Private Information Retrieval from Multiple Access Channels

Karim Banawan Sennur Ulukus
Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742
kbanawan@umd.edu *ulukus@umd.edu*

***Abstract*—We consider the private information retrieval problem from multiple access channels (MAC-PIR). In MAC-PIR, there are N databases, each storing the same set of M messages. The database responses reach the user through a multiple access channel (MAC) that may mix the responses together in a stochastic way. We show that for the additive MAC and the conjunction/disjunction MAC, channel coding and retrieval scheme are *inseparable* unlike the noisy private information retrieval problem (NPIR). We show that the retrieval scheme depends on the properties of the MAC, in particular on the linearity aspect. For both cases, we provide schemes that achieve the full capacity without any loss due to the privacy constraint, which implies that the user can exploit the nature of the channel in its favor. Finally, we show that the full capacity is not always attainable by determining the capacity of the selection channel.**

I. Introduction

Motivated by the need for absolute privacy guarantees against data-mining techniques, private information retrieval (PIR) is considered an important research thrust for future networks. The PIR problem is introduced by Chor et al. [1] to study the privacy of the downloaded content from public databases. In classical PIR, a user wishes to retrieve a file privately from N non-colluding databases each storing the same set of M messages (files). To that end, the user submits queries for the databases that do not reveal the user's interest in the desired file. The databases respond with *correct* answer strings via *noiseless orthogonal links*, from which the user reconstructs the desired file. PIR schemes are designed to maximize the retrieval rate, which is the ratio of the number of downloaded bits from the desired message to the total number of downloaded bits. Recently, there has been a growing interest within the information theory society in PIR [2]–[6]. Sun and Jafar have characterized the capacity of the classical PIR problem [7], where the capacity is defined as the supremum over all possible retrieval rates. Following [7], many interesting variants of the classical PIR problem have been considered, such as [8]–[34].

In all previous works, the links from the databases to the user are *noiseless* and the answer strings are returned via *orthogonal links*. Yet, in some applications, the answer strings may be mixed before reaching the user. For example: if the user is retrieving the desired file from wireless base stations, the answer strings would be combined in the air before reaching the user. Another example is retrieval from a cloud, where the returned packets may collide and superimpose each other. These practical settings can be abstracted by a cooperative multiple access channel (MAC) model, where the databases are cooperating to convey the desired message to the user, while the user receives a stochastic mapping from the database responses in general. This motivates the problem of PIR from MAC, and poses many interesting questions, such as: How to devise schemes that mitigate the errors introduced by the MAC with a small sacrifice from the retrieval rate? Is there a separation between the channel coding needed for reliable transmission over MACs and the private retrieval scheme, or if there is a necessity for joint processing? How do the statistical properties of MACs fundamentally affect the retrieval rate?

In this paper, we introduce the PIR problem from multiple access channel (MAC-PIR). In the MAC-PIR problem, the responses of the databases reach the user through a discrete memoryless MAC with a known transition probability $p(y|x_1,\cdots,x_N)$. In this case, the output of the channel is a noisy mixture of all databases' responses. The user needs to decode the desired message with vanishingly small probability of error from the channel output. The most closely related work to the MAC-PIR problem is the noisy PIR (NPIR) with orthogonal links in [34], where the user receives N distinct noisy answer strings from N orthogonal noisy links connected to the databases. Reference [34] shows that channel coding and retrieval are *almost separable*, i.e., each database performs the optimal coding scheme against the channel errors independently taking into consideration adaption of the traffic ratio from the databases according to channel capacities.

Interestingly, in this paper, we show that this conclusion is false in general for MAC-PIR, i.e., we show that channel coding and retrieval strategy are *inseparable* unlike in NPIR. We show this fact by deriving the PIR capacity of two simple MACs, namely: additive MAC, and logical conjunction/disjunction MAC. In these two cases, we show that *privacy for free* can be attained by designing retrieval strategies that exploit the properties of the channel to maximize the retrieval rate. Interestingly, we show that for the additive MAC, the optimal PIR scheme is linear, while we show a non-linear PIR scheme for the logical conjunction/disjunction MAC, that requires $N \geq 2^{M-1}$ to achieve $C_{PIR} = 1$. We conclude by showing that full capacity may not be attainable for all

This work was supported by NSF Grants CNS 13-14733, CCF 14-22111, CNS 15-26608 and CCF 17-13977.

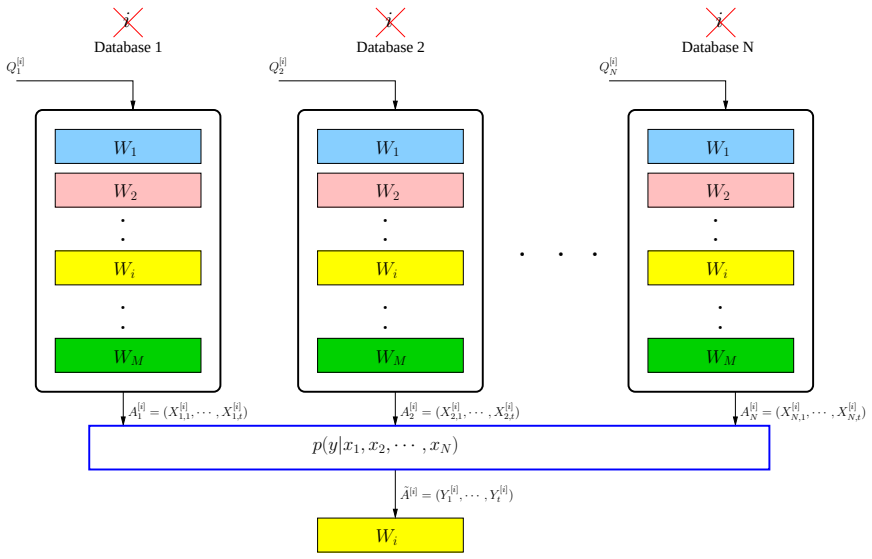

Fig. 1. The MAC-PIR problem.

MACs by giving a counterexample, which is the selection channel. The capacity of MAC-PIR for an arbitrary transition probability distribution remains an open problem.

II. System Model

Consider a classical PIR model with N replicated and non-colluding databases storing M messages. Each database stores the same set of messages $W_{1:M} = \{W_1, \cdots, W_M\}$. The mth message W_m is an L-length vector picked uniformly from $\mathbb{F}_2^L$. The messages $W_{1:M}$ are i.i.d., i.e., for $m \in \{1, \cdots, M\}$,

$$H(W_m) = L, \quad H(W_{1:M}) = ML \tag{1}$$

In PIR, in order to retrieve W_i, the user submits N queries $Q_{1:N}^{[i]} = \{Q_1^{[i]}, \cdots, Q_N^{[i]}\}$, one for each database. The queries and the messages are statistically independent because the user has no knowledge about $W_{1:M}$,

$$I(W_{1:M}; Q_{1:N}^{[i]}) = 0, \quad i \in \{1, \cdots, M\} \tag{2}$$

The nth database responds with an answer string $A_n^{[i]} = (X_{n,1}^{[i]}, \cdots, X_{n,t}^{[i]})$. The nth answer string is a deterministic function of $(W_{1:M}, Q_n^{[i]})$, i.e., for all n and i,

$$H(A_n^{[i]} | W_{1:M}, Q_n^{[i]}) = 0 \tag{3}$$

The user receives a noisy observation $\tilde{A}^{[i]} = (Y_1^{[i]}, \cdots, Y_t^{[i]})$, where the responses of the databases $(A_1^{[i]}, A_2^{[i]}, \cdots, A_N^{[i]})$ pass through a discrete memoryless channel with a transition probability distribution $p(y|x_1, \cdots, x_N)$, i.e.,

$$P\left(\tilde{A}^{[i]} | A_1^{[i]}, \cdots, A_N^{[i]}\right) = \prod_{\eta=1}^{t} p\left(y_\eta^{[i]} | x_{1,\eta}^{[i]}, \cdots, x_{N,\eta}^{[i]}\right) \tag{4}$$

In this sense, the retrieval is performed via a *cooperative multiple access channel*, as the databases cooperate to convey the message W_i to a common receiver (the user). The full cooperation is realized via the user queries. Furthermore, in MAC-PIR, the database responses are mixed together to have the noisy observation $\tilde{A}^{[i]}$ in contrast to NPIR in [34].

To ensure the privacy, $Q_n^{[i]}$ should not reveal any information about i. We can write the privacy constraint as,

$$(Q_n^{[i]}, A_n^{[i]}, W_{1:M}) \sim (Q_n^{[j]}, A_n^{[j]}, W_{1:M}), \; i, j \in \{1, \cdots, M\} \tag{5}$$

In addition, for the MAC-PIR problem, the user should reliably reconstruct W_i with vanishingly small probability of error by observing the noisy and mixed output $\tilde{A}^{[i]}$,

$$H(W_i | Q_{1:N}^{[i]}, \tilde{A}^{[i]}) \leq o(L) \tag{6}$$

where $\frac{o(L)}{L} \to 0$ as $L \to \infty$.

The retrieval rate R is achievable if there exists a sequence of retrieval schemes, indexed by L, that satisfy (5), (6) with a noisy (mixed) answer string of length t, thus, $R = \lim_{L\to\infty} \frac{L}{t}$. The PIR capacity $C_{\text{PIR}} = \sup \; R$ over all retrieval schemes.

III. Related Work

In this section, we review some capacity results for NPIR [34] for comparison with the results of MAC-PIR. In NPIR, the user receives N orthogonal noisy observations which are not mixed as the response from the nth database passes through an orthogonal noisy channel with transition probability $p(y_n|x_n)$. The capacity result for $M = 2, 3$ messages is summarized in the following theorem.

Theorem 1 (Capacity for $M = 2, 3$ messages [34]) *For noisy PIR with orthogonal links with capacities $\mathbf{C} = (C_1, \cdots, C_N)$, the capacity $C_{PIR}(\mathbf{C})$ for $M = 2$ is:*

$$C_{PIR}(\mathbf{C}) = \max_{n_i \in [N]} \frac{n_0 n_1}{\sum_{n=1}^{n_0} \frac{n_0+1}{C_n} + \sum_{n=n_0+1}^{n_1} \frac{n_0}{C_n}} \tag{7}$$

and for $M = 3$, $C_{PIR}(\mathbf{C})$ is:

$$\max_{n_i \in [N]} \frac{n_0 n_1 n_2}{\sum_{n=1}^{n_0} \frac{n_0 n_1 + n_0 + 1}{C_n} + \sum_{n=n_0+1}^{n_1} \frac{n_0 n_1 + n_0}{C_n} + \sum_{n=n_1+1}^{n_2} \frac{n_0 n_1}{C_n}} \tag{8}$$

Theorem 1 implies that the PIR capacity does not depend explicitly on the transition probability $p(y_n|x_n)$ but rather on the link capacity C_n. Furthermore, the result implies *almost separation* between the channel coding and the retrieval scheme as each database performs the channel coding needed for combating errors independently. The problem is coupled through adapting the traffic ratio from each database. The result implies that the noisy channel cannot enhance the retrieval rate, as $C_{\text{PIR}}(\mathbf{C})$ is maximized by communicating through noiseless channels ($C_n = 1$). In the sequel, we show that these implications are false for MAC-PIR by deriving the capacities of two special cases of MAC-PIR, namely, the additive MAC and the conjunction/disjunction MAC.

IV. Capacity of PIR from Additive MAC

In the first special case, we consider the additive MAC. At each time instant η, the responses of the databases are added together (modulo-2 addition) in addition to a random variable $Z_\eta \sim$ Bernoulli(p), which is independent of $(W_{1:M}, Q^{[i]}_{1:N})$ and corresponds to a random additive noise, i.e.,

$$Y_\eta = \sum_{n=1}^{N} X_{n,\eta} + Z_\eta \tag{9}$$

Theorem 2 derives the PIR capacity of the additive MAC.

Theorem 2 *For the MAC-PIR problem from discrete memoryless additive channel, the PIR capacity is given by:*

$$C_{PIR} = 1 - H(p) \tag{10}$$

where p is the flipping probability of the additive channel.

Remark 1 *For additive noiseless channel, i.e., $p = 0$ and $Y_\eta = \sum_{n=1}^{N} X_{n,\eta}$, the PIR capacity $C_{PIR} = 1$. This implies that there is no penalty due to the privacy constraint, i.e., the user can have privacy for free. Interestingly, this is the first instance where the PIR capacity is independent of the number of databases N and the number of messages M.*

Remark 2 *For additive noiseless channel, i.e., $p = 0$, separation between channel coding and retrieval process is not optimal unlike NPIR. In fact, the retrieval scheme is dependent on the structure of the channel. To see this, the user generates a random binary vector $\mathbf{h} = [h_1\ h_2\ \cdots\ h_M] \in \{0,1\}^M$. The user sends $\mathbf{h}$ to database 1, flips the ith position of $\mathbf{h}$ and sends it to database 2, and does not send anything to the remaining databases. Thus, the responses of the databases are,*

$$A_1^{[i]} = \sum_{m=1}^{M} h_m W_m, \qquad A_2^{[i]} = \sum_{m=1}^{M} h_m W_m + W_i \tag{11}$$

This is exactly the scheme in [1]. Since the channel is additive and noiseless, $\tilde{A}^{[i]} = A_1^{[i]} + A_2^{[i]} = W_i$. Hence, $R = 1$. Here, we note that, the channel performs the processing at the user for free. This implies that by careful design of queries, the user can exploit the channel to maximize the retrieval rate.

Proof: We prove the converse and achievability.

a) The converse proof: We assume that W_1 is the desired message without loss of generality. Then, we have,

$$\begin{aligned}
L &= H(W_1) && (12)\\
&\overset{(1),(2)}{=} H(W_1|W_{2:M}, Q^{[1]}_{1:N}) && (13)\\
&\overset{(6)}{leq} H(W_1|W_{2:M}, Q^{[1]}_{1:N}) - H(W_1|W_{2:M}, Q^{[1]}_{1:N}, \tilde{A}^{[1]}) \\
&\quad + o(L) && (14)\\
&= I(W_1; \tilde{A}^{[1]}|Q^{[1]}_{1:N}, W_{2:M}) + o(L) && (15)\\
&= H(\tilde{A}^{[1]}|Q^{[1]}_{1:N}, W_{2:M}) - H(\tilde{A}^{[1]}|Q^{[1]}_{1:N}, W_{1:M}) + o(L) && (16)\\
&\overset{(3)}{\leq} H(\tilde{A}^{[1]}) - H(\tilde{A}^{[1]}|Q^{[1]}_{1:N}, W_{1:M}, A^{[1]}_{1:N}) + o(L) && (17)\\
&= t - H(\tilde{A}^{[1]}|A^{[1]}_{1:N}) + o(L) && (18)\\
&= t - \sum_{\eta=1}^{t} H(Y^{[1]}_\eta|X^{[1]}_{1,\eta}, X^{[1]}_{2,\eta}, \cdots, X^{[1]}_{N,\eta}) + o(L) && (19)\\
&= t - \sum_{\eta=1}^{t} H(Z_\eta|X^{[1]}_{1,\eta}, X^{[1]}_{2,\eta}, \cdots, X^{[1]}_{N,\eta}) + o(L) && (20)\\
&= t(1 - H(p)) + o(L) && (21)
\end{aligned}$$

where (13) follows from the independence of the messages and the queries, (14) follows from the reliability constraint, (17) follows from the fact that the answer string $A^{[1]}_n$ is a deterministic function of the messages and the queries, (18) follows from the fact that $(W_{1:M}, Q^{[1]}_{1:N}) \to A^{[1]}_{1:N} \to \tilde{A}^{[1]}$ is a Markov chain, (19) follows from the fact that the channel is memoryless, and (21) follows from the independence of Z_η and $(X^{[1]}_{1,\eta}, X^{[1]}_{2,\eta}, \cdots, X^{[1]}_{N,\eta})$ as a consequence of the independence of $(Z_\eta, W_{1:M}, Q^{[1]}_{1:N})$.

Hence, by reordering terms and taking $L \to \infty$, we have $R = \frac{L}{t} \leq 1 - H(p)$. Note that we can interpret the upper bound as the cooperative MAC bound, i.e., $R \leq I(Y; X_1, X_2, \cdots, X_N) = 1 - H(p)$.

b) The achievability proof: To show the general achievability, the user submits queries to database 1 and database 2 only and ignores the remaining databases. We note that the additive channel in this case boils down to $Y_\eta = X_{1,\eta} + X_{2,\eta} + Z_\eta$, which means that the channel $p(y|x_1, x_2)$ is BSC(p).

To that end, let the mth message be a vector $W_m = [W_m(1)\quad W_m(2)\quad \cdots\quad W_m(L)]$ of length L. The user repeats the following scheme L times. For the jth repetition of the scheme, the user generates a random binary vector $\mathbf{h}(j) = [h_1(j)\quad h_2(j)\quad \cdots\quad h_M(j)] \in \{0,1\}^M$. The user sends the following queries to the databases:

$$Q_1^{[i]}(j) = \mathbf{h}(j), \qquad Q_2^{[i]}(j) = \mathbf{h}(j) + \mathbf{e}_i \tag{22}$$

where $\mathbf{e}_i$ is the unit vector containing 1 only at the ith position. The queries are private since $Q_n^{[i]}$ is a vector picked uniformly from $\{0,1\}^M$ for any message i.

For the jth repetition of the scheme, the database uses the received query vector as a combining vector for the jth elements of all messages. The nth database concatenates all

responses in a vector $U_n^{[i]}$ of length L, hence

$$U_1^{[i]} = \left[\sum_{m=1}^{M} h_m(1) W_m(1) \quad \sum_{m=1}^{M} h_m(2) W_m(2) \quad \cdots \quad \sum_{m=1}^{M} h_m(L) W_m(L) \right] \tag{23}$$

$$U_2^{[i]} = \left[\sum_{m=1}^{M} h_m(1) W_m(1) + W_i(1) \quad \sum_{m=1}^{M} h_m(2) W_m(2) + W_i(2) \quad \cdots \quad \sum_{m=1}^{M} h_m(L) W_m(L) + W_i(L) \right] \tag{24}$$

Using Shannon's theorem for BSC(p) [35, Theorem 4.17, Corollary 4.18], for $p \in (0, 0.5)$, all but ρ linear $[t, L]$ block codes $\mathcal{C}$, where $\frac{L}{t} = r < 1 - H(p)$, have $P_e(\mathcal{C}) < \frac{2}{\rho} \cdot 2^{-t\Delta(p,r)}$ for $\Delta(p, r) > 0$. Then, the databases agree on the same $[t, L]$ code from the family of the good codes, where $t = \frac{L}{[1-H(p)]}$. The nth database encodes $U_n^{[i]}$ independently by the same $[t, L]$ linear block code to output $A_n^{[i]}$.

The noisy observation at the user is given by:

$$\tilde{A}^{[i]} = A_1^{[i]} + A_2^{[i]} + Z_{1:t} = \hat{A}^{[i]} + Z_{1:t} \tag{25}$$

Since the two databases employ the same linear block code, the sum of the two codewords $\hat{A}^{[i]} = A_1^{[i]} + A_2^{[i]}$ is also a valid codeword corresponding to the sum $U_1^{[i]} + U_2^{[i]}$.

Consequently, as $L \to \infty$, $t \to \infty$, the probability of error in decoding $U_1^{[i]} + U_2^{[i]}$ is $P_e(L) \to 0$. By observing that $U_1^{[i]} + U_2^{[i]} = W_i$, the reliability proof follows. ■

Remark 3 *The PIR scheme relies on the additivity of the channel, as the scheme uses a linear block code to exploit the fact that the sum of two codewords from a linear block code is a valid codeword. Thus, the retrieval process depends on the channel transition probability explicitly unlike NPIR.*

V. Capacity of PIR from Conjunction MAC

We show that we can achieve privacy for free for channels other than the additive channels. We illustrate by considering the MAC-PIR problem through channels that output the logical conjunctions (AND)/disjunctions (OR) of the inputs. Let $\wedge$, $\vee$, $\neg$ denote the conjunction, disjunction, negation operators, respectively. The input-output relation of the discrete memoryless logical conjunction channel is given as:

$$Y_\eta = \bigwedge_{n=1}^{N} X_{n,\eta} \tag{26}$$

For the conjunction channel, we have the following result.

Theorem 3 *For the MAC-PIR problem from discrete memoryless logical conjunction channel, if $N \geq 2^{M-1}$, then the PIR capacity is $C_{PIR} = 1$.*

Remark 4 *Similar to the additive channel, there is no loss due to the presence of a privacy constraint for the conjunction channel. In this case, the capacity depends on the number of messages M, and the number of databases N (which is related to M as $N = 2^{M-1}$) unlike the additive channel.*

Proof: It suffices to show only the achievability as the retrieval rate is trivially upper bounded by 1. The user submits queries to 2^{M-1} databases only. The user generates the random variables $(Z_1, \cdots, Z_M)$ independently, privately, and uniformly from $\{0, 1\}$. $Z_m \sim$ Bernoulli($\frac{1}{2}$) represents the negation state of the mth message literal in the first query $Q_1^{[i]}$. Let $\tilde{W}_m$ be the requested literal from the mth message in $Q_1^{[i]}$, hence,

$$\tilde{W}_m = \begin{cases} W_m, & Z_m = 1 \\ \neg W_m, & Z_m = 0 \end{cases} \tag{27}$$

Assume that W_1 is the desired message. From database 1, the user requests the disjunction $X_1 = \bigvee_{m=1}^{M} \tilde{W}_m$. From other databases, the user requests the same literal $\tilde{W}_1$ with a disjunction of the remaining messages with different negation pattern than the first query. Denote the disjunction of messages $W_{2:M}$ from the nth database by F_n, hence,

$$Y = \left(\bigvee_{m=1}^{M} \tilde{W}_m \right) \wedge \left(\tilde{W}_1 \vee \neg \tilde{W}_2 \vee \bigvee_{m \in [M] \setminus \{1,2\}} \tilde{W}_m \right) \wedge \left(\tilde{W}_1 \vee \neg \tilde{W}_3 \vee \bigvee_{m \in [M] \setminus \{1,3\}} \tilde{W}_m \right) \wedge \cdots \tag{28}$$

$$= \tilde{W}_1 \vee \bigwedge_{i=1}^{2^{M-1}} F_i = \tilde{W}_1 \tag{29}$$

where (29) follows from successively applying the Boolean relation $(\tilde{W}_1 \vee G_1) \wedge (\tilde{W}_1 \vee G_2) = \tilde{W}_1 \vee (G_1 \wedge G_2)$ for logical expressions G_1, G_2 and the fact that there exist 2^{M-1} different negation states for the literals from $W_{2:M}$. Each negation state is requested from one database in the form of logical expression F_i, hence the conjunction of all these logical expressions $\bigwedge_{i=1}^{2^{M-1}} F_i = 0$. This satisfies the reliability constraint. Intuitively, the queries cover *exactly half* the Karnaugh map, which is reduced to either W_1 or $\neg W_1$.

Since the negation state for every message is chosen uniformly, independently, and privately, the probability of receiving specific query from the user is $\frac{1}{2^M}$ irrespective to the desired message, which guarantees the privacy. ■

Remark 5 *As an explicit example, let $M = 3$, $N = 2^{M-1} = 4$, then the user requests the following:*

$$X_1 = \tilde{W}_1 \vee \tilde{W}_2 \vee \tilde{W}_3, \quad X_2 = \tilde{W}_1 \vee \neg \tilde{W}_2 \vee \tilde{W}_3 \tag{30}$$

$$X_3 = \tilde{W}_1 \vee \tilde{W}_2 \vee \neg \tilde{W}_3, \quad X_4 = \tilde{W}_1 \vee \neg \tilde{W}_2 \vee \neg \tilde{W}_3 \tag{31}$$

Hence, the output of the channel can be written as:

$$Y = X_1 \wedge X_2 \wedge X_3 \wedge X_4 \tag{32}$$

$$= (\tilde{W}_1 \vee \tilde{W}_3) \wedge (\tilde{W}_1 \vee \neg \tilde{W}_3) = \tilde{W}_1 \tag{33}$$

W_1 is decodable as the user knows the negation of $\tilde{W}_1$.

Remark 6 *The achievable scheme is a non-linear retrieval scheme that depends on the non-linear characteristics of the channel in contrast to the linear retrieval scheme used for the additive channel. This confirms the non-separability between our retrieval scheme and the channel coding.*

Remark 7 *We note that the result is still true if the channel is replaced by a disjunction channel, i.e.,* $Y_\eta = \bigvee_{n=1}^{N} X_{n,\eta}$*. The proof follows by replacing every disjunction operator with a conjunction operator in the submitted queries.*

VI. Capacity of PIR from Selection MAC

We illustrate that the *privacy for free* phenomenon may not be always feasible for any arbitrary channel in the MAC-PIR problem. To illustrate, we consider the selection channel. In this channel, the user selects to connect to one database only at random and sticks to it throughout the transmission, i.e.,

$$Y_\eta = X_{n,\eta}, \quad n \sim \text{uniform } \{1, \cdots, N\} \tag{34}$$

This implies that the user faces a single-database PIR problem at every channel use. In this case, the capacity is attained by downloading all the messages (M messages) from the connected database. Thus, the PIR capacity is given by $C_{\text{PIR}} = \frac{1}{M}$.

On the other hand if the user selects to connect to one database at random at every channel use, i.e.,

$$Y_\eta = X_{n(\eta),\eta}, \quad n(\eta) \sim \text{uniform } \{1, \cdots, N\} \tag{35}$$

where $n(\eta)$ is the database index at channel use η, then, $C_{PIR} \leq C = (1+\frac{1}{N}+\cdots+\frac{1}{N^{M-1}})^{-1}$ trivially as the user can choose to ignore all the responses except one in classical PIR [7], and hence C is a valid upper bound. For the achievability, the user can repeat the scheme in [7] ν times, which results in using the selection channel $t = \nu\frac{L}{C} = \nu\frac{N(N^M-1)}{N-1}$. At channel use η, the user chooses a new query element from $Q_{n(\eta)}^{[i]}$ and submits it to database $n(\eta)$. As $\nu \to \infty$, by strong law of large numbers, each database will be visited t_n times, where $t_n \to \frac{t}{N}$ for every n. Hence, all bits are decodable by the decodability of the scheme in [7] and $C_{PIR} = C = (1 + \frac{1}{N} + \cdots + \frac{1}{N^{M-1}})^{-1} < 1$ as well.

References

[1] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. Private information retrieval. *Journal of the ACM*, 45(6):965–981, 1998.

[2] N. B. Shah, K. V. Rashmi, and K. Ramchandran. One extra bit of download ensures perfectly private information retrieval. In *IEEE ISIT*, June 2014.

[3] G. Fanti and K. Ramchandran. Efficient private information retrieval over unsynchronized databases. *IEEE Journal of Selected Topics in Signal Processing*, 9(7):1229–1239, October 2015.

[4] T. Chan, S. Ho, and H. Yamamoto. Private information retrieval for coded storage. In *IEEE ISIT*, June 2015.

[5] A. Fazeli, A. Vardy, and E. Yaakobi. Codes for distributed PIR with low storage overhead. In *IEEE ISIT*, June 2015.

[6] R. Tajeddine and S. El Rouayheb. Private information retrieval from MDS coded data in distributed storage systems. In *IEEE ISIT*, July 2016.

[7] H. Sun and S. A. Jafar. The capacity of private information retrieval. *IEEE Trans. on Info. Theory*, 63(7):4075–4088, July 2017.

[8] H. Sun and S. A. Jafar. The capacity of robust private information retrieval with colluding databases. *IEEE Trans. on Info. Theory*, 64(4):2361–2370, April 2018.

[9] H. Sun and S. Jafar. The capacity of symmetric private information retrieval. 2016. Available at arXiv:1606.08828.

[10] K. Banawan and S. Ulukus. The capacity of private information retrieval from coded databases. *IEEE Trans. on Info. Theory*, 64(3):1945–1956, March 2018.

[11] H. Sun and S. A. Jafar. Optimal download cost of private information retrieval for arbitrary message length. *IEEE Trans. on Info. Forensics and Security*, 12(12):2920–2932, Dec 2017.

[12] Q. Wang and M. Skoglund. Symmetric private information retrieval for MDS coded distributed storage. 2016. Available at arXiv:1610.04530.

[13] H. Sun and S. A. Jafar. Multiround private information retrieval: Capacity and storage overhead. *IEEE Trans. on Info. Theory*, 64(8):5743–5754, August 2018.

[14] R. Freij-Hollanti, O. Gnilke, C. Hollanti, and D. Karpuk. Private information retrieval from coded databases with colluding servers. *SIAM Journal on Applied Algebra and Geometry*, 1(1):647–664, 2017.

[15] H. Sun and S. A. Jafar. Private information retrieval from mds coded data with colluding servers: Settling a conjecture by Freij-Hollanti et al. *IEEE Trans. on Info. Theory*, 64(2):1000–1022, Feb. 2018.

[16] R. Tajeddine, O. W. Gnilke, D. Karpuk, R. Freij-Hollanti, C. Hollanti, and S. El Rouayheb. Private information retrieval schemes for coded data with arbitrary collusion patterns. 2017. Available at arXiv:1701.07636.

[17] K. Banawan and S. Ulukus. Multi-message private information retrieval: Capacity results and near-optimal schemes. *IEEE Trans. on Info. Theory*. To appear. Also available at arXiv:1702.01739.

[18] Y. Zhang and G. Ge. A general private information retrieval scheme for MDS coded databases with colluding servers. 2017. Available at arXiv: 1704.06785.

[19] Y. Zhang and G. Ge. Multi-file private information retrieval from MDS coded databases with colluding servers. 2017. Available at arXiv: 1705.03186.

[20] K. Banawan and S. Ulukus. The capacity of private information retrieval from Byzantine and colluding databases. *IEEE Trans. on Info. Theory*. To appear. Also available at arXiv:1706.01442.

[21] Q. Wang and M. Skoglund. Secure symmetric private information retrieval from colluding databases with adversaries. 2017. Available at arXiv:1707.02152.

[22] R. Tandon. The capacity of cache aided private information retrieval. 2017. Available at arXiv: 1706.07035.

[23] Q. Wang and M. Skoglund. Linear symmetric private information retrieval for MDS coded distributed storage with colluding servers. 2017. Available at arXiv:1708.05673.

[24] S. Kadhe, B. Garcia, A. Heidarzadeh, S. El Rouayheb, and A. Sprintson. Private information retrieval with side information. 2017. Available at arXiv:1709.00112.

[25] Y.-P. Wei, K. Banawan, and S. Ulukus. Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching. 2017. Available at arXiv:1709.01056.

[26] Z. Chen, Z. Wang, and S. Jafar. The capacity of private information retrieval with private side information. 2017. Available at arXiv:1709.03022.

[27] Y.-P. Wei, K. Banawan, and S. Ulukus. The capacity of private information retrieval with partially known private side information. 2017. Available at arXiv:1710.00809.

[28] H. Sun and S. A. Jafar. The capacity of private computation. 2017. Available at arXiv:1710.11098.

[29] M. Mirmohseni and M. A. Maddah-Ali. Private function retrieval. 2017. Available at arXiv:1711.04677.

[30] M. Abdul-Wahid, F. Almoualem, D. Kumar, and R. Tandon. Private information retrieval from storage constrained databases–coded caching meets PIR. 2017. Available at arXiv:1711.05244.

[31] Y.-P. Wei, K. Banawan, and S. Ulukus. Cache-aided private information retrieval with partially known uncoded prefetching: Fundamental limits. *Jour. on Selected Areas in Communications*, 2017. To appear.

[32] K. Banawan and S. Ulukus. Asymmetry hurts: Private information retrieval under asymmetric-traffic constraints. *IEEE Trans. on Info. Theory*. Submitted January 2018. Also available at arXiv:1801.03079.

[33] K. Banawan and S. Ulukus. Private information retrieval through wiretap channel II: Privacy meets security. *IEEE Trans. on Info. Theory*. Submitted January 2018. Also available at arXiv:1801.06171.

[34] K. Banawan and S. Ulukus. Noisy private information retrieval. In *IEEE Asilomar 2018*, October 2018.

[35] R. Roth. *Introduction to Coding Theory*. Cambridge University Press, 2006.