \Omega(K), P = 1$. Here $\gamma = M/K$. The alphabet $\mathcal{U} = \{U_1, U_2, \cdots, U_K\}$ is represented on a straight line.

$$\lim_{K\to\infty} \frac{M'}{K'} = \lim_{K\to\infty} \frac{K^P - (K-M)^P}{K^P} = 0, \tag{69}$$

i.e. $M' = o(K')$. Similarly, if $M = \Omega(K), M' = \Omega(K')$. Therefore in the following proof of OR private search, we only consider the case $P = 1$.

Define $\gamma \triangleq M/K < 1$. When $M = o(K)$, regard each M-element set as one new symbol and consider messages corresponding to disjoint search patterns. For example, suppose the alphabet set is $\{1, 2, \cdots, K\}$, $M = 2$, regard $\{1, 2\}, \{1, 3\}, \{2, 3\}, \cdots$ as new symbols. Consider the messages corresponding to $\{1, 2\}, \{3, 4\}, \{5, 6\}, \cdots$. There are $K' = \lfloor K/2 \rfloor$ such messages. As $K \to \infty$, $K' \to \infty$. Based on the proof for the exact search setting, these messages satisfy (15). Therefore the asymptotic capacity is $1 - 1/N$. For the general case, consider the $K' = \lfloor K/M \rfloor$ messages corresponding to disjoint search patterns. Invoking Theorem 3 we conclude that the asymptotic capacity is $1 - 1/N$.

For $M = \Omega(K)$, by symmetry of the truth function, searching for a given set is the same as searching for its complement. The entropy $H_2(\gamma) = H_2(1-\gamma)$ and the capacity as a function of γ, is symmetric around $\gamma = 1/2$. Thus we only need to consider $\gamma = M/K \le 1/2$. Let us find a sequence of dependent messages such that (43) is satisfied. Choose W_1 corresponding to $S_1 = \{U_1, U_2, \cdots, U_M\}$. It separates the alphabet set $\mathcal{U}$ into 2 parts: S_1 of size γK, and $\mathcal{U} \backslash S_1$ of size $(1-\gamma)K$. Note that $\gamma K = M$ is an integer. Choose the second message W_2 so that it is comprised of $\lfloor \gamma M \rfloor$ elements of S_1 and $M - \lfloor \gamma M \rfloor$ elements of $\mathcal{U} \backslash S_1$. Repeating this step we get a series of dependent messages, as in Fig. 3.

Note that

$$H(W_l) = LH_2(\gamma), \ \forall l. \tag{70}$$

Since $\gamma \le 1/2$, $\frac{\gamma M - 1}{M} \le \frac{\lfloor \gamma M \rfloor}{M} \le 1/2$. In terms of function $H_2(x)$, $\frac{\gamma M - 1}{M}$ and $\frac{\lfloor \gamma M \rfloor}{M}$ are both in the monotonically

increasing range. When $\frac{M-\lfloor \gamma M \rfloor}{K-M} \leq 1/2$,

$$H(W_2|W_1)$$
$$= LH_2\left(\frac{\lfloor \gamma M \rfloor}{M}\right)\frac{M}{K} + LH_2\left(\frac{M-\lfloor \gamma M \rfloor}{K-M}\right)\frac{K-M}{K} \tag{71}$$
$$\geq LH_2\left(\frac{\gamma M - 1}{M}\right)\frac{M}{K} + LH_2\left(\frac{\gamma(K-M)-1}{K-M}\right)\frac{K-M}{K} \tag{72}$$
$$= LH_2\left(\frac{\gamma^2 K - 1}{\gamma K}\right)\frac{M}{K} + LH_2\left(\frac{\gamma(1-\gamma)K - 1}{(1-\gamma)K}\right)\frac{K-M}{K}. \tag{73}$$

Then we have

$$\lim_{K\to\infty} H(W_2|W_1) \geq LH_2(\gamma) = H(W_1), \tag{74}$$
$$\Rightarrow \lim_{K\to\infty} H(W_2|W_1) = H(W_1). \tag{75}$$

When $\frac{M-\lfloor \gamma M \rfloor}{K-M} \geq 1/2$, $\frac{M-\lfloor \gamma M \rfloor}{K-M}$ and $\frac{\gamma(K-M)-1}{K-M}$ are in non-monotonic range, (72) is still true. Due to the symmetry, we only need to show $\frac{M-\lfloor \gamma M \rfloor}{K-M}$ is closer to $1/2$ than $\frac{\gamma(K-M)-1}{K-M}$. We have

$$\frac{M-\lfloor \gamma M \rfloor}{K-M} - \frac{1}{2}$$
$$= \frac{M-\lfloor \gamma M \rfloor}{K-M} - \gamma + \gamma - \frac{1}{2} \tag{76}$$
$$= \frac{M-\lfloor \gamma M \rfloor}{K-M} - \frac{M-\gamma M}{K-M} + \gamma - \frac{1}{2} \tag{77}$$
$$= \frac{\gamma M - \lfloor \gamma M \rfloor}{K-M} - \frac{1}{2} + \gamma \tag{78}$$
$$\leq \frac{1}{K-M} - \frac{1}{2} + \gamma \tag{79}$$
$$\leq \frac{1}{K-M}. \tag{80}$$

And

$$\frac{1}{2} - \frac{M-\gamma M-1}{K-M}$$
$$= \frac{1}{2} - \frac{M-\gamma M-1}{K-M} + \gamma - \gamma \tag{81}$$
$$= \frac{1}{2} - \frac{M-\gamma M-1}{K-M} + \frac{M-\gamma M}{K-M} - \gamma \tag{82}$$
$$= \frac{1}{2} + \frac{1}{K-M} + \gamma \tag{83}$$
$$\geq \frac{1}{K-M}. \tag{84}$$

Combining (80) and (84), (72) is satisfied.

Since $M = \Omega(K)$, there exists a constant $0 < c < 1$ such that $\gamma = M/K \geq c$ for sufficiently large K. For a given K, consider the search of only the restricted messages $\{W_l : l \leq \log_{1/c}\sqrt{K}\}$. Note that the number of the restricted messages goes to infinity as $K \to \infty$. Next we prove

$$\lim_{K\to\infty} \frac{H(W_l|W_{[l-1]})}{LH_2(\gamma)} = 1, \ \forall l \leq \log_{1/c}\sqrt{K}. \tag{85}$$

According to our choice of the message W_l in Fig. 3, we partition the alphabet into 2^l parts at step l. Thus there are 2^{l-1} terms in $H(W_l|W_{[l-1]})$. In particular, $\forall i \in [2^{l-1}]$, the i-th term corresponds to the event that the record symbol is in the i-th part, and we use ξ_i to denote its probability. To bound the i-th term, let us use a binary number to represent $i-1$. Let the number of "1"s in the binary number be m_i and $m_i \in [l-1]$. For example, if $l = 4$ and $i = 2$, then $i - 1 = (001)_2$, and $m_i = 1$. The size of the i-th part is between $\gamma^{l-m_i}(1-\gamma)^{m_i}K - l + 1$ and $\gamma^{l-m_i}(1-\gamma)^{m_i}K + l - 1$. Then the i-th term of $H(W_l|W_1, \cdots, W_{l-1})$ is greater than or equal to

$$LH_2\left(\frac{\gamma^{l-m_i+1}(1-\gamma)^{m_i}K - l + 1}{\gamma^{l-m_i}(1-\gamma)^{m_i}K + l - 1}\right)\cdot \xi_i \tag{86}$$
$$= LH_2\left(\frac{\gamma - \frac{l-1}{\gamma^{l-m_i}(1-\gamma)^{m_i}K}}{1 + \frac{l-1}{\gamma^{l-m_i}(1-\gamma)^{m_i}K}}\right)\cdot \xi_i. \tag{87}$$

When $K \to \infty$, $\forall i \in [2^{l-1}]$, $l \leq \log_{1/c}\sqrt{K}$,

$$\lim_{K\to\infty} \frac{l-1}{\gamma^{l-m_i}(1-\gamma)^{m_i}K} \leq \lim_{K\to\infty} \frac{l-1}{\gamma^l K} = 0. \tag{88}$$

Therefore,

$$\lim_{K\to\infty} LH_2\left(\frac{\gamma - \frac{l-1}{\gamma^{l-m_i}(1-\gamma)^{m_i}K}}{1 + \frac{l-1}{\gamma^{l-m_i}(1-\gamma)^{m_i}K}}\right) = \lim_{K\to\infty} LH_2(\gamma). \tag{89}$$

Summing up all the terms, we obtain

$$\lim_{K\to\infty} H(W_l|W_1, \cdots, W_{l-1}) \geq \lim_{K\to\infty} LH_2(\gamma), \tag{90}$$
$$\Rightarrow \lim_{K\to\infty} H(W_l|W_1, \cdots, W_{l-1}) = \lim_{K\to\infty} H(W_l). \tag{91}$$

Invoking Theorem 3 at this point, we conclude that the asymptotic capacity is $1 - 1/N$.

4) AND Private Search: For AND search, the record size $P \geq M$, otherwise no record matches. Similar to OR search, we translate it to OR search with the record size $P' = 1$. For example, consider the alphabet set is $\{A, B, C\}$, i.e., $K = 3$ and record length $P = 2$. A user wishes to search for A and B. It is equivalent to the problem that the alphabet set is all ordered tuples consisting of $\{A, B, C\}$, i.e. $\{AA, AB, AC, BA, BB, BC, CA, CB, CC\}$ and the record length is $P' = 1$. The user wishes to search for AB or BA.

In general case, for AND search problem with alphabet size K, search set size M and record length $P \geq M$, it is equivalent to the OR search problem with alphabet size $K' = K^P$, record length $P' = 1$ and search set size $M' = o(K')$. Here M' is the number of matching cases for an AND search, which is a function of P, M, K, i.e. $M' = \Gamma(P, M, K)$. To show $M' = o(K')$, we first calculate the value of $\Gamma(P, M, K)$. Notice the fact that there are two cases where a P-symbol length record matches: 1) The first $P-1$ symbols already contain all of the M chosen symbols, and 2) The first $P-1$ symbols only contain $M-1$ chosen symbols. For the first case, the last symbol can be any one in the alphabet while for the second case, the last symbol must be the missing symbol and any one of the M symbols could

be missing. Thus, the value of $\Gamma(P, M, K)$ can be calculated by recursive equation

$$\Gamma(P, M, K) = K\Gamma(P-1, M, K) + M\Gamma(P-1, M-1, K-1), \quad (92)$$

with base cases

$$\Gamma(P, 1, K) = K^P - (K-1)^P, \quad (93)$$
$$\Gamma(M, M, K) = M!. \quad (94)$$

Recall that P and M are constants, which do not grow with K. When $K \to \infty$, $\Gamma(P, 1, K) = o(K^P)$ and $\Gamma(M, M, K) = O(1)$. Suppose $\Gamma(P-1, M, K) = o(K^{P-1})$ and $\Gamma(P-1, M-1, K-1) = o((K-1)^{P-1})$,

$$\begin{aligned}\Gamma(P, M, K) &= K \cdot o\left(K^{P-1}\right) + M \cdot o\left((K-1)^{P-1}\right) \\ &= o\left(K^P\right). \end{aligned} \quad (95)$$

Based on mathematical induction, $\forall M$ and $\forall P \geq M$,

$$\Gamma(P, M, K) = o(K^P) = o(K'). \quad (96)$$

Since the asymptotic capacity of OR search is $1 - 1/N$, the asymptotic capacity of AND search is $1 - 1/N$.

5) Sequence Private Search: For sequence search, the non-trival case is under the condition $P \geq M$, otherwise no record matches. Suppose the chosen sequence is the tuple $S = (U_{\theta_1}, \cdots, U_{\theta_M})$. Note that here U_{θ_i} and U_{θ_j} can be the same symbol even through $i \neq j$. For sequence search, again we translate it to OR search with the record size $P' = 1$. Consider the same example where alphabet set is $\{A, B, C\}$, i.e., $K = 3$ and record length $P = 2$. A user wishes to search for a sequence AB. It is equivalent to the problem that the alphabet set is all ordered tuples consisting of $\{A, B, C\}$, i.e. $\{AA, AB, AC, BA, BB, BC, CA, CB, CC\}$ and the record length is $P' = 1$. The user wishes to search for AB.

In general case, for sequence search problem with alphabet size K, search set size M and record length $P \geq M$, it is equivalent to the OR search problem with alphabet size $K' = K^P$, record length $P' = 1$ and search set size $M' = o(K')$. Here M' is the number of matching cases for a sequence search, which is a function of P, M, K, i.e. $M' = \Psi(P, M, K)$. To show $M' = o(K')$, note that if a sequence is contained in a record, every character in the sequence must be contained in that record,

$$\Psi(P, M, K) \leq \Gamma(P, m, K) = o(K^P) = o(K'), \quad (97)$$

where $m = |\{U_{\theta_1}, \cdots, U_{\theta_M}\}| \leq M$ is the number of distinct searched symbols. Therefore the asymptotic capacity of sequence search is $1 - 1/N$.

D. Proof of Proposition 1

Consider the even values of K as it approaches infinity so that we have $H(W_k(\eta)) = H(1/2) = 1$ bit, i.e., each message bit is marginally uniform. We prove the proposition by contradiction. Suppose the asymptotic capacity is $1 - \frac{1}{N}$, namely, $\lim_{K\to\infty} \frac{D}{LH(1/2)} = \left(1 - \frac{1}{N}\right)^{-1}$, and suppose the bound (12) is tight for some sequence $k_1, k_2, \cdots$. Note that

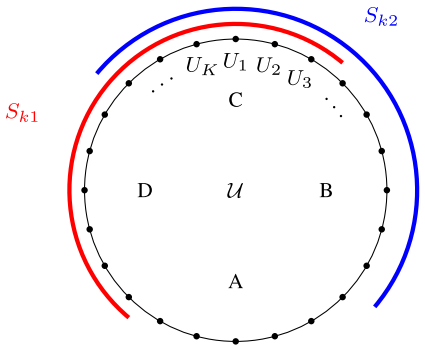

Fig. 4. Alphabet circle.

message W_{k_i} corresponds to search set S_{k_i}. Then we have the following equation.

$$\begin{aligned}&\lim_{K\to\infty} 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots = \lim_{K\to\infty} \frac{D}{LH(\frac{1}{2})} = \lim_{K\to\infty} \frac{D}{L} \\ &\overset{(12)}{=} \lim_{K\to\infty} H(W_{k_1}(\eta)) + \frac{1}{N} H(W_{k_2}(\eta) \mid W_{k_1}(\eta)) \\ &+ \frac{1}{N^2} H\left(W_{k_3}(\eta) \mid W_{k_1}(\eta), W_{k_2}(\eta)\right) + \ldots \end{aligned} \quad (98)$$

Therefore,

$$\begin{aligned}0 = &\lim_{K\to\infty} \frac{1}{N} \left(1 - H(W_{k_2}(\eta) \mid W_{k_1}(\eta))\right) \\ &+ \frac{1}{N^2} \left(1 - H\left(W_{k_3}(\eta) \mid W_{k_1}(\eta), W_{k_2}(\eta)\right)\right) + \cdots,\end{aligned}$$

which implies that

$$\lim_{K\to\infty} H\left(W_{k_2}(\eta) \mid W_{k_1}(\eta)\right) = 1, \quad (99)$$
$$\lim_{K\to\infty} H\left(W_{k_3}(\eta) \mid W_{k_1}(\eta), W_{k_2}(\eta)\right) = 1. \quad (100)$$

Let us represent $U_1, U_2, \cdots, U_K$ on an alphabet circle $\mathcal{U}$ shown in Fig. 4.

Since S_{k_1} is a contiguous set of $K/2$ points on the circle, without loss of generality it may be represented by the red semi-circle. $W_{k_1}(\eta)$ and $W_{k_2}(\eta)$ are binary random variables. So if $\lim_{K\to\infty} H\left(W_{k_2}(\eta) \mid W_{k_1}(\eta)\right) = 1$, then

$$\lim_{K\to\infty} H\left(W_{k_2}(\eta) \mid W_{k_1}(\eta) = 0\right) = 1, \quad (101)$$
$$\lim_{K\to\infty} H\left(W_{k_2}(\eta) \mid W_{k_1}(\eta) = 1\right) = 1. \quad (102)$$

This is equivalent to, within S_{k_1} half of the points must be in S_{k_2} and half of the points must be outside S_{k_2}, when K approaches infinity. Similar for the points outside S_{k_1}. Therefore, without loss of generality, S_{k_2} is represented by the blue semi-circle on the alphabet circle. Note that this divides the alphabet circle into 4 parts, labeled as A, B, C, D, corresponding to $(W_{k_1}(\eta), W_{k_2}(\eta)) = (0,0), (0,1), (1,1), (1,0)$, respectively. Note that each of these spans $K/4 + o(K)$ points.

Since $\lim_{K\to\infty} H\left(W_{k_3}(\eta) \mid W_{k_1}(\eta), W_{k_2}(\eta)\right) = 1$, then

$$\lim_{K\to\infty} H\left(W_{k_3}(\eta) \mid (W_{k_1}(\eta), W_{k_2}(\eta)) = (0,0)\right) = 1, \quad (103)$$
$$\lim_{K\to\infty} H\left(W_{k_3}(\eta) \mid (W_{k_1}(\eta), W_{k_2}(\eta)) = (0,1)\right) = 1, \quad (104)$$
$$\lim_{K\to\infty} H\left(W_{k_3}(\eta) \mid (W_{k_1}(\eta), W_{k_2}(\eta)) = (1,1)\right) = 1, \quad (105)$$
$$\lim_{K\to\infty} H\left(W_{k_3}(\eta) \mid (W_{k_1}(\eta), W_{k_2}(\eta)) = (1,0)\right) = 1. \quad (106)$$

Consider (103), it is the sector of the $\mathcal{U}$ circle labeled A. Within this sector $W_{k_3}(\eta)$ must be uniform, i.e., half of A must be in S_{k_3} and half of A must be outside S_{k_3}. Similarly, conditions (104), (105) and (106) imply that half of B, C, D must be in S_{k_3} and half of B, C, D must be outside S_{k_3}. But S_{k_3} is a contiguous semicircle, a continuous semi-circle cannot overlap with half of each of A, B, C, D. Therefore we have a contradiction. The contradiction means that either the asymptotic capacity of OR search with special patterns is not $1 - 1/N$ or Theorem 1 is not tight for this OR private search.

V. Concluding Remarks

We introduced the private search problem, which requires PIR with dependent messages (DPIR). We derived a general converse bound for DPIR, studied its asymptotic behavior, and combined it with a known general achievability result in order to characterize the asymptotic capacity of various forms of private search, which include exact search, OR search, AND search, NOT search and sequence search. We also showed through an example that even asymptotic capacity characterizations for private search are difficult for additionally constrained message structures.

We note that the sufficient condition in Theorem 3 is applicable to T-colluding servers [17] or (N, T)-MDS coded servers $(T < N)$ [18], i.e., for DPIR with T-colluding servers or (N, T)-MDS coded servers, if there exists an increasing sequence k_i that satisfies (15), then the asymptotic capacity is $1 - \frac{T}{N}$. The converse proof is similar to the proof of Theorem 3. Following [17], [18], we can obtain a download lower bound similar to Equation (34),

$$D \geq \left(1 + \frac{T}{N} + \frac{T^2}{N^2} + \cdots + \frac{T^{m-1}}{N^{m-1}}\right) LH(w) - \frac{TI(W_{k_2}; W_{k_1})}{N} - \cdots - \frac{T^{m-1} I\left(W_{k_m}; W_{k_{[1:m-1]}}\right)}{N^{m-1}}. \tag{107}$$

With the method in Section IV-B, it is easily proven that $\lim_{\mu\to\infty} \sum_{l=1}^{m-1} \frac{T^l I\left(W_{k_{l+1}}; W_{k_{[1:l]}}\right)}{LH(w)N^l} = 0$. In terms of the achievable scheme, one can use the scheme of [35] and set the parameters $K_c = 1, X = 0, T = T, B = 0, U = 0$ for T-colluding servers, and set parameters $K_c = T, X = 0$, $T = 1, B = 0, U = 0$ for (N, T)-MDS coded servers.

One future direction is the capacity of private search over restricted search patterns discussed in section IV-D. Another future direction is the capacity in non-asymptotic regime. In contrast to the outer bound matching the achieving rate in the asymptotic regime, there is a gap between the outer bound and the achieving rate in the non-asymptotic regime. Take exact search with $P = 1$ as an example, when there are only $K = 2$ messages, the result is trivial because W_1 is a function of W_2. So there is no privacy for $K = 2$. Consider $K = 3$ and $N = 2$, suppose the desired message is W_1, an achievable scheme is shown in Table II.

TABLE II
Achievable Scheme

Server 1	Server 2
a_1, b_1	a_2, b_2
$a_3 + b_2$	$a_5 + b_1$
$a_4 + c_2$	$a_6 + c_2$
$b_4 + c_3$	$b_6 + c_5$
$a_7 + b_6 + c_5$	$a_8 + b_4 + c_3$

The "+" in the scheme means XOR operation. a_i notates $W_1(i)$, which is the i-th symbol of the first message. Similarly, b_i, c_i notate $W_2(i)$ and $W_3(i)$. Based on the problem setting, dependency only exists among a_i, b_i, c_i. The correctness and privacy of this scheme are inherited from the achievable scheme of PIR [16] and the dependence.

In this scheme, on one hand due to the dependency among a_i, b_i, c_i, we achieve the rate $R = 0.6617$. On the other hand, according to our outer bound (12), $C_{PS}(3, 1, 1, 2) \leq 0.7337$. There is a gap between the achievable rate and the outer bound. Bound (12) is an outer bound for general DPIR problems, and Proposition 1 and this example show that the outer bound may not be tight for private search. To close the gap, one might need to improve the converse bound in the future. Note that for $N = 2, K = 3$, the PIR capacity is $4/7 < 0.6617$. It shows that in the non-asymptotic regime, dependency among messages can increase the capacity, which is different from that in the asymptotic case.

Appendix

A. Proof of (45)

Let $f(p) = H_2(p) = -p\log_2(p) - (1-p)\log_2(1-p)$. When $K \to \infty$, $\frac{1}{K-t} \to 0$ and $\frac{1}{K} \to 0$. Since $\lim_{K\to\infty} f\left(\frac{1}{K-t}\right) = 0$, $\lim_{K\to\infty} f\left(\frac{1}{K}\right) = 0$ and both of them are differentiable, L'Hôpital's rule is applicable. Consider the derivative of $f(p)$,

$$\frac{\mathrm{d}f}{\mathrm{d}p} = -\log_2 p - \frac{p}{p\ln 2} + \log_2(1-p) + \frac{1-p}{(1-p)\ln 2} \tag{108}$$

$$= \log_2(1-p) - \log_2 p = \log_2 \frac{1-p}{p}. \tag{109}$$

Let $p = \frac{1}{K-t}$ and $q = \frac{1}{K}$. According to L'Hôpital's rule,

$$\lim_{K\to\infty} \frac{f\left(\frac{1}{K-t}\right)}{f\left(\frac{1}{K}\right)} = \lim_{K\to\infty} \frac{\frac{\mathrm{d}f}{\mathrm{d}p}\frac{\mathrm{d}p}{\mathrm{d}K}}{\frac{\mathrm{d}f}{\mathrm{d}q}\frac{\mathrm{d}q}{\mathrm{d}K}} \tag{110}$$

$$= \lim_{K\to\infty} \frac{\frac{-1}{(K-t)^2}\log_2\frac{1-p}{p}}{\frac{-1}{K^2}\log_2\frac{1-q}{q}} \tag{111}$$

$$= \lim_{K\to\infty} \frac{\frac{-1}{(K-t)^2}\log_2(K-t-1)}{\frac{-1}{K^2}\log_2(K-1)}. \tag{112}$$

Since $\lim_{K\to\infty} \frac{\frac{-1}{(K-t)^2}}{\frac{-1}{K^2}} = 1$ and $\lim_{K\to\infty} \frac{\log_2(K-t-1)}{\log_2(K-1)} = 1$, we obtain

$$\lim_{K\to\infty} \frac{f\left(\frac{1}{K-t}\right)}{f\left(\frac{1}{K}\right)} = 1. \tag{113}$$

B. Proof of (55)

Let $f(p) = H_2(p) = -p\log_2(p) - (1-p)\log_2(1-p)$. When $K \to \infty$, $\frac{(K-t-1)^P}{(K-t)^P} \to 1$ and $\frac{(K-1)^P}{K^P} \to 1$. Since $\lim_{K\to\infty} f\left(\frac{(K-t-1)^P}{(K-t)^P}\right) = 0$, $\lim_{K\to\infty} f\left(\frac{(K-1)^P}{K^P}\right) = 0$ and both of them are differentiable, L′Hôpital′s rule is applicable. Let $p = \frac{(K-t-1)^P}{(K-t)^P}$ and $q = \frac{(K-1)^P}{K^P}$. According to L′Hôpital′s rule,

$$\lim_{K\to\infty} \frac{f\left(\frac{(K-t-1)^P}{(K-t)^P}\right)}{f\left(\frac{(K-1)^P}{K^P}\right)} = \lim_{K\to\infty} \frac{\frac{\mathrm{d}f}{\mathrm{d}p}\frac{\mathrm{d}p}{\mathrm{d}K}}{\frac{\mathrm{d}f}{\mathrm{d}q}\frac{\mathrm{d}q}{\mathrm{d}K}} \tag{114}$$

$$= \lim_{K\to\infty} \frac{\frac{P(K-t-1)^{P-1}(K-t)^{P-1}}{(K-t)^{2P}} \log_2 \frac{1-p}{p}}{\frac{P(K-1)^{P-1}K^{P-1}}{K^{2P}} \log_2 \frac{1-q}{q}} \tag{115}$$

$$= \lim_{K\to\infty} \frac{\frac{P(K-t-1)^{P-1}(K-t)^{P-1}}{(K-t)^{2P}} \log_2 \left(\frac{(K-t)^P}{(K-t-1)^P} - 1\right)}{\frac{P(K-1)^{P-1}K^{P-1}}{K^{2P}} \log_2 \left(\frac{K^P}{(K-1)^P} - 1\right)}. \tag{116}$$

Since

$$\lim_{K\to\infty} \frac{\frac{P(K-t-1)^{P-1}(K-t)^{P-1}}{(K-t)^{2P}}}{\frac{P(K-1)^{P-1}K^{P-1}}{K^{2P}}} = 1 \tag{117}$$

and

$$\lim_{K\to\infty} \frac{\log_2 \left(\frac{(K-t)^P}{(K-t-1)^P} - 1\right)}{\log_2 \left(\frac{K^P}{(K-1)^P} - 1\right)} \tag{118}$$

$$\overset{\text{L′Hôpital′s rule}}{=} \lim_{K\to\infty} \frac{\frac{1}{\left(\frac{(K-t)^P}{(K-t-1)^P}-1\right)\ln 2} \cdot \frac{-P(K-t)^{P-1}}{(K-t-1)^{P+1}}}{\frac{1}{\left(\frac{K^P}{(K-1)^P}-1\right)\ln 2} \cdot \frac{-PK^{P-1}}{(K-1)^{P+1}}} = 1, \tag{119}$$

we obtain

$$\lim_{K\to\infty} \frac{f\left(\frac{(K-t-1)^P}{(K-t)^P}\right)}{f\left(\frac{(K-1)^P}{K^P}\right)} = 1. \tag{120}$$

REFERENCES

[1] Z. Chen, Z. Wang, and S. Jafar, "The asymptotic capacity of private search," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 2122–2126.

[2] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in *Advances in Cryptology—EUROCRYPT*. Berlin, Germany: Springer, 2004, pp. 506–522.

[3] E.-J. Goh, "Secure indexes," *IACR Cryptol. ePrint Arch.*, vol. 2003, p. 216, 2003. [Online]. Available: https://eprint.iacr.org/2003/216.pdf

[4] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in *Proc. IEEE INFOCOM*, Mar. 2010, pp. 1–5.

[5] M. Kuzu, M. S. Islam, and M. Kantarcioglu, "Efficient similarity search over encrypted data," in *Proc. IEEE 28th Int. Conf. Data Eng.*, Apr. 2012, pp. 1156–1167.

[6] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in *Proc. Appl. Cryptogr. Netw. Secur.*, 2004, pp. 31–45.

[7] J. Katz, A. Sahai, and B. Waters, "Secure conjunctive keyword search over encrypted data," in *Advances in Cryptology*. 2008, pp. 146–162.

[8] C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 8, pp. 1467–1479, Aug. 2012.

[9] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private information retrieval," *J. ACM*, vol. 45, no. 6, pp. 965–981, 1998.

[10] B. Chor, N. Gilboa, and M. Naor, "Private information retrieval by keywords," Dept. Comput. Sci., Technion, Haifa, Israel, Tech. Rep. TR-CS0917, 1997.

[11] R. Ostrovsky and W. E. Skeith, III, "Private searching on streaming data," in *Advances in Cryptology*. Berlin, Germany: Springer, 2005, pp. 223–240.

[12] M. Finiasz and K. Ramchandran, "Private stream search at the same communication cost as a regular search: Role of LDPC codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 2556–2560.

[13] J. Bethencourt, D. Song, and B. Waters, "New constructions and practical applications for private stream searching," in *Proc. IEEE Symp. Secur. Privacy (S&P)*, May 2006, p. 139.

[14] G. Fanti, "Private media search on public databases," M.S. thesis, Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, Berkeley, CA, USA, 2012.

[15] G. Fanti, "Privacy-preserving messaging and search: A collaborative approach," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, Berkeley, CA, USA, 2015.

[16] H. Sun and S. A. Jafar, "The capacity of private information retrieval," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 4075–4088.

[17] H. Sun and S. A. Jafar, "The capacity of robust private information retrieval with colluding databases," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2361–2370, Apr. 2018.

[18] K. Banawan and S. Ulukus, "The capacity of private information retrieval from coded databases," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1945–1956, Mar. 2018.

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

[20] M. Abdul-Wahid, F. Almoualem, D. Kumar, and R. Tandon, "Private information retrieval from storage constrained databases—Coded caching meets PIR," 2017, *arXiv:1711.05244*. [Online]. Available: http://arxiv.org/abs/1711.05244

[21] H. Sun and S. A. Jafar, "The capacity of symmetric private information retrieval," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 322–329, Jan. 2019.

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

[23] R. Tandon, "The capacity of cache aided private information retrieval," in *Proc. 55th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2017, pp. 1078–1082.

[24] Y.-P. Wei, K. Banawan, and S. Ulukus, "Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 3215–3232, May 2019.

[25] S. Kadhe, B. Garcia, A. Heidarzadeh, S. El Rouayheb, and A. Sprintson, "Private information retrieval with side information," 2017, *arXiv:1709.00112*. [Online]. Available: http://arxiv.org/abs/1709.00112

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

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

[28] H. Sun and S. A. Jafar, "The capacity of private computation," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3880–3897, Jun. 2019.

[29] M. Mirmohseni and M. A. Maddah-Ali, "Private function retrieval," in *Proc. Iran Workshop Commun. Inf. Theory (IWCIT)*, Apr. 2018, pp. 1–6.

[30] S. A. Obead and J. Kliewer, "Achievable rate of private function retrieval from MDS coded databases," 2018, *arXiv:1802.08223*. [Online]. Available: http://arxiv.org/abs/1802.08223

[31] S. A. Obead, H.-Y. Lin, E. Rosnes, and J. Kliewer, "Capacity of private linear computation for coded databases," 2018, *arXiv:1810.04230*. [Online]. Available: http://arxiv.org/abs/1810.04230

[32] N. Raviv and D. A. Karpuk, "Private polynomial computation from Lagrange encoding," 2018, *arXiv:1812.04142*. [Online]. Available: http://arxiv.org/abs/1812.04142

[33] M. H. Mousavi, M. A. Maddah-Ali, and M. Mirmohseni, "Private inner product retrieval for distributed machine learning," 2019, *arXiv:1902.06319*. [Online]. Available: http://arxiv.org/abs/1902.06319

[34] N. B. Shah, K. V. Rashmi, and K. Ramchandran, "One extra bit of download ensures perfectly private information retrieval," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 856–860.

[35] Z. Jia and S. A. Jafar, "X-secure T-private information retrieval from MDS coded storage with Byzantine and unresponsive servers," 2019, *arXiv:1908.10854*. [Online]. Available: http://arxiv.org/abs/1908.10854

Zhen Chen (Student Member, IEEE) received the B.S. and M.S. degrees in electronic information engineering from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of California Irvine, Irvine, CA, USA. His research interests include coded distributed computation, information theory, and its applications to security and privacy.

Zhiying Wang (Member, IEEE) received the B.Sc. degree in information electronics and engineering from Tsinghua University in 2007, and the M.Sc. and Ph.D. degrees in electrical engineering from the California Institute of Technology in 2009 and 2013, respectively. She was a Post-Doctoral Fellow with the Department of Electrical Engineering, Stanford University. She is currently an Assistant Professor with the Center for Pervasive Communications and Computing, University of California Irvine, Irvine, CA, USA. Her research focuses on information theory, coding theory, with an emphasis on coding for storage devices and systems. Dr. Wang received the NSF Center for Science of Information (CSoI) Postdoctoral Research Fellow in 2013. She was a recipient of IEEE Communication Society Data Storage Best Paper Award in 2013.

Syed Ali Jafar (Fellow, IEEE) received the B.Tech. degree from IIT Delhi, India, in 1997, the M.S. degree from Caltech, Pasadena, CA, USA, in 1999, and the Ph.D. degree from Stanford University, Stanford, CA, USA, in 2003, all in electrical engineering.

His industry experience includes positions at Lucent Bell Labs and Qualcomm. He is currently a Professor with the Department of Electrical Engineering and Computer Science, University of California Irvine, Irvine, CA, USA. His research interests include multiuser information theory, wireless communications, and network coding.

Dr. Jafar was a recipient of the New York Academy of Sciences Blavatnik National Laureate in physical sciences and engineering, the NSF CAREER Award, the ONR Young Investigator Award, the UCI Academic Senate Distinguished Mid-Career Faculty Award for Research, the School of Engineering Mid-Career Excellence in Research Award, and the School of Engineering Maseeh Outstanding Research Award. His coauthored articles have received the IEEE Information Theory Society Paper Award, IEEE Communication Society and Information Theory Society Joint Paper Award, IEEE Communications Society Best Tutorial Paper Award, IEEE Communications Society Heinrich Hertz Award, IEEE Signal Processing Society Young Author Best Paper Award, IEEE Information Theory Society Jack Wolf ISIT Best Student Paper Award, and three IEEE GLOBECOM Best Paper Awards. He received the UC Irvine EECS Professor of the Year award six times from the Engineering Students Council in 2006, 2009, 2011, 2012, 2014, and 2017, the School of Engineering Teaching Excellence Award in 2012, and a Senior Career Innovation in Teaching Award in 2018. He was a University of Canterbury Erskine Fellow in 2010 and an IEEE Communications Society Distinguished Lecturer from 2013 to 2014. Since 2019, he has been an IEEE Information Theory Society Distinguished Lecturer. He was recognized as a Thomson Reuters/Clarivate Analytics Highly Cited Researcher and included by Sciencewatch among the World's Most Influential Scientific Minds in 2014, 2015, 2016, 2017, and 2018. He served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS from 2004 to 2009, the IEEE COMMUNICATIONS LETTERS from 2008 to 2009, and the IEEE TRANSACTIONS ON INFORMATION THEORY from 2009 to 2012.