

Intermittent Private Information Retrieval With Application to Location Privacy

Fangwei Ye[✉], *Member, IEEE*, and Salim El Rouayheb, *Senior Member, IEEE*

Abstract—We study the problem of intermittent private information retrieval with multiple servers, in which a user consecutively requests one of K messages from N replicated databases such that part of requests need to be protected while others do not need privacy. Motivated by the location privacy application, the correlation between requests is modeled by a Markov chain. We propose an intermittent private information retrieval scheme that concatenates an obfuscation scheme and a private information retrieval scheme for the time period when privacy is not needed, to prevent leakage incurred by the correlation over time. In the end, we illustrate how the proposed scheme for the problem of intermittent private information retrieval with Markov structure correlation can be applied to design a location privacy protection mechanism in the location privacy problem.

Index Terms—Information-theoretic privacy, private information retrieval, location privacy.

I. INTRODUCTION

PRVACY-PRESERVING mechanism [1]–[3] has been intensively studied because of the upsurge in privacy concerns. An emerging application is location privacy, since location-based service becomes an integral part of daily life. Location privacy has attracted significant attention recently [4]–[15], and particularly from an information-theoretic perspective [11], [12]. However, it has not been fully addressed that how the correlation between locations corrodes privacy, especially when a user may not need privacy all the time due to the overheads incurred by the privacy-preserving mechanism. To capture the impact of correlation in location privacy systematically, we formulate a theoretical problem in the framework of information-theoretic private information retrieval (PIR) [16], [17]. In particular, we are interested in the Markov structure correlation, as a commonly adopted mobility model that models the correlation between locations is the Markov model [7]–[12].

Private information retrieval (PIR) [16], [17] has attracted significant attention recently due to its key role in understanding privacy in downloading scenarios. The PIR capacity, that is, the utility metric to measure download cost from

databases, was characterized by Sun and Jafar [17], in which the canonical setting is that a user is interested in retrieving one of the K messages from N replicated database while hiding the identity of the desired message. Many variants of the ordinary PIR problem have been studied in [18]–[32].

The new variation to be studied in this paper, namely intermittent private information retrieval, is motivated by the fact that privacy usually comes at a cost so a user may not need privacy all the time. Privacy-preserving mechanisms typically incur higher overheads in terms of computation, memory, and delay, etc. These incurred burdens may motivate the user to choose whether he/she needs privacy or not at certain times. For example, people may switch between normal and incognito modes in browsers depending on network connection and sensitivity of contents, etc.

Under the intermittent PIR setting, when a user needs privacy, he/she has to use a PIR scheme. The question is what should be done when the user does not need privacy. One natural answer is a straightforward scheme, i.e., a scheme without any concern of privacy, which suffers from the fact that the user's behavior is usually correlated over time and hence a careless downloading at the current time will leak information about the request at the time instance that needs privacy. Another natural answer is a PIR scheme, which surely preserves privacy due to the one-shot nature of the PIR scheme [17]. However, this conservative strategy generally sacrifices the efficiency, i.e., increasing the download cost, since it over-protects a request that does not need privacy.

In this paper, we study the problem of intermittent private information retrieval with Markov structure correlation, in which a user consecutively requests one of K time-varying messages from N replicated databases at each time, such that part of requests (at some time periods) need privacy while others do not need privacy. The requests over time are correlated and we model the correlation by a first-order Markov chain as said.

We propose a solution that can be considered as a concatenation of an obfuscation scheme and a PIR scheme. In particular, the scheme can be viewed as a PIR scheme over a randomized subset of messages, where the subset is optimized according to the given Markov structure correlation between requests. Also, we bound the download cost of the concatenation scheme. The obfuscation scheme that optimizes the randomly chosen subsets first appeared as a primitive component in the ON-OFF privacy problem [33], [34] proposed by the authors, where the ON-OFF privacy problem can be regarded as an intermittent PIR problem with a *single server* in the language of this paper. Therefore, the proposed scheme

Manuscript received May 31, 2021; revised November 1, 2021; accepted December 21, 2021. Date of publication January 12, 2022; date of current version February 17, 2022. The work of Salim El Rouayheb was supported in part by NSF Grant CCF 1817635. (Corresponding author: Fangwei Ye.)

Fangwei Ye was with the Department of Electrical and Computer Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ 08854 USA. He is now with the Broad Institute of MIT and Harvard, Cambridge, MA 02142 USA (e-mail: fye@broadinstitute.org).

Salim El Rouayheb is with the Department of Electrical and Computer Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ 08854 USA (e-mail: salim.elrouayheb@rutgers.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JSAC.2022.3142301>.

Digital Object Identifier 10.1109/JSAC.2022.3142301

0733-8716 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See <https://www.ieee.org/publications/rights/index.html> for more information.

in this paper can be considered as extending the obfuscation scheme for the intermittent PIR with a single server therein to the setting of the intermittent PIR with *multiple servers*.

To echo the location privacy motivation at the beginning, we will illustrate how the proposed scheme for the problem of intermittent PIR with Markov structure correlation can be applied to design a location privacy protection mechanism in the location privacy problem at the end.

Organization: The rest of the paper is organized as follows. In Section II, we formulate the problem of intermittent private information retrieval with Markov structure correlation. The canonical case of two requests is discussed in Section III, and the general case of a Markov chain is discussed in Section IV. In Section V, we show how to apply the intermittent PIR scheme to the location privacy application. We conclude the paper in Section VI.

Notation: Throughout this paper, the probability distribution for a random variable X that takes values in an alphabet \mathcal{X} is denoted by $\{p_X(x) : x \in \mathcal{X}\}$ with $p_X(x) = \mathbb{P}\{X = x\}$. When there is no ambiguity, $p_X(x)$ will be abbreviated as $p(x)$.

II. PROBLEM FORMULATION

We follow the terminology in [17] to introduce the setting of intermittent private information retrieval with multiple servers accompanying with the correlation between requests over time.

We assume that there are N servers and K *time-varying* messages in the system. At each discrete time t , the messages $W_{1,t}, \dots, W_{K,t}$ are generated independently by K information sources. At time t , each of the servers stores a replica of all K updated messages $W_{1,t}, \dots, W_{K,t}$. We slightly abuse the notations by dropping the time index t from $W_{i,t}$ for notational simplicity, and the underlying time index t will be clear in the context. Assume that K messages (at each time) are mutually independent and each of the messages consists of L independent bits that uniformly take values in the binary alphabet $\{0, 1\}$.

At each time t , the user is interested in retrieving a message from $\{W_1, \dots, W_K\}$. Let $\{X^{(t)} : t = 0, 1, 2, \dots\}$ denote the requests, where each $X^{(t)}$ takes values in $[K] := \{1, \dots, K\}$. As mentioned in the introduction, the correlation model of requests is an essential attribute in the problem, and we are particularly interested in the case where the requests $X^{(t)}$ for $t = 0, 1, \dots$ form a Markov chain.

The intermittence introduced in this paper is described as follows. The user may or may not wish to hide the identity of the message of interest at time t . Specifically, let $S^{(t)}$ denote the privacy status at time t , where $S^{(t)} = 1$ means that the user wishes to keep $X^{(t)}$ private and $S^{(t)} = 0$ means that the user is not concerned with privacy. We assume that the privacy status $S^{(t)}$ is completely chosen by the user, i.e., $S^{(t)}$ is viewed as a given parameter that is independent of the user's request $X^{(t)}$. We also assume that the privacy status $S^{(t)}$ is shared by both the servers and the user. In other words, we are not interested in hiding the privacy status in our formulation. Without loss of generality, we assume that $S^{(0)} = 1$, i.e., $X^{(0)}$ needs privacy.

The same as the classical PIR setting, suppose that a user wants to retrieve a message $W_{X^{(t)}}$ at time t . To retrieve the

message, the user generates N queries $Q_1^{(t)}, \dots, Q_N^{(t)}$ and the query $Q_i^{(t)}$ will be sent to the i -th server. To clarify, the user may generate the query for the request $X^{(t)}$ by utilizing all the *causal* information, i.e., all the previous and the current requests $X^{(j)}$ for $j \leq t$, all the previous and the current privacy status $S^{(j)}$ for $j \leq t$, and all the previous queries $Q_i^{(j)}$ for $j < t$. More rigorously, the query at time t is supposed to be generated by the query function that maps $\{X^{(j)}, S^{(j)} : j \leq t\}$, $Q_i^{(j)}$ for $j < t$ and some random key $F^{(t)}$, to the query $Q_i^{(t)}$, i.e.,

$$\Phi_i : \{1, \dots, K\}^{t+1} \times \{0, 1\}^{t+1} \times \mathcal{Q}^t \times \mathcal{F} \rightarrow \mathcal{Q}, \quad (1)$$

where \mathcal{Q} is supposed to be a common alphabet of queries for conciseness, and $F^{(t)}$ denotes the random key¹ on the alphabet \mathcal{F} .

Upon receiving the a query $Q_i^{(t)}$, the i -th server generates an answer $A_i^{(t)}$ to response to the query. We require that the answer $A_i^{(t)}$ is a deterministic function of the query $Q_i^{(t)}$ provided the stored messages. After receiving answers $A_1^{(t)}, \dots, A_N^{(t)}$, the user should be able to decode the desired message $W_{X^{(t)}}$ with zero error probability.

We would like to clarify two points about the setting. First, for any given privacy status, the query $Q_i^{(t)}$ may be viewed as a stochastic function of all the causal requests. Second, the messages are assumed to be time-varying, more precisely independent over time, so the answer $A_i^{(t)}$ only depends on the current messages and the query $Q_i^{(t)}$.

As said, the user should be able to decode the message of interest, which is referred to as *correctness* requirement [17]. The correctness requirement is defined in the same way in this paper, i.e.,

$$H(W_{X^{(t)}} | X^{(t)}, F^{(t)}, Q_{1:N}^{(t)}, A_{1:N}^{(t)}) = 0, \quad (2)$$

where $Q_{1:N}^{(t)} := \{Q_i^{(t)} : i = 1, \dots, N\}$ and $A_{1:N}^{(t)} := \{A_i^{(t)} : i = 1, \dots, N\}$.

The other requirement of the system is the *privacy* requirement. For our intermittent PIR setting, we require that for any time t , given all previous queries received by the i -th server, the query $Q_i^{(t)}$ should not reveal any information about the causal requests that need privacy, i.e.,

$$[\text{Privacy}] \quad I(X^{(\mathcal{P}_t)}; Q_i^{(t)} | Q_i^{(1)}, \dots, Q_i^{(t-1)}) = 0, \quad (3)$$

for $i \in [N]$, where $X^{(\mathcal{P}_t)} := \{X^{(j)} : j \in \mathcal{P}_t\}$ and

$$\mathcal{P}_t := \{j : S^{(j)} = 1, j \leq t\}. \quad (4)$$

Note that $0 \in \mathcal{P}_t$ for any t from the assumption $S^{(0)} = 1$.

The conditioning in (3) serves to ensure causality by design. Barring this conditioning, privacy could be alternatively defined by

$$I(X^{(\mathcal{P}_t)}; Q_i^{(1)}, \dots, Q_i^{(t)}) = 0. \quad (5)$$

However, this alternative definition implies that at some point $j < t$, the query $Q_i^{(j)}$ may be required to protect some *future*

¹The key in this paper may be context-dependent, i.e., generated dependent of the input.

request $X^{(j')}$ such that $j' > j$ and $j' \in \mathcal{P}_t$, i.e.,

$$I(X^{(j')}; Q_i^{(j)}) = 0,$$

induced by (5). This generally means that the adversary may attempt to infer future requests. Given the correlation between the requests and the assumption that the user does not know (or infer) the future requests in advance (adhere to causality in our formulation), the alternative privacy definition enforces a less interesting solution that the query $Q_i^{(j)}$ must be independent of $X^{(j)}$, i.e., a standard PIR scheme all the time. For this reason, we adopt (3) as the privacy requirement, which leads to a theoretically interesting and meaningful problem.

Conventionally, the utility metric is defined by the normalized download cost. Let $\ell_i^{(t)}$ denote the length of the answer $A_i^{(t)}$, and the normalized download cost of the i -th server is given by

$$\alpha_i^{(t)} := \frac{\mathbb{E}[\ell_i^{(t)}]}{L},$$

that is the expected amount of downloaded data per bit desired message from the i -th server. Correspondingly, the total download cost is

$$\alpha^{(t)} = \sum_{i=1}^N \alpha_i^{(t)}.$$

Clearly, we are aimed to minimize the total download cost.

III. CANONICAL CASE: TWO REQUESTS

In this section, we start from the canonical case of two requests, which is the first step to understand the impact of the correlation in the intermittent private information retrieval problem. Also, we will see later it indeed serves as the key component to solve the general problem where the requests are modeled by a Markov chain.

Let $X^{(0)}$ and $X^{(1)}$ be two random variable taking values in $[K]$, representing two requests at time $t = 0$ and $t = 1$ respectively. Suppose that $S^{(0)} = 1$ and $S^{(1)} = 0$, i.e., $X^{(0)}$ at time $t = 0$ is a request that needs privacy while $X^{(1)}$ at time $t = 1$ is a request that does not need privacy. The initial probability distribution of $X^{(0)}$ is denoted by π_0 , and the transition probabilities $p(x^{(1)}|x^{(0)})$ for $x^{(0)}, x^{(1)} \in [K]$ are known.

By invoking the privacy requirement in (3), the designed queries $Q_i^{(0)}$ and $Q_i^{(1)}$ for $i \in [N]$ at two time periods $t = 0$ and $t = 1$ should satisfy

$$I(X^{(0)}; Q_i^{(0)}) = 0, \quad (6)$$

and

$$I(X^{(0)}; Q_i^{(1)}|Q_i^{(0)}) = 0. \quad (7)$$

Note that $\mathcal{P}_0 = \mathcal{P}_1 = \{0\}$ provided the privacy status $S^{(0)} = 1$ and $S^{(1)} = 0$.

First, we notice that the sub-problem of minimizing the download cost $\alpha^{(0)}$ at $t = 0$, satisfying (6), is exactly a original PIR problem, and the minimum download cost is known in [17], i.e., $\min \alpha^{(0)} = C(N, K)$, where

$$C(N, K) := 1 + N^{-1} + N^{-2} + \dots + N^{-K+1}, \quad (8)$$

which can be achieved by the PIR-capacity achieving scheme therein.

Therefore, the interesting part is to ask if there exists a better retrieval mechanism at $t = 1$ when the privacy is not needed, while preserving the privacy of $X^{(0)}$.

Provided that $Q_i^{(0)}$ is the query of the PIR capacity-achieving scheme, i.e.,

$$I(X^{(0)}; Q_i^{(0)}) = 0 \quad \text{and} \quad I(Q_i^{(1)}; Q_i^{(0)}|X^{(0)}) = 0,$$

where the latter one follows because the query $Q_i^{(0)}$ of a PIR scheme only depends on $X^{(0)}$ and some random key, the privacy requirement (7) can then be written by

$$I(X^{(0)}; Q_i^{(1)}) = 0, \quad \forall i \in [N], \quad (9)$$

which is formally stated in the following proposition and the justification is deferred to the appendix.

Proposition 1: For any $i \in [N]$, given $I(X^{(0)}; Q_i^{(0)}) = 0$ and $I(Q_i^{(1)}; Q_i^{(0)}|X^{(0)}) = 0$, we know that

$$I(X^{(0)}; Q_i^{(1)}|Q_i^{(0)}) = 0$$

if and only if

$$I(X^{(0)}; Q_i^{(1)}) = 0.$$

Therefore, we will focus on designing queries $Q_i^{(1)}$, $i \in [N]$ satisfying the requirement (9) in the sequel. We start by introducing some necessary notations and stating the result.

Suppose the transition probabilities $p(x^{(1)}|x^{(0)})$ for $x^{(0)}, x^{(1)} \in [K]$ are given. For any $i \in [K]$, suppose that

$$\begin{aligned} \mathbb{P}\{X^{(1)} = i|X^{(0)} = v_{i,1}\} \\ \leq \dots \leq \mathbb{P}\{X^{(1)} = i|X^{(0)} = v_{i,K}\}, \end{aligned} \quad (10)$$

i.e., ordering the likelihood probabilities of $p(x^{(1)}|x^{(0)})$, where $v_{i,1}, \dots, v_{i,K}$ are K distinct elements in $[K]$. Let λ_j be the summation of the j -th minimal likelihood probabilities for each possible value of $X^{(1)}$, i.e.,

$$\lambda_j := \sum_{i \in [K]} \mathbb{P}\{X^{(1)} = i|X^{(0)} = v_{i,j}\}, \quad j = 1, \dots, K. \quad (11)$$

Also, let

$$\sigma := \max\{j : \lambda_j \leq 1\}, \quad (12)$$

and

$$\theta_j := \min\{1, \lambda_j\} - \min\{1, \lambda_{j-1}\}, \quad (13)$$

i.e., $\lambda_j - \lambda_{j-1}$ for $j \leq \sigma$, $1 - \lambda_\sigma$ for $j = \sigma + 1$, and 0 for $j > \sigma + 1$. All these parameters can be obtained from the given transition probabilities $p(x^{(1)}|x^{(0)})$ for $x^{(0)}, x^{(1)} \in [K]$.

Theorem 1: For any given transition probabilities $p(x^{(1)}|x^{(0)})$ for $x^{(0)}, x^{(1)} \in [K]$, there exists an intermittent private information retrieval scheme with download cost

$$\alpha^{(1)} = \mathbb{E} \left[\left(1 - \frac{1}{N}\right)^{-1} \left(1 - \frac{1}{N^{|\mathcal{U}|}}\right) \right], \quad (14)$$

for some random variable U that takes value in the power set of $[K]$ such that

$$\mathbb{P}\{|U| \leq i\} \geq \sum_{j=1}^i \theta_j, \quad \forall i = 1, \dots, K. \quad (15)$$

To clarify, Theorem 1 states that there exists some U satisfying (15), such that the download cost of the scheme at time $t = 1$ is $\mathbb{E}[C(N, |U|)]$, as shown in (14). In fact, the auxiliary random variable U represents an obfuscation scheme in our design. Therefore, (14) implies that the download cost of our intermittent private information retrieval scheme depends on the design of the obfuscation scheme, and (15) guarantees that there exists an obfuscation scheme satisfying (15).

As $C(N, |U|)$, i.e., the expression with the expectation in (14), is increasing with $|U|$ for a given N , it suggests that if the probability of U of small size is larger, then the download cost is generally smaller. However, due to the privacy requirement, it may not be possible to make $|U|$ too small, e.g., the extreme case is that $|U| = 1$ with probability 1. Nevertheless, (15) guarantees the existence of a random variable U such that the distribution of U satisfies (15), where the worst case is that

$$\mathbb{P}\{|U| = i\} = \theta_i, \quad i = 1, \dots, K. \quad (16)$$

As such, we have the following corollary immediately from the theorem.

Corollary 1: For any given transition probabilities $p(x^{(1)}|x^{(0)})$ for $x^{(0)}, x^{(1)} \in [K]$, there exists an intermittent private information retrieval scheme with download cost

$$\alpha^{(1)} \leq \sum_{i=1}^K \theta_i \left(1 - \frac{1}{N}\right)^{-1} \left(1 - \frac{1}{N^i}\right). \quad (17)$$

It is clear that the corollary can be established by showing that (16) is indeed the worst case of (15), in terms of the corresponding download cost, which is justified as follows:

$$\begin{aligned} \mathbb{E}[C(N, |U|)] &= \sum_{i=1}^K C(N, i) \mathbb{P}\{|U| = i\} \\ &= \sum_{i=1}^K (C(N, i) - C(N, i-1)) \sum_{j=i}^K \mathbb{P}\{|U| = j\} \\ &= \sum_{i=1}^K (C(N, i) - C(N, i-1)) \mathbb{P}\{|U| \geq i\}. \end{aligned}$$

Since (15) implies that

$$\mathbb{P}\{|U| \geq i\} \leq \sum_{j=i}^K \theta_j$$

by the fact that $\sum_{j=1}^K \theta_j = 1$ from the definition of θ_j (c.f.(13)), and $C(N, i)$ is increasing with i ,

we immediately obtain that

$$\begin{aligned} \mathbb{E}[C(N, |U|)] &= \sum_{i=1}^K (C(N, i) - C(N, i-1)) \mathbb{P}\{|U| \geq i\} \\ &\leq \sum_{i=1}^K (C(N, i) - C(N, i-1)) \sum_{j=i}^K \theta_j \\ &= \sum_{i=1}^K C(N, i) \theta_i, \end{aligned}$$

which completes the justification.

Another immediate observation of the theorem is that the right-hand side of (14) is exactly the same as the inverse of the PIR capacity in [17], when $|U| = K$ certainly, i.e., $U = [K]$ with probability 1. As said, the expression within the expectation is increasing with $|U|$ and a trivial upper bound on $|U|$ is K , which implies that the download cost specified by (14) and (15) is always better than the download cost of a standard PIR scheme in general.

Example: To better illustrate the impact of correlation, we present an example here to show the relation between the download cost $\alpha^{(1)}$ and the given transition probabilities $p(x^{(1)}|x^{(0)})$. We study the simplest case $N = K = 2$, and we write $p(x^{(1)}|x^{(0)})$ explicitly by the probability transition matrix

$$P = \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix},$$

such that $0 \leq \alpha, \beta \leq 1$. $P_{i,j}$ denotes $\mathbb{P}\{X^{(1)} = j | X^{(0)} = i\}$. By inspecting the definition (13), we know that

$$\theta_1 = \min\{\alpha + \beta, 2 - \alpha - \beta\}, \quad \theta_2 = 1 - \theta_1.$$

Without loss of generality, we assume that $\alpha + \beta \leq 1$. By (17), we have

$$\alpha^{(1)} \leq \frac{3}{2} - \frac{1}{2}(\alpha + \beta),$$

where $3/2$ is the download cost of a PIR scheme over $K = 2$ messages for $N = 2$ servers and $\alpha + \beta$ somehow represents the correlation. We can clearly see two extreme cases. If $\alpha + \beta = 1$, i.e., requests are independent, then $\alpha^{(1)} = 1$, which implies that we can retrieve the desired message directly (or view it as a PIR scheme for 1 message and 2 servers). If $\alpha + \beta = 0$, i.e., requests are deterministic by each other, then $\alpha^{(1)} = 3/2$, which corresponds to the download cost of a PIR scheme for $K = N = 2$.

In the next section, we will describe a scheme achieving the download cost shown in Theorem 1. In particular, we will show there exists some U (obfuscation scheme) satisfying (15), such that the average download cost of the resulting intermittent private information retrieval scheme is $\mathbb{E}[C(N, |U|)]$.

A. Concatenation Scheme

We consider a concatenation of an obfuscation scheme and a standard PIR capacity-achieving² scheme [17], to achieve the download cost as shown in Theorem 1.

²Any capacity-achieving PIR scheme works, and we choose the pioneering one [17] for concreteness.

TABLE I
JOINT PROBABILITY DISTRIBUTION $p(x^{(0)}, x^{(1)})$

$X^{(0)} \backslash X^{(1)}$	1	2
1	3/8	1/8
2	1/8	3/8

TABLE II
CONDITIONAL PROBABILITIES $p(u|x^{(0)}, x^{(1)})$

$(X^{(0)}, X^{(1)}) \backslash U$	{1}	{2}	{1, 2}
(1, 1)	1/3	0	2/3
(1, 2)	0	1	0
(2, 1)	1	0	0
(2, 2)	0	1/3	2/3

A helpful observation on PIR capacity [17] is that the capacity is decreasing with the number of messages, so the general idea here is that we randomly choose a subset $U \subset [K]$ of messages, and implement the PIR scheme over the selected subset of messages. Generally speaking, the download cost is smaller when the size of the subset is smaller. However, privacy may not hold when the size of the subset is too small. For example, if $|U| = 1$ certainly, i.e., only downloading the desired message, the privacy may be broken since the server immediately knows which message is being retrieved. On the other hand, if $|U| = K$ certainly, i.e., always using a standard PIR scheme over K messages, the privacy holds but with a high download cost. Therefore, we have to optimize the randomized way of choosing such a subset U to reduce its size while preserving privacy.

More precisely, we first obfuscate the request $X^{(1)}$ to a set $U \subseteq [K]$ that includes $X^{(1)}$, and then retrieve the message $W_{X^{(1)}}$ privately by taking the PIR capacity-achieving scheme over a subset of messages $\{W_i : i \in U\}$. In this way, the PIR scheme preserves the identity of $X^{(1)}$ provided U , i.e., only information about U is leaked, and the obfuscation is designed to guarantee that no information about $X^{(0)}$ can be obtained from U .

Example: Before describing the scheme in details, we study the simplest example $N = K = 2$ to illustrate the idea. The setting of servers is the same as the example in [17], i.e., each server stores a full copy two messages (at time $t = 1$) (a_1, a_2, a_3, a_4) and (b_1, b_2, b_3, b_4) .

Suppose that the joint distribution of $X^{(0)}$ and $X^{(1)}$ is given in Table I. The obfuscation set U can be designed according to the conditional probabilities in Table II.

Assume that $X^{(0)} = X^{(1)} = 1$. With probability $\frac{2}{3}$, the user will request the first message via a standard $N = K = 2$ PIR scheme, e.g., querying for $(a_1, b_1, a_3 + b_2)$ from the first server and $(a_2, b_2, a_4 + b_1)$ from the second server, i.e., totally 6 bits downloaded for a message of 4 bits. With probability $\frac{1}{3}$, the user will directly request the message 1 as desired, e.g., directly querying for (a_1, a_2) from the first server and (a_3, a_4) from the second server. We can check that $\frac{16}{3}$ bits are downloaded on average to retrieve the first message when $X^{(0)} = X^{(1)} = 1$.

Similarly, if $X^{(0)} = 2$ and $X^{(1)} = 1$, the user will directly request the message 1 as desired certainly from Table II, e.g., querying for (a_1, a_2) from the first server and (a_3, a_4) from the second server.

TABLE III
TIME-SHARING OF TWO SCHEMES FOR $X^{(0)} = X^{(1)} = 1$ BASED ON U

Probability	DB1	DB2
$\frac{2}{3}$	a_1, b_1 $a_3 + b_2$	a_2, b_2 $a_4 + b_1$
$\frac{1}{3}$	a_1 a_2	a_3 a_4

When one of the servers, e.g., the first server, receives the queries for (a_1, a_2) , although it immediately knows that the request at this time is $X^{(1)} = 1$, the privacy of $X^{(0)}$ is still preserved, since $X^{(0)} = 1$ and $X^{(0)} = 2$ are equally likely when (a_1, a_2) is retrieved, i.e.,

$$\begin{aligned} \mathbb{P}\{Q_1^{(1)} = (a_1, a_2) | X^{(0)} = 1\} \\ = \mathbb{P}\{Q_1^{(1)} = (a_1, a_2) | X^{(0)} = 2\} = \frac{1}{4}, \end{aligned}$$

due to the design of U for the given correlation between $X^{(0)}$ and $X^{(1)}$.

Now, we describe the concatenation scheme in details as follows.

Obfuscation: Suppose that U is a subset of $[K]$, i.e., U takes values in the power set of $[K]$, denoted by \mathcal{P}_K . Choose U based on $X^{(1)}$ and $X^{(0)}$, more precisely the conditional probability $p(u|x^{(1)}, x^{(0)})$ for any given $p(x^{(1)}, x^{(0)})$, to be a solution to the following “optimization” problem:

$$\begin{aligned} &\text{minimize } \mathbb{E}[C(N, |U|)] \\ &\text{subject to } X^{(1)} \in U, \\ &\quad U \text{ is independent of } X^{(0)}. \end{aligned} \quad (18)$$

Note that U is a random variable and the expectation in the objective function is over U . A more standard formulation of this optimization problem, i.e., describing the decision variables explicitly, is deferred to the end of this section. Here, we keep this neat formulation to illustrate the basic idea of the obfuscation scheme. The constraint $X^{(1)} \in U$ represents that $X^{(1)} \in U$ certainly, or more precisely $p(u, x^{(1)}) = 0$ for $x^{(1)} \notin u$. The two constraints are indeed imposed to closely depict the intuitive idea of the scheme, i.e., obfuscating the request $X^{(1)}$ to a set U that includes $X^{(1)}$ (necessary for the next PIR phase) and preserving the privacy of $X^{(0)}$. The discussion on solving this optimization problem is also deferred to the end of this section, and now let us just assume that the problem is solvable and the solution $p(u|x^{(1)}, x^{(0)})$ can be obtained. After obtaining the solution $p(u|x^{(1)}, x^{(0)})$, sample an obfuscation set u according to $p(u|x^{(1)}, x^{(0)})$ based on the observed requests $x^{(1)}$ and $x^{(0)}$.

Retrieval: Given the request $X^{(1)}$ and the obfuscation set U , retrieve the message $W_{X^{(1)}}$ by using the standard PIR capacity-achieving scheme [17] for $|U|$ messages specified by U , i.e., constructing queries $Q_i^{(1)}$ for $i \in [N]$ from a PIR scheme with N servers and $|U|$ messages.

Let us first examine the correctness and the privacy of this concatenated scheme. The correctness is an immediate consequence of the first constraint of (18), since the retrieval scheme is just a private retrieval scheme to retrieve $W_{X^{(1)}}$ from $|U|$ messages including the desired message.

For the privacy requirement, the obfuscation step constructs U that is independent of $X^{(0)}$ as a constraint, so we have

$$I(U; X^{(0)}) = 0. \quad (19)$$

Since the retrieval scheme is a standard PIR capacity-achieving scheme, we have

$$I(Q_i^{(1)}; X^{(1)}|U) = 0, \quad \forall i \in [N], \quad (20)$$

by examining the PIR scheme [17], i.e., the query to an individual server i does not leak any information about the request given the subset of messages that is of interest.

With (19) and (20), we claim that

$$I(Q_i^{(1)}; X^{(0)}) = 0,$$

which is the privacy requirement to be justified. Towards this end, consider

$$\begin{aligned} I(Q_i^{(1)}; X^{(0)}) &\leq I(Q_i^{(1)}, U; X^{(0)}) \\ &= I(U; X^{(0)}) + I(Q_i^{(1)}; X^{(0)}|U) \\ &\leq I(U; X^{(0)}) + I(Q_i^{(1)}; X^{(0)}, X^{(1)}|U) \\ &= I(Q_i^{(1)}; X^{(0)}, X^{(1)}|U), \end{aligned}$$

where $I(U; X^{(0)}) = 0$ follows from (19). Since

$$\begin{aligned} I(Q_i^{(1)}; X^{(0)}, X^{(1)}|U) &= I(Q_i^{(1)}; X^{(1)}|U) + I(Q_i^{(1)}; X^{(0)}|X^{(1)}, U) \\ &= 0, \end{aligned}$$

where $I(Q_i^{(1)}; X^{(1)}|U) = 0$ follows from (20), and $I(Q_i^{(1)}; X^{(0)}|X^{(1)}, U) = 0$ follows because $Q_i^{(1)}$ is only dependent of the random key given $X^{(1)}$ and U for the private retrieval scheme, which implies

$$X^{(0)} \rightarrow X^{(1)}, U \rightarrow Q_i^{(1)},$$

and thus $I(Q_i^{(1)}; X^{(0)}|X^{(1)}, U) = 0$. Hence, we can obtain that

$$I(Q_i^{(1)}; X^{(0)}) = 0,$$

that is to be proved.

Finally, let us evaluate the download cost of the scheme. Direct from the capacity result in [17], the download cost for a given obfuscation set U is

$$C(N, |U|) = 1 + N^{-1} + N^{-2} + \dots + N^{-|U|+1},$$

i.e., the (inverse) PIR capacity for N servers and $|U|$ messages. Hence, the download cost of this concatenation scheme is given by

$$\alpha^{(1)} = \mathbb{E}[C(N, |U|)] = \mathbb{E}\left[\left(1 - \frac{1}{N}\right)^{-1} \left(1 - \frac{1}{N^{|U|}}\right)\right],$$

and the probability distribution of U is specified by the solution to problem (18), which indeed explains why we choose $\mathbb{E}[C(N, |U|)]$ as the objective function in (18).

Therefore, we have justified that the concatenation scheme satisfies the correctness and the privacy requirements. The download cost is $\alpha^{(1)} = \mathbb{E}[C(N, |U|)]$, where U represents an obfuscation scheme and can be any feasible solution to

the problem (18). Referring to Theorem 1, the remaining part is to show the existence of some U satisfying (15), i.e., the problem (18) has a feasible solution satisfying (15).

B. Existence of an Admissible Obfuscation

As said, the last step is show that there exists a solution to the problem (18) such that the resulting U satisfies (15). Towards this end, we first interpret the problem (18) as a linear programming (LP), and write the decision variables and the objective function in a more explicit form.

The problem (18) can be viewed as a linear programming by treating each conditional probability $p(u|x^{(1)}, x^{(0)})$ as a decision variable for any given $p(x^{(1)}, x^{(0)})$, $x^{(0)}, x^{(1)} \in [K]$ and $u \in \mathcal{P}_K$. To see this, we first inspect the constraints. The first constraint $X^{(1)} \in U$ can be equivalently written by

$$p(u|x^{(1)}, x^{(0)}) = 0, \quad \forall x^{(1)} \notin u. \quad (21)$$

The second (independence) constraint can be written by

$$\begin{aligned} \sum_{x^{(1)} \in [K]} p(u|x^{(1)}, x^{(0)})p(x^{(1)}|x^{(0)}) \\ = \sum_{x^{(1)} \in [K]} p(u|x^{(1)}, \tilde{x}^{(0)})p(x^{(1)}|\tilde{x}^{(0)}), \end{aligned} \quad (22)$$

for any $x^{(0)}, \tilde{x}^{(0)} \in [K]$ and $u \in \mathcal{P}_K$. For given $p(x^{(1)}, x^{(0)})$, both constraints are clearly linear with decision variables $p(u|x^{(1)}, x^{(0)})$.

Lastly, let us examine on the objective function. Although $C(N, |U|)$ seems a power function with $|U|$, $\mathbb{E}[C(N, |U|)]$ is indeed linear with decision variables $p(u|x^{(1)}, x^{(0)})$, i.e.,

$$\begin{aligned} \mathbb{E}[C(N, |U|)] &= \sum_{u \in \mathcal{P}_K} p(u) C(N, |u|) \\ &= \sum_{u \in \mathcal{P}_K} \sum_{x^{(0)}, x^{(1)} \in [K]} p(x^{(0)}, x^{(1)})p(u|x^{(0)}, x^{(1)})C(N, |u|) \\ &= \sum_{x^{(0)}, x^{(1)}} p(x^{(0)}, x^{(1)}) \sum_{c=1}^K C(N, c) \\ &\quad \times \left(\sum_{u: |u|=c} p(u|x^{(0)}, x^{(1)}) \right), \end{aligned}$$

which is linear with $p(u|x^{(0)}, x^{(1)})$ for given $p(x^{(0)}, x^{(1)})$.

By these interpretations, we write the optimization problem in a more explicit form:

$$\begin{aligned} &\underset{p(u|x^{(1)}, x^{(0)})}{\text{minimize}} \quad \mathbb{E}[C(N, |U|)] \\ &\text{subject to (21), (22)} \\ &\quad \sum_{u \in \mathcal{P}_K} p(u|x^{(1)}, x^{(0)}) = 1, \quad \forall x^{(1)}, x^{(0)}, \\ &\quad p(u|x^{(1)}, x^{(0)}) \geq 0, \quad \forall u, x^{(1)}, x^{(0)}. \end{aligned} \quad (23)$$

It is worth noting that the problem is always feasible since

$$p(u|x^{(1)}, x^{(0)}) = \begin{cases} 1, & u = [K], \\ 0, & u \neq [K], \end{cases}$$

for $x^{(1)}, x^{(0)} \in [K]$ is always a feasible solution for any given $p(x^{(1)}, x^{(0)})$. In the context of PIR, it indicates that using the private retrieval scheme over K messages is always feasible when querying for the request $X^{(1)}$.

It should be noted that a similar LP formulation was first discussed in [33] when authors studied a so-called ON-OFF privacy problem that can also be considered as the problem of intermittent PIR with a single server, although in a slightly different setting where two random variables $X^{(0)}$ and $X^{(1)}$ may not have the same alphabet therein.

Due to the exponential blowup in the number of decision variables and constraints, solving the LP instance numerically is generally intractable. Nevertheless, the following lemma, which is borrowed from [33] and interpreted with the notation in this paper, guarantees the existence of some solution that corresponds to (15) as claimed in Theorem 1.

Lemma 1 ([33, Lemma 3]): For any given random variables $X^{(0)}, X^{(1)} \in [K]$, there exists a random variable $U \in \mathcal{P}_K$ satisfying that U is independent of $X^{(0)}$, $p(u|x^{(1)}, x^{(0)}) = 0$ for $x^{(1)} \notin u$, and

$$\mathbb{P}\{|U| \leq i\} \geq \sum_{j=1}^i \theta_j, \quad \forall i = 1, \dots, K. \quad (24)$$

In parlance of the optimization, Lemma 1 claims the existence of some feasible solution to the problem (23) (or equivalently (18)), for any given $X^{(1)}$ and $X^{(0)}$ (or $p(x^{(1)}, x^{(0)})$), such that the resulting U (or $p(u)$) satisfies the constraints in (24). It is clear that (24) is exactly the same as (15), that is to be proved in Theorem 1.

The lemma is established by a constructive proof, i.e., constructing an admissible $p(u|x^{(1)}, x^{(0)})$ for $x^{(1)}, x^{(0)} \in [K]$ and $u \in \mathcal{P}_K$, provided the given $p(x^{(1)}, x^{(0)})$, or $p(x^{(1)}|x^{(0)})$ (implying that the lemma holds for any initial probability distribution π_0). Instead of showing the detailed proof that can be found in [33], we present an example to illustrate the basic idea of the construction, or roughly speaking the basic idea of finding some particular feasible solution to the optimization problem (23).

Example: Suppose that the transition probabilities $p(x^{(1)}|x^{(0)})$ are given by

$$P = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.5 & 0.4 & 0.1 \\ 0.2 & 0.5 & 0.3 \end{bmatrix},$$

where $P_{i,j} = \mathbb{P}\{X^{(1)} = j|X^{(0)} = i\}$.

The designed probabilities $p(u, x^{(1)}|x^{(0)})$ are represented in Table IV, where the shaded cells of value 0 come immediately from the condition $p(u, x^{(1)}|x^{(0)}) = 0$ for $x^{(1)} \notin u$. Throughout this example, we will show how to fill in the values of other cells.

- $|U| = 1$: For each $i \in [K]$, choose $U = \{i\}$, and let

$$\begin{aligned} \mathbb{P}\{U = \{i\}, X^{(1)} = i|X^{(0)} = j\} \\ = \mathbb{P}\{X^{(1)} = i|X^{(0)} = v_{i,1}\} \end{aligned}$$

for all $j \in [K]$, i.e., 0.1, 0.3 and 0.1 for $i = 1, 2, 3$, respectively, where $v_{i,1}$ is defined in (10).

- $|U| = 2$: For each $i \in [K]$ and $v_{i,1}$, find a column index (of P) c_i such that

$$\begin{aligned} \mathbb{P}\{X^{(1)} = c_i|X^{(0)} = v_{i,1}\} \\ \geq \mathbb{P}\{X^{(1)} = c_i|X^{(0)} = v_{c_i,2}\} + \mu_i, \end{aligned}$$

where

$$\begin{aligned} \mu_i = \mathbb{P}\{X^{(1)} = i|X^{(0)} = v_{i,2}\} \\ - \mathbb{P}\{X^{(1)} = i|X^{(0)} = v_{i,1}\}. \end{aligned}$$

Choose $U = \{i, c_i\}$ and let

$$\mathbb{P}\{U = \{i, c_i\}, X^{(1)} = x|X^{(0)} = v_{i,j}\} = \mu_i, \quad (25)$$

for $j \geq 2$ and $x = i$ or $j < 2$ and $x = c_i$. As in this example, for $i = 1$, we have $\mu_i = 0.1$, i.e., the second minimal value minus the minimum value in the first column of P , where $v_{i,1} = 1$ and $v_{i,2} = 3$. Let $c_i = 3$. Then we can check that

$$\begin{aligned} 0.6 = \mathbb{P}\{X^{(1)} = 3|X^{(0)} = 1\} \\ \geq \mathbb{P}\{X^{(1)} = 3|X^{(0)} = v_{3,2}\} + 0.1, \end{aligned}$$

where $v_{3,2} = 2$ and hence $\mathbb{P}\{X^{(1)} = 3|X^{(0)} = v_{3,2}\} = 0.1$. The process for $i = 1$ finally configures the value 0.1 for $U = \{1, 3\}$ in the table.

This generally explains why we call it an obfuscation scheme. For each $i \in [K]$, we carefully find an index c_i for $v_{i,1}$ and mix it with i to form a set U such that when observing U , there exists a pair $(x^{(1)}, x^{(0)})$ generating U for all $x^{(0)} \in [K]$. Note that since for different $i \in [K]$, the set U may be the same, e.g., $U = \{1, 3\}$ for both $i = 1$ and $i = 3$, so $p(u, x^{(1)}|x^{(0)})$ is configured in an augmented way, i.e., the right-hand side of (25) is added to the left-hand side instead of being overwritten, such as $0.1 + 0.2$ in the cell.

- $|U| = 3$: Configure all remaining values constrained by $p(x^{(1)}|x^{(0)})$, i.e., the summation of each row in the table.

Remark 1: The general algorithm would basically extend the above process for $|U| = 2$. Roughly speaking, for $|U| = c = 1, \dots, \sigma$ and each $i \in [K]$, find an index $c_{i,j}$ for each $v_{i,j}$ such that $j \leq c - 1$. Then choose $U = \{i, c_{i,j} : j \leq c - 1\}$ and configure

$$\begin{aligned} \mathbb{P}\{U, X^{(1)} = x|X^{(0)} = v_{i,j}\} \\ = \mathbb{P}\{X^{(1)} = i|X^{(0)} = v_{i,c}\} \\ - \mathbb{P}\{X^{(1)} = i|X^{(0)} = v_{i,c-1}\}, \end{aligned}$$

for $j \geq c$ and $x = i$ or $j < c$ and $x = c_{i,j}$. It is worth noting that $c_{i,j}$ may be the same for different j , so the size of U may be smaller than c . This observation indeed leverages the inequality (24) in the lemma, where the worst case is

$$\mathbb{P}\{|U| = i\} = \theta_i, \quad \forall i = 1, \dots, K,$$

as mentioned.

TABLE IV
THE CONSTRUCTED $p(u, x^{(1)}|x^{(0)})$ FOR THE GIVEN $p(x^{(1)}|x^{(0)})$

$X^{(0)} \backslash U$		$\{1\}$	$\{2\}$	$\{3\}$	$\{1, 2\}$	$\{1, 3\}$	$\{2, 3\}$	$\{1, 2, 3\}$	$P_{i,j}$
1	1	0.1	0	0	0	0	0	0	0.1
	2	0	0.3	0	0	0	0	0	0.3
	3	0	0	0.1	0	0.1 + 0.2	0.1	0.1	0.6
2	1	0.1	0	0	0	0.1 + 0.2	0	0.1	0.5
	2	0	0.3	0	0	0	0.1	0	0.4
	3	0	0	0.1	0	0	0	0	0.1
3	1	0.1	0	0	0	0.1	0	0	0.2
	2	0	0.3	0	0	0	0.1	0.1	0.5
	3	0	0	0.1	0	0.2	0	0	0.3

IV. GENERAL CASE: MARKOV CHAIN

In this section, we will show how to use the two-requests scheme in Section III-A as a building block to design an intermittent PIR scheme over time when the requests $X^{(t)}$, $t = 0, 1, \dots$ form a Markov chain.

First, let

$$\tau(t) := \max\{j : j \in \mathcal{P}_t\}, \quad (26)$$

where \mathcal{P}_t is defined in (4). We may write $\tau(t)$ by τ for notational simplicity when the time index t is clear in the context. Note that \mathcal{P}_t is completely determined by the privacy status $S^{(t)}$, which is chosen by the user. Roughly speaking, $\tau(t)$ represents the latest time that the user needed privacy at time t .

Then the following proposition is a direct but useful consequence of the assumption of Markov structure correlation, and its proof is deferred to the appendix.

Proposition 2: For any $i \in [N]$, if $Q_i^{(t)}$ is independent of $X^{(\tau(t))}$ conditioning on $Q_i^{(0)}, \dots, Q_i^{(t-1)}$, then $Q_i^{(t)}$ is independent of $X^{(\mathcal{P}_t)}$ conditioning on $Q_i^{(0)}, \dots, Q_i^{(t-1)}$, i.e.,

$$I(X^{(\tau(t))}; Q_i^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}) = 0, \quad (27)$$

implies that

$$I(X^{(\mathcal{P}_t)}; Q_i^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}) = 0. \quad (28)$$

From Proposition 2, we know that it is sufficient to design queries $Q_i^{(t)}$ for $i \in [N]$ satisfying (27), in order to guarantee the desired privacy (3). Roughly speaking, at time t , we need to design queries $Q_i^{(t)}, i \in [N]$ for the request $X^{(t)}$ while preserving the privacy of the request $X^{(\tau)}$.

Recall the scheme for the two-requests case in Section III, where we design queries $Q_i^{(1)}, i \in [N]$ for the request $X^{(1)}$ while preserving the privacy of the request $X^{(0)}$. The roles of $X^{(0)}$ and $X^{(1)}$ are similar to $X^{(\tau)}$ and $X^{(t)}$ in this section, where the queries are designed for the retrieval purpose of the current request $X^{(t)}$ but they have to preserve privacy of some previous request $X^{(\tau)}$. Therefore, the scheme for the general Markov case is indeed similar to the canonical case of two requests that was discussed, where the main difference is that the prior distribution (e.g., $p(x^{(1)}|x^{(0)})$ in the previous section) as an input has to be updated at each time. In particular, let $U^{(t)}$ denote the obfuscation set at time t , and

$$p_{X^{(t)}|X^{(\tau)}}(x^{(t)}|x^{(\tau)}, u^{(0)}, \dots, u^{(t-1)}), \quad (29)$$

which serves the same role as $p(x^{(1)}|x^{(0)})$ in the two-requests case, has to be updated according to the generated $u^{(t)}$ at each time period. For notational simplicity, let $U^{[0:t-1]} := \{U^{(0)}, \dots, U^{(t-1)}\}$.

We summarize the above intuition by presenting a result that is similar to Theorem 1. Before that, we introduce a necessary notation that is similar to the one defined in Section III. Let $\theta_j(u^{([0:t-1])})$ be defined the same as θ_j in (13), but for the conditional probabilities $p_{X^{(t)}|X^{(\tau)}}(x^{(t)}|x^{(\tau)}, u^{(0)}, \dots, u^{(t-1)})$ as shown in (29) for any given realizations $u^{([0:t-1])}$. Note that the random variables here are $X^{(t)}$ and $X^{(\tau)}$, serving the same roles as $X^{(1)}$ and $X^{(0)}$ in the definitions (10), (11), (12) and (13).

Theorem 2: Suppose that requests $X^{(t)}$ for $t = 0, 1, \dots$ form a Markov chain and the privacy status $S^{(t)}$ for $t = 0, 1, \dots$ are given. There exists an intermittent private information retrieval scheme with download cost

$$\alpha^{(t)} = \mathbb{E} \left[\left(1 - \frac{1}{N}\right)^{-1} \left(1 - \frac{1}{N^{|U^{(t)}|}}\right) \right], \quad (30)$$

for some random variable $U^{(t)}$ that takes value in the power set of $[K]$ such that

$$\mathbb{P} \left\{ |U^{(t)}| \leq i | U^{([0:t-1])} = u^{([0:t-1])} \right\} \geq \sum_{j=1}^i \theta_j(u^{([0:t-1])}), \quad (31)$$

for $i = 1, \dots, K$ and any $u^{([0:t-1])}$.

Remark 2: Similar to Corollary 1, a slightly weaker but more explicit form of (31) is that

$$\mathbb{P} \left\{ |U^{(t)}| = j | U^{([0:t-1])} = u^{([0:t-1])} \right\} = \theta_j(u^{([0:t-1])}), \quad (32)$$

for $j = 1, \dots, K$ and any $u^{([0:t-1])}$.

As said, the concatenation scheme that justifies the theorem can be described in the same manner as we did in the two-requests case, i.e.,

- 1) Design an obfuscation set $U^{(t)}$ that includes $X^{(t)}$ and is independent of $X^{(\tau)}$ conditioning on $U^{([0:t-1])}$, i.e., sampling an obfuscation set $u^{(t)}$ according to

$$p_{U^{(t)}|X^{(t)}, X^{(\tau)}}(u^{(t)}|x^{(t)}, x^{(\tau)}, u^{([0:t-1])}),$$

that corresponds to an obfuscation scheme which is discussed right after the enumeration.

- 2) Query for the request $X^{(t)}$ by using a standard PIR scheme over messages specified by $U^{(t)}$.

In particular, we modify the optimization problem (23) therein associated with the obfuscation scheme (the first step of the above) to incorporate the previously released $u^{([0:t-1])}$, where (21) and (22) can be correspondingly modified by

$$p_{U^{(t)}|X^{(t)},X^{(\tau)}}(u^{(t)}|x^{(t)},x^{(\tau)},u^{([0:t-1])})=0, \quad \forall x^{(t)} \notin u^{(t)}, \quad (33)$$

and

$$\begin{aligned} & \sum_{x^{(t)}} p(u^{(t)}|x^{(t)},x^{(\tau)},u^{([0:t-1])}) p(x^{(t)}|x^{(\tau)},u^{([0:t-1])}) \\ &= \sum_{x^{(t)}} p(u^{(t)}|x^{(t)},\tilde{x}^{(\tau)},u^{([0:t-1])}) p(x^{(t)}|\tilde{x}^{(\tau)},u^{([0:t-1])}), \end{aligned} \quad (34)$$

for any $x^{(\tau)}, \tilde{x}^{(\tau)} \in [K]$ and $u^{(t)} \in \mathcal{P}_K$. Note that the probability simplex (the third constraint in (23)) and the nonnegativity of probabilities (the fourth constraint in (23)) always have to be guaranteed. For conciseness, we omit to write them explicitly. Hence, the optimization problem can be written by

$$\begin{aligned} & \text{minimize} \quad \mathbb{E} [C(N, |U^{(t)}|)] \\ & \text{subject to (33), (34).} \end{aligned} \quad (35)$$

In short, the problem (35) can be simply modified from (23) by replacing the given distribution $p_{X^{(1)}|X^{(0)}}(x^{(1)}|x^{(0)})$ and the decision variables $p_{U^{(1)}|X^{(1)},X^{(0)}}(u^{(1)}|x^{(1)},x^{(0)})$ in (23) by $p_{X^{(t)}|X^{(\tau)}}(x^{(t)}|x^{(\tau)},u^{([0:t-1])})$ and

$$p_{U^{(t)}|X^{(t)},X^{(\tau)}}(u^{(t)}|x^{(t)},x^{(\tau)},u^{([0:t-1])}),$$

respectively.

As we know from Lemma 1, for any given $u^{([0:t-1])}$, if $p_{X^{(t)}|X^{(\tau)}}(x^{(t)}|x^{(\tau)},u^{([0:t-1])})$ is known, then the problem (35) has a feasible solution

$$p_{U^{(t)}|X^{(t)},X^{(\tau)}}(u^{(t)}|x^{(t)},x^{(\tau)},u^{([0:t-1])})$$

for $u^{(t)} \in \mathcal{P}_K$ and $x^{(t)}, x^{(\tau)} \in [K]$, such that the resulting $U^{(t)}$ satisfies that

$$\mathbb{P} \{ |U^{(t)}| \leq i | U^{([0:t-1])} = u^{([0:t-1])} \} \geq \sum_{j=1}^i \theta_j(u^{([0:t-1])}),$$

for all $i = 1, \dots, K$, as shown in (31). Finally, we use the obtained $p_{U^{(t)}|X^{(t)},X^{(\tau)}}(u^{(t)}|x^{(t)},x^{(\tau)},u^{([0:t-1])})$ to sample an obfuscation set $u^{(t)}$.

It should be noted that solving (35), or more precisely obtaining the obfuscation set sampling distribution $p_{U^{(t)}|X^{(t)},X^{(\tau)}}(u^{(t)}|x^{(t)},x^{(\tau)},u^{([0:t-1])})$ as discussed relies on knowing $p_{X^{(t)}|X^{(\tau)}}(x^{(t)}|x^{(\tau)},u^{([0:t-1])})$. However, these quantities can not be obtained directly from the transition probabilities of the given Markov chain, since it encompasses the previously generated obfuscation sets $u^{([0:t-1])}$. As such, we need to track $p_{X^{(t)}|X^{(\tau)}}(x^{(t)}|x^{(\tau)},u^{([0:t-1])})$ over time t . Roughly speaking, $p_{X^{(t)}|X^{(\tau)}}(x^{(t)}|x^{(\tau)},u^{([0:t-1])})$ is the ‘‘prior’’ distribution at time t that needs to be updated at each time t .

Remark 3: If $S^{(t)} = 1$, i.e., the request $X^{(t)}$ needs privacy, then we know that $\tau(t) = t$ by definition, which implies that

$$\theta_j(u^{([0:t-1])}) = \begin{cases} 0, & j < K, \\ 1, & j = K, \end{cases}$$

for any $u^{([0:t-1])}$. Referring to (32), it suggests that

$$\mathbb{P} \{ |U^{(t)}| = K | U^{([0:t-1])} = u^{([0:t-1])} \} = 1,$$

which is consistent with our early observation, i.e., using a standard PIR scheme (over all K messages) if privacy is needed at time t . In other words, the above analysis of designing an obfuscation set $U^{(t)}$ unifies both cases $S^{(t)} = 0$ and $S^{(t)} = 1$.

Now, we show how to track $p_{X^{(t)}|X^{(\tau)}}(x^{(t)}|x^{(\tau)},u^{([0:t-1])})$, that is equivalent to $p_{X^{(t)}|X^{(\tau)}}(x^{(t)}|x^{(\tau)},u^{([0:t-1])})$ as we need, for $t = 0, 1, \dots$. The process is essentially similar to the standard forward algorithm [35] by utilizing the Markov structure and incorporating the designed $p_{U^{(t)}|X^{(t)},X^{(\tau)}}(u^{(t)}|x^{(t)},x^{(\tau)},u^{([0:t-1])})$.

Recall the assumption that $S^{(0)} = 1$ (implying $\tau(0) = 0$) and initial probability distribution π_0 of $X^{(0)}$ is known, so $p_{X^{(t)}|X^{(\tau)}}(x^{(t)}|x^{(\tau)},u^{([0:t-1])})$ is known for $t = 0$, i.e.,

$$p_{X^{(t)}|X^{(\tau)}}(x^{(t)}|x^{(\tau)},u^{([0:t-1])}) = \pi_0.$$

At each time $t = 1, 2, \dots$, we consider $S^{(t)} = 0$ or 1 separately.

If $S^{(t)} = 1$, then we know that $\tau(t) = t$ by the definition of $\tau(t)$. Consider

$$\begin{aligned} & p(x^{(t+1)}, x^{(\tau(t+1))} | u^{([0:t])}) \\ & \stackrel{(a)}{=} \sum_{x^{(t)}} p(x^{(t)} | u^{([0:t])}) p(x^{(t+1)}, x^{(\tau(t+1))} | x^{(t)}) \\ & \stackrel{(b)}{=} \sum_{x^{(t)}} p(x^{(t)}, x^{(\tau(t))} | u^{([0:t-1])}) p(x^{(t+1)}, x^{(\tau(t+1))} | x^{(t)}), \end{aligned}$$

where (a) follows because $\tau(t+1)$ is either t or $t+1$ provided that $S^{(t)} = 1$, (b) follows because $U^{(t)} = [K]$ that is a constant and $\tau(t) = t$. Since $p(x^{(t+1)}, x^{(\tau(t+1))} | x^{(t)})$ can be obtained straightforwardly from the transition probabilities of the Markov chain, $p(x^{(t+1)}, x^{(\tau(t+1))} | u^{([0:t])})$ can be updated from $p(x^{(t)}, x^{(\tau(t))} | u^{([0:t-1])})$.

If $S^{(t)} = 0$, then we know that $\tau(t) = \tau(t-1)$ by definition. Consider

$$\begin{aligned} & p(x^{(t+1)}, x^{(\tau(t+1))} | u^{([0:t])}) \\ & \stackrel{(a)}{=} \sum_{x^{(t)}} p(x^{(t)}, x^{(\tau(t))} | u^{([0:t])}) p(x^{(t+1)}, x^{(\tau(t+1))} | x^{(t)}, x^{(\tau(t))}) \\ & \propto \sum_{x^{(t)}} p(x^{(t)}, x^{(\tau(t))}, u^{(t)} | u^{([0:t-1])}) \\ & \quad \times p(x^{(t+1)}, x^{(\tau(t+1))} | x^{(t)}, x^{(\tau(t))}) \\ & = \sum_{x^{(t)}} p(x^{(t)}, x^{(\tau(t))} | u^{([0:t-1])}) p(u^{(t)} | x^{(t)}, x^{(\tau(t))}, u^{([0:t-1])}) \\ & \quad \times p(x^{(t+1)}, x^{(\tau(t+1))} | x^{(t)}, x^{(\tau(t))}), \end{aligned}$$

where (a) follows because $\tau(t+1)$ is either $\tau(t)$ or $t+1$. Since $p(u^{(t)} | x^{(t)}, x^{(\tau(t))}, u^{([0:t-1])})$ is the obfuscation sampling distribution by design, and $p(x^{(t+1)}, x^{(\tau(t+1))} | x^{(t)}, x^{(\tau(t))})$ can

be directly obtained from the transition probabilities of the Markov chain, $p(x^{(t+1)}, x^{(\tau(t+1))} | u^{([0:t])})$ can be updated from $p(x^{(t)}, x^{(\tau(t))} | u^{([0:t-1])})$.

Therefore, the above process keeps tracking the probability distribution $p_{X^{(t)}, X^{(\tau(t))}}(x^{(t)}, x^{(\tau(t))} | u^{([0:t-1])})$ for $t = 0, 1, \dots$, that is needed for the obfuscation set design.

We summarize the proposed intermittent PIR scheme as follows: at time t , with the known probability distribution $p_{X^{(t)}, X^{(\tau(t))}}(x^{(t)}, x^{(\tau(t))} | u^{([0:t-1])})$, where $u^{([0:t-1])}$ are previously generated obfuscation sets from time 0 to $t-1$,

- 1) Design a sampling distribution of the obfuscation set $U^{(t)}$, i.e.,

$$p_{U^{(t)} | X^{(t)}, X^{(\tau(t))}}(u^{(t)} | x^{(t)}, x^{(\tau(t))}, u^{([0:t-1])}),$$

such that $\mathbb{P}\{X^{(t)} \in U^{(t)}\} = 1$ and $U^{(t)}$ is independent of $X^{(\tau(t))}$ given the previous obfuscation sets $u^{([0:t-1])}$, i.e.,

$$I(X^{(\tau(t))}; U^{(t)} | U^{([0:t-1])} = u^{([0:t-1])}) = 0. \quad (36)$$

The sampling distribution of the obfuscation set $U^{(t)}$ can be any feasible solution to the optimization problem (35) for known $p_{X^{(t)}, X^{(\tau(t))}}(x^{(t)}, x^{(\tau(t))} | u^{([0:t-1])})$, and there is an existence guarantee of a feasible solution such that the resulting $U^{(t)}$ satisfies (31) for any given $p_{X^{(t)}, X^{(\tau(t))}}(x^{(t)}, x^{(\tau(t))} | u^{([0:t-1])})$.

- 2) Generate an obfuscation set $u^{(t)}$ according to the designed sampling distribution

$$p_{U^{(t)} | X^{(t)}, X^{(\tau(t))}}(u^{(t)} | x^{(t)}, x^{(\tau(t))}, u^{([0:t-1])}),$$

based on the requests $x^{(\tau(t))}$, $x^{(t)}$ and previously generated $u^{([0:t-1])}$.

- 3) Query for the request $x^{(t)}$ by using a standard PIR scheme over messages specified by $u^{(t)}$.
- 4) Compute

$$p_{X^{(t+1)}, X^{(\tau(t+1))}}(x^{(t+1)}, x^{(\tau(t+1))} | u^{([0:t])})$$

from the known

$$p_{X^{(t)}, X^{(\tau(t))}}(x^{(t)}, x^{(\tau(t))} | u^{([0:t-1])}),$$

the given transition probabilities of the Markov chain, and the designed obfuscation set sampling distribution

$$p_{U^{(t)} | X^{(t)}, X^{(\tau(t))}}(u^{(t)} | x^{(t)}, x^{(\tau(t))}, u^{([0:t-1])}).$$

Finally, let us verify that the scheme satisfies the correctness requirement and the privacy requirement (c.f. (3)) formally. The proposed scheme guarantees that the desired message can be retrieved successfully by design, since the retrieval phase is a standard PIR scheme (for the request $X^{(t)}$) over messages specified by $U^{(t)}$. From Proposition 2, we know that if the scheme satisfies (27) then it satisfies the desired privacy requirement in (3). Since the immediate privacy guarantee of the scheme is that

$$I(X^{(\tau)}; U^{(t)} | U^{([0:t-1])}) = 0,$$

which is guaranteed by (36) during the design, we need to show that it implies that $I(X^{(\tau)}; Q_i^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}) = 0$, that is (27) to be justified. Towards this end, consider

$$\begin{aligned} & I(X^{(\tau)}; Q_i^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}) \\ &= H(X^{(\tau)} | Q_i^{([0:t-1])}) - H(X^{(\tau)} | Q_i^{(t)}, Q_i^{([0:t-1])}) \\ &\stackrel{(a)}{\leq} H(X^{(\tau)} | Q_i^{([0:t-1])}) - H(X^{(\tau)} | U^{([0:t])}) \\ &\stackrel{(b)}{\leq} H(X^{(\tau)} | U^{([0:t-1])}) - H(X^{(\tau)} | U^{([0:t])}) \\ &= I(X^{(\tau)}; U^{(t)} | U^{([0:t-1])}) \\ &= 0, \end{aligned}$$

where (a) follows because the fact that $Q_i^{(t)}$ is generated by a PIR scheme over messages in $U^{(t)}$ implies the following chain

$$X^{(\tau)} \rightarrow X^{(t)} \rightarrow U^{(t)} \rightarrow Q_i^{(t)},$$

and (b) follows because $U^{(t)}$ is deterministic of $Q_i^{(t)}$ in a standard PIR scheme, i.e., knowing one of the messages in $U^{(t)}$ is being retrieved from the query $Q_i^{(t)}$ to the i -th server, which finishes the justification.

It is generally hard to obtain a closed-form formula of the download cost $\alpha^{(t)}$ at time t , since $\theta_j(u^{([0:t-1])})$ is not simply a function of the given transition probabilities as in the two-requests case. For this reason, we present an evaluation at the end to illustrate the download cost.

Evaluation: We evaluate the download cost $\alpha^{(t)}$ specified by (30) and (32) for the simplest case $N = K = 2$ as an illustration in Figure 1. Similar to the previous example, suppose that the probability transition matrix of the Markov chain is

$$P = \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix},$$

such that $0 \leq \alpha, \beta \leq 1$. Since the privacy status $S^{(t)}$ affects the download cost via $\tau(t)$ (c.f. (26)), we simulate the download cost, as a function of $t - \tau(t)$ for several values of $\alpha + \beta$, where the maximum value 1.5 of y-axis is the download cost of a standard PIR scheme for $N = K = 2$, and the minimum value 1 corresponds to retrieval of the desired message directly. We can observe that as $\alpha + \beta$ approaches 1, the correlation between the requests decreases, which leads to a decrease in the download cost. As $t - \tau$ goes larger, the correlation between the current request and the latest request that needs privacy decreases, which also leads to a decrease in the download cost.

V. APPLICATION TO LOCATION PRIVACY

As said, the reason why we are particularly interested in the Markov structure correlation is because of the motivating location privacy application. In this section, we will show how we apply the proposed intermittent PIR scheme to design an obfuscation-based location privacy protection mechanism, and discuss some specific aspects of the location privacy problem.

As we mentioned, a commonly adopted mobility model of the location trace is the Markov model [7]–[12], i.e., the location at time (discrete time-stamp) t is denoted by $X^{(t)}$ and $X^{(t)}, t = 0, 1, \dots$ form a first-order Markov chain.

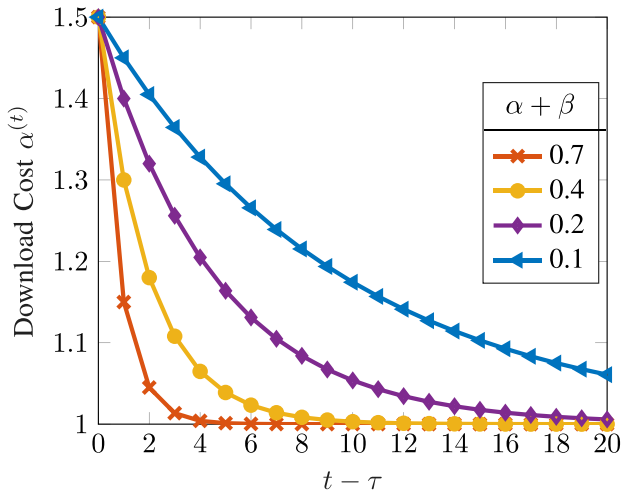


Fig. 1. The download cost $\alpha^{(t)}$ for $N = K = 2$, as a function of $t - \tau$ for different values of $\alpha + \beta$.

Assume that each $X^{(t)}$ takes values in a common alphabet $[K] = \{1, \dots, K\}$.

The user may want to share his/her location with some service providers (SPs), in order to receive location-based services. In this section, we model the provided service by an information retrieval, i.e., the user sends his/her location to a SP, and then the SP responds by sending some contents according to the location. In other words, we are interested in the case such that downloading is a concern for the service quality. Also, we assume that there are multiple service providers who can provide alternative services, e.g., querying through a cloud.

To protect the location privacy, a user may send a perturbed location to the SPs instead of the true location by sacrificing the service quality to some degree while preserving the privacy in some range. Many works [11]–[13] have been done to study the location privacy problem with different notions of privacy and utility metrics. The closest one to this paper is [11], where the privacy notion is information-theoretic, i.e., defined by the mutual information between true location trace and the released perturbation of locations, and the utility is defined by a non-specified distortion function.

As we keep motivating in this paper, the user may only be concerned about the privacy of some locations while others can be released without any concern about the privacy. For a time t , let $\mathcal{P}_t \subseteq [0 : t]$ be the given set such that $X^{(t)}$ requires privacy if and only if $t \in \mathcal{P}$. The set \mathcal{P}_t is supposed to be determined by the user, and viewed as a given parameter. The same as our discussion about intermittent PIR, the essential difference between the situation here and protecting a single location [15], is that the user has to be careful when releasing the location he/she does not care about the privacy, since the location that needs privacy may be inferred due to the temporal correlation in the location trace.

Our focus is on an extreme operational point such that privacy leakage is zero and the utility is maximized. Different from the distortion-based mechanism [11], [12], the location privacy protection mechanism in this section is obfuscation-based, i.e., mixing the true location with certain

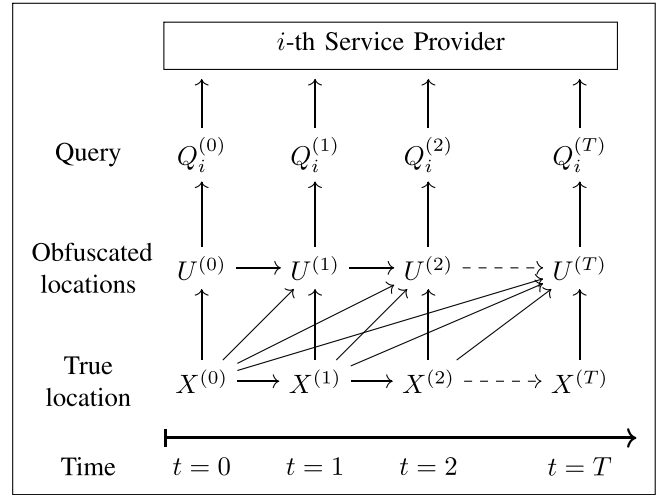


Fig. 2. An obfuscation-based location privacy protection mechanism.

perturbed locations together and requesting the obfuscation set from SPs.

The application of our proposed intermittent private information retrieval scheme to this specific location privacy problem is straightforward, by viewing the true location $X^{(t)}$ as the request in previous sections, i.e.,

$$\text{true location} \xleftrightarrow{X^{(t)}} \text{request}.$$

Therefore, we can directly transplant the proposed scheme in previous sections to obtain an obfuscation-based location privacy protection mechanism, as shown in Figure 2, such that at time t ,

$$I(X^{(\mathcal{P}_t)}; Q_i^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}) = 0, \quad (37)$$

where $X^{(\mathcal{P}_t)} = \{X^{(i)} : i \in \mathcal{P}_t\}$, i.e., previous locations (before time t) that need to be protected.

The mapping from the intermittent PIR scheme to the location privacy protection mechanism as illustrated in Figure 2 should be straightforward, so we skip repeating details that can be found in previous sections. Instead, we discuss some issues regarding the privacy and utility metric in the location privacy context.

Privacy Metric: The privacy notion, i.e., left-hand side of (37), of our approach is essentially the same as the so-called *online privacy* in [11], where the privacy is measured by

$$\sum_t I(X^{(\mathcal{P}_t)}; Q_i^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}),$$

i.e., the accumulation of that in (37). Since we require the stringent zero leakage, the summation over t makes no difference due to the nonnegativity of the mutual information. Roughly speaking, the online privacy (37) guarantees that given all previously released queries, the current query leaks zero information of all previous true locations that need privacy.

A similar privacy notion, namely *offline privacy*,

$$I(X^{(\mathcal{P}_T)}; Q_i^{(0)}, \dots, Q_i^{(T)}), \quad (38)$$

by assuming a finite time period T for the sake of definition, was introduced in [11], where the authors argued that offline privacy is generally intractable to manage.

The online privacy and the offline privacy are indeed the same as the privacy requirement (3) and (5) discussed in Section II. As we explained, the online privacy requirement closely adheres to the causal nature of the scheme, where the query $Q_i^{(t)}$ has to be generated at time t instantly with causal information only. Also, under the stringent privacy requirement of zero leakage considered in this paper, the offline privacy metric may induce a trivial solution that the query is independent of $X^{(t)}$, which generally sacrifices the service quality too much.

Therefore, we consider the online privacy metric in this section. However, we have to admit that the offline privacy is theoretically interesting under a relaxed privacy requirement where privacy leakage is allowed. Indeed, [12] studied this notion of privacy in the framework of a Markov decision process with states $X^{(t)}$. Conceptually, the notion of offline privacy encompasses the concept of preventing the adversary from inferring future locations, while the notion of online privacy only considers the protection of locations that have been sampled.

Utility Metric: Since we model the provided service by an information retrieval process that accommodates the obfuscation-based mechanism, the content associated with the true location can be obtained perfectly, i.e., query accuracy is perfect, by downloading more than necessary, which is different from the distortion-based mechanism that asks for the content of a perturbed location. In this sense, we consider the download cost as a utility metric to fit the obfuscation-based framework.

We would like to slightly clarify the utility metric, as it looks different from the conventional notion, e.g., [11], [12], where the utility is measured by a single-letter distortion between the query and the true location. Since the location privacy protection mechanism is operated from another perspective in this paper, where the location privacy protection mechanism would share an obfuscated version of the true location, the quality of service is largely decided by the overhead of the content downloaded from the SPs. A motivating example here is that the user may download the map information for a larger range than he/she needs to hide the true location in some situations, e.g., augmented reality games and self-driving cars.

VI. CONCLUSION

In this paper, we study the problem of intermittent private information retrieval with Markov structure correlation, where only part of the requests need privacy. We propose an intermittent private information retrieval scheme concatenating an obfuscation scheme and a standard PIR scheme to prevent leakage over time. The download cost is reduced compared to a standard PIR scheme, at the time when privacy is not needed. Since the Markov structure correlation is motivated by the location privacy problem, we end up by applying the proposed intermittent private information retrieval scheme to design a location privacy protection mechanism and discussing some specific issues in the location privacy problem.

APPENDIX A PROOF OF PROPOSITION 1

Consider

$$\begin{aligned} I(X^{(0)}; Q_i^{(1)}) &= I(X^{(0)}, Q_i^{(0)}; Q_i^{(1)}) - I(Q_i^{(0)}; Q_i^{(1)} | X^{(0)}) \\ &= I(X^{(0)}, Q_i^{(0)}; Q_i^{(1)}) \\ &= I(Q_i^{(0)}; Q_i^{(1)}) + I(X^{(0)}; Q_i^{(1)} | Q_i^{(0)}) \\ &= I(Q_i^{(0)}; Q_i^{(1)}) + I(X^{(0)}; Q_i^{(1)}, Q_i^{(0)}) - I(X^{(0)}; Q_i^{(0)}) \\ &= I(Q_i^{(0)}; Q_i^{(1)}) + I(X^{(0)}; Q_i^{(1)}, Q_i^{(0)}). \end{aligned}$$

Since

$$\begin{aligned} I(Q_i^{(0)}; Q_i^{(1)}) &\leq I(Q_i^{(0)}; Q_i^{(1)}, X^{(0)}) \\ &= I(Q_i^{(0)}; X^{(0)}) + I(Q_i^{(1)}; Q_i^{(0)} | X^{(0)}) \\ &= 0, \end{aligned}$$

we know that $I(Q_i^{(0)}; Q_i^{(1)}) = 0$ by the nonnegativity of the mutual information, and hence

$$I(X^{(0)}; Q_i^{(1)}) = I(X^{(0)}; Q_i^{(1)}, Q_i^{(0)}),$$

which implies that $I(X^{(0)}; Q_i^{(1)}, Q_i^{(0)}) = 0$ if and only if $I(X^{(0)}; Q_i^{(1)}) = 0$.

APPENDIX B PROOF OF PROPOSITION 2

The proof follows simply from the Markov structure. Consider

$$\begin{aligned} I(X^{(\mathcal{P}_t)}; Q_i^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}) &= I(X^{(\tau(t))}; Q_i^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}) \\ &\quad + I(X^{(\mathcal{P}_t \setminus \tau(t))}; Q_i^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}, X^{(\tau(t))}). \end{aligned}$$

The second term can be bounded by

$$\begin{aligned} &I(X^{(\mathcal{P}_t \setminus \tau(t))}; Q_i^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}, X^{(\tau(t))}) \\ &\leq I(X^{(\mathcal{P}_t \setminus \tau(t))}; X^{(t)}, Q_i^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}, X^{(\tau(t))}) \\ &= I(X^{(\mathcal{P}_t \setminus \tau(t))}; X^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}, X^{(\tau(t))}) \\ &\quad + I(X^{(\mathcal{P}_t \setminus \tau(t))}; Q_i^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}, X^{(\tau(t))}, X^{(t)}) \\ &\stackrel{(a)}{=} I(X^{(\mathcal{P}_t \setminus \tau(t))}; X^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}, X^{(\tau(t))}) \\ &\stackrel{(b)}{=} 0, \end{aligned}$$

where (a) follows because $Q_i^{(t)}$ is a stochastic function of $X^{(\tau(t))}$, $X^{(t)}$ and $Q_i^{(0)}, \dots, Q_i^{(t-1)}$, and (b) follows because the Markov structure of $X^{(t)}$ for $t = 0, 1, \dots$ and $t \geq \tau(t) \geq \max \mathcal{P}_t \setminus \tau(t)$ by the definition of $\tau(t)$.

Therefore, we obtain that

$$\begin{aligned} I(X^{(\mathcal{P}_t)}; Q_i^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}) &\leq I(X^{(\tau(t))}; Q_i^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}). \end{aligned}$$

Due to the nonnegativity of the mutual information, it is clear that

$$I(X^{(\tau(t))}; Q_i^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}) = 0,$$

implies that

$$I(X^{(P_t)}; Q_i^{(t)} | Q_i^{(0)}, \dots, Q_i^{(t-1)}) = 0,$$

which completes the proof.

REFERENCES

- [1] L. Sweeney, “ k -anonymity: A model for protecting privacy,” *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 10, no. 5, pp. 557–570, 2002.
- [2] C. Dwork, “Differential privacy: A survey of results,” in *Proc. Int. Conf. Theory Appl. Models Comput.* Berlin, Germany: Springer, 2008, pp. 1–19.
- [3] M. Bezzi, “An information theoretic approach for privacy metrics,” *Trans. Data Privacy*, vol. 3, no. 3, pp. 199–215, Dec. 2010.
- [4] R. Shokri, C. Troncoso, C. Diaz, J. Freudiger, and J.-P. Hubaux, “Unraveling an old cloak: k -anonymity for location privacy,” in *Proc. 9th Annu. ACM Workshop Privacy Electron. Soc. (WPES)*, 2010, pp. 115–118.
- [5] B. Gedik and L. Liu, “Protecting location privacy with personalized k -anonymity: Architecture and algorithms,” *IEEE Trans. Mobile Comput.*, vol. 7, no. 1, pp. 1–18, Jan. 2008.
- [6] J. Hua, W. Tong, F. Xu, and S. Zhong, “A geo-indistinguishable location perturbation mechanism for location-based services supporting frequent queries,” *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 5, pp. 1155–1168, May 2018.
- [7] K. Chatzikokolakis, C. Palamidessi, and M. Stronati, “A predictive differentially-private mechanism for mobility traces,” in *Proc. Int. Symp. Privacy Enhancing Technol. Symp.* Cham, Switzerland: Springer, 2014, pp. 21–41.
- [8] Y. Xiao and L. Xiong, “Protecting locations with differential privacy under temporal correlations,” in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2015, pp. 1298–1309.
- [9] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, “Quantifying location privacy,” in *Proc. IEEE Symp. Secur. Privacy*, May 2011, pp. 247–262.
- [10] R. Shokri, G. Theodorakopoulos, and C. Troncoso, “Privacy games along location traces: A game-theoretic framework for optimizing location privacy,” *ACM Trans. Privacy Secur.*, vol. 19, no. 4, pp. 1–31, Feb. 2017.
- [11] W. Zhang, M. Li, R. Tandon, and H. Li, “Online location trace privacy: An information theoretic approach,” *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 1, pp. 235–250, Jan. 2019.
- [12] E. Erdemir, P. L. Dragotti, and D. Gunduz, “Privacy-aware time-series data sharing with deep reinforcement learning,” *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 389–401, 2021.
- [13] S. Oya, C. Troncoso, and F. Pérez-González, “Back to the drawing board: Revisiting the design of optimal location privacy-preserving mechanisms,” in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2017, pp. 1959–1972.
- [14] V. Bindschaedler and R. Shokri, “Synthesizing plausible privacy-preserving location traces,” in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2016, pp. 546–563.
- [15] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, “Protecting location privacy: Optimal strategy against localization attacks,” in *Proc. ACM Conf. Comput. Commun. Secur. (CCS)*, 2012, pp. 617–627.
- [16] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, “Private information retrieval,” in *Proc. IEEE 36th Annu. Found. Comput. Sci.*, Oct. 1995, pp. 41–50.
- [17] H. Sun and S. A. Jafar, “The capacity of private information retrieval,” *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4075–4088, Jul. 2017.
- [18] 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.
- [19] 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.
- [20] R. Tajeddine, O. W. Gnilke, and S. E. 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.
- [21] S. Kadhe, B. Garcia, A. Heidarzadeh, S. E. Rouayheb, and A. Sprintson, “Private information retrieval with side information,” *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2032–2043, Apr. 2020.
- [22] 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.
- [23] K. Banawan and S. Ulukus, “The capacity of private information retrieval from Byzantine and colluding databases,” *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 1206–1219, Feb. 2018.
- [24] Y.-P. Wei, K. Banawan, and S. Ulukus, “The capacity of private information retrieval with partially known private side information,” *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8222–8231, Dec. 2019.
- [25] M. A. Attia, D. Kumar, and R. Tandon, “The capacity of private information retrieval from uncoded storage constrained databases,” *IEEE Trans. Inf. Theory*, vol. 66, no. 11, pp. 6617–6634, Nov. 2020.
- [26] C. Tian, H. Sun, and J. Chen, “Capacity-achieving private information retrieval codes with optimal message size and upload cost,” *IEEE Trans. Inf. Theory*, vol. 65, no. 11, pp. 7613–7627, Nov. 2019.
- [27] C. Tian, “On the storage cost of private information retrieval,” *IEEE Trans. Inf. Theory*, vol. 66, no. 12, pp. 7539–7549, Dec. 2020.
- [28] S. Li and M. Gastpar, “Single-server multi-user private information retrieval with side information,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1954–1958.
- [29] S. Li and M. Gastpar, “Single-server multi-message private information retrieval with side information: The general cases,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 1083–1088.
- [30] S. Li and M. Gastpar, “Converse for multi-server single-message PIR with side information,” in *Proc. 54th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2020, pp. 1–6.
- [31] R. Zhou, C. Tian, H. Sun, and T. Liu, “Capacity-achieving private information retrieval codes from MDS-coded databases with minimum message size,” *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 4904–4916, Aug. 2020.
- [32] Q. Wang, H. Sun, and M. Skoglund, “The capacity of private information retrieval with eavesdroppers,” *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 3198–3214, May 2019.
- [33] F. Ye, C. Naim, and S. E. Rouayheb, “ON-OFF privacy in the presence of correlation,” *IEEE Trans. Inf. Theory*, vol. 67, no. 11, pp. 7438–7457, Nov. 2021.
- [34] F. Ye, C. Naim, and S. E. Rouayheb, “ON-OFF privacy against correlation over time,” *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 2104–2117, 2021.
- [35] L. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

Fangwei Ye (Member, IEEE) received the B.Eng. degree in information engineering from Southeast University in 2013 and the Ph.D. degree from the Department of Information Engineering, The Chinese University of Hong Kong, in 2018. From 2018 to 2020, he was a Post-Doctoral Associate with the Department of Electrical and Computer Engineering, Rutgers University. He is currently with the Broad Institute of MIT and Harvard. His research interests include information theory and its applications to privacy, bioinformatics, and coding opportunities in learning.

Salim El Rouayheb (Senior Member, IEEE) received the Diploma degree in electrical engineering from the Faculty of Engineering, Lebanese University, Roumieh, Lebanon, in 2002, the M.S. degree from the American University of Beirut, Lebanon, in 2004, and the Ph.D. degree in electrical engineering from Texas A&M University, College Station, in 2009. He is currently an Associate Professor with the ECE Department, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA. He was a Postdoctoral Research Fellow with UC Berkeley from 2010 to 2011 and a Research Scholar with Princeton University from 2012 to 2013. He was an Assistant Professor with the ECE Department, Illinois Institute of Technology, from 2013 to 2017. His research interests include information theory and coding theory with applications to reliability, security, and privacy in distributed systems. He was a recipient of NSF Career Award.