Private Information Retrieval Through Wiretap Channel II

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 problem of private information retrieval through a wiretap channel II (PIR-WTC-II). In PIR-WTC-II, a user wants to retrieve a message (or file) privately out of M messages, which are stored in N replicated and non-communicating databases. An eavesdropper observes a fraction μ_n of the traffic exchanged between the nth database and the user. The databases should encode the returned answer strings such that the eavesdropper learns nothing about the *contents* of the databases. We aim at characterizing the capacity of the PIR-WTC-II under these joint privacy and security constraints. We obtain an upper bound in the form of a max-min optimization problem. We propose an achievability scheme that satisfies the security constraint by encoding a secret key into an artificial noise vector using an MDS code. The user and the databases operate at one of the corner points of the achievable scheme of the PIR under asymmetric traffic constraints such that the retrieval rate is maximized under the imposed security constraint. The upper bound and the lower bound match for the cases of $M = 2$ and $M = 3$ messages, for any number of databases N, and any μ_n.**

I. Introduction

Private information retrieval (PIR) [1] is a canonical problem, which considers the privacy of the content downloaded from public databases. The classical PIR model includes N non-colluding databases storing the same set of M messages and a user who privately requests a file from these databases, i.e., without revealing the user's interest in a specific file. The user submits a query to each database and receives an answering string. From all answering strings, the user should be able to decode the desired file. The retrieval rate is the ratio between the number of desired message symbols and the total number of the downloaded symbols. There has been a growing interest in the PIR problem in the information theory society [2]–[7]. In [8], Sun and Jafar characterize the capacity of the classical PIR problem, which is defined as the supremum of all PIR rates over all achievable retrieval schemes. Following [8], the capacities of many interesting variants of the classical PIR problem have been considered, such as [9]–[32].

The sole requirement of these previous works is to protect the identity of the desired message from the public databases. Another interesting dimension to the PIR problem is when the content of the requested message needs to be protected against an external eavesdropper, who wishes to learn about the contents of the databases by observing the queries and answer strings. In this paper, we impose an extra constraint to the PIR problem, namely, the secrecy constraint, which ensures that the queries and the answer strings do not leak any information about the contents of the databases to the eavesdropper. A few works exist in the secure PIR problem: [33] considers the problem of information storage and retrieval, guaranteeing that storing of information is secure. [34] considers a *symmetric* PIR setting where there is a passive eavesdropper who can tap in on the incoming and outgoing transmissions of any E servers. Interestingly, the secret key needed for the *symmetric* PIR process is used as an encryption key to secure the contents of the databases from the eavesdropper. This problem is further investigated for the classical PIR problem in [35], which derives inner and outer bounds for this problem, in addition to the minimum amount of common randomness required.

We investigate PIR through a wiretap channel II (PIR-WTC-II). Ozarow and Wyner [36] introduced wiretap channel II. In PIR-WTC-II, the user observes answer strings through a noiseless channel from the nth database. The eavesdropper can observe a fraction μ_n from the nth answer string by choosing any set of positions $\mathcal{S}_n \subset \{1, \cdots, t_n\}$, such that $|\mathcal{S}_n| = \mu_n t_n$. The databases encode the answer strings such that the eavesdropper learns nothing from its observations. Naturally, the nth database dedicates $\mu_n t_n$ portion of the answer string to confuse the eavesdropper, constraining the *meaningful* portion of the answer to be $(1 - \mu_n)t_n$. This fundamentally relates PIR-WTC-II to the PIR problem under asymmetric traffic constraints [37], as lengths of answer strings can no longer be symmetric.

In this paper, we obtain an upper bound for the PIR-WTC-II problem, which takes the form of a max-min optimization problem. The inner minimization problem derives the tightest upper bound for the retrieval rate for an arbitrary traffic ratio vector $\boldsymbol{\tau}$, while the outer maximization problem optimizes over $\boldsymbol{\tau}$. For the achievability, we design the *meaningful* portion of the queries to operate at one of the corner points of the PIR problem under asymmetric traffic constraints [37]. To satisfy the security constraint, each database generates a secret key with $\mu_n t_n$ length, encodes it into an artificial noise vector using a $(t_n, \mu_n t_n)$ MDS code and encrypts the returned answer string with this artificial noise vector. Interestingly, our achievable rate does not need any shared randomness between the databases or the user. Our upper and lower bounds match for $M = 2$ and $M = 3$. We only provide sketches of the proofs here due to space limitations; proof details, illustrative remarks, extra examples and some figures can be found in the longer version [38].

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

II. System Model

Consider a PIR model, in which there are N non-colluding databases, each storing the same content of M messages (or files). The message W_m is represented as a vector of length L picked from a finite field $\mathbb{F}_q^L$ with a sufficiently large alphabet. The messages $W_{1:M} = \{W_1, \cdots, W_M\}$ are i.i.d., hence

$$H(W_m) = L, \quad H(W_{1:M}) = ML, \quad (q\text{-ary bits}) \tag{1}$$

A user wants to retrieve a message W_i from the N databases without revealing the identity of the message i to any individual database. The user sends the query $Q_n^{[i]}$ to the nth database. Since the user has no knowledge about the realizations of $W_{1:M}$, the queries and the messages are independent, i.e.,

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

where $Q_{1:N}^{[i]} = \{Q_1^{[i]}, \cdots, Q_N^{[i]}\}$. To ensure the privacy of W_i, the user should constrain the query intended to retrieve W_i to be indistinguishable from the query intended to retrieve any other message W_j at any individual database. Thus,

$$(Q_n^{[i]}, A_n^{[i]}, W_{1:M}) \sim (Q_n^{[j]}, A_n^{[j]}, W_{1:M}),\ j \in \{1,\cdot\cdot,M\} \tag{3}$$

where $\sim$ denotes statistical equivalence.

The nth database, after receiving the query $Q_n^{[i]}$, responds with a t_n-length answering string $A_n^{[i]}$. The answer string is generally a *stochastic* mapping of the messages $W_{1:M}$ and the received query $Q_n^{[i]}$, hence

$$H(A_n^{[i]} | Q_n^{[i]}, W_{1:M}, \mathcal{G}_n) = 0, \quad n \in \{1, \cdots, N\} \tag{4}$$

where $\mathcal{G}_n$ is independent of all other random variables, whose realization is known at the nth database only. We denote the traffic ratio vector by $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_N)$. The traffic ratio at the nth database τ_n is given by $\tau_n = \frac{t_n}{\sum_{i=1}^N t_i}$. We assume that the answer strings are transmitted through a WTC-II. In PIR-WTC-II, the user observes the t_n-length answer string $A_n^{[i]}$ from the nth database through a noiseless channel. On the other hand, the eavesdropper can observe a fraction μ_n from the nth answer string. More specifically, the eavesdropper arbitrarily chooses any set of positions $\mathcal{S}_n \subset \{1, \cdots, t_n\}$ to observe from the nth answer string, such that $|\mathcal{S}_n| = \mu_n t_n$, i.e., the output of the eavesdropper channel is given by,

$$Z_n^{[i]} = A_n^{[i]}(\mathcal{S}_n), \quad n \in \{1, \cdots, N\} \tag{5}$$

We denote the unobserved portion of the answer string by $Y_n^{[i]} = A_n^{[i]}(\bar{\mathcal{S}}_n)$, where $\bar{\mathcal{S}}_n = \{1, \cdots, t_n\} \setminus \mathcal{S}_n$, thus $A_n^{[i]} = (Y_n^{[i]}, Z_n^{[i]})$. We write the eavesdropping ratios, which are fixed and given, as a vector $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)$. Without loss of generality, we assume $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_N$.

The databases encode the answer strings such that the eavesdropper learns nothing from the queries and its observations. Consequently, we write the security constraint as,

$$I(W_{1:M}; Z_{1:N}^{[i]}, Q_{1:N}^{[i]}) = 0 \tag{6}$$

The user should be able to reconstruct the desired message W_i from the collected answer strings with arbitrarily small probability of error. We write the reliability constraint as,

$$H(W_i | Q_{1:N}^{[i]}, A_{1:N}^{[i]}) = o(L) \tag{7}$$

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

For a fixed N, M, and eavesdropping ratio vector $\boldsymbol{\mu}$, a retrieval rate $R(\boldsymbol{\mu})$ is achievable if there exists a PIR scheme which satisfies (3), (6), and (7) for some message lengths $L(\boldsymbol{\mu})$ and answer strings of lengths $\{t_n(\boldsymbol{\mu})\}_{n=1}^N$, where,

$$R(\boldsymbol{\mu}) = \frac{L(\boldsymbol{\mu})}{\sum_{n=1}^N t_n(\boldsymbol{\mu})} \tag{8}$$

The message length $L(\boldsymbol{\mu})$ can grow arbitrarily large.

The capacity of PIR-WTC-II, $C(\boldsymbol{\mu})$, is defined as the supremum of all achievable retrieval rates, i.e., $C(\boldsymbol{\mu}) = \sup R(\boldsymbol{\mu})$.

III. Main Results and Discussions

Theorem 1 (Upper bound) *For PIR-WTC-II under eavesdropping capabilities $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)$, the capacity is upper bounded by $\bar{C}(\boldsymbol{\mu})$ which is given by:*

$$\max_{\boldsymbol{\tau} \in \mathbb{T}} \min_{n_i \in \{1, \cdots, N\}} \frac{\phi(0) + \frac{\phi(n_1)}{n_1} + \frac{\phi(n_2)}{n_1 n_2} + \cdots + \frac{\phi(n_{M-1})}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{9}$$

where $\mathbb{T} = \Big\{\boldsymbol{\tau} : \tau_n \geq 0 \quad \forall n \in [1:N], \quad \sum_{n=1}^N \tau_n = 1\Big\}$, and $\phi(\ell) = \sum_{n=\ell+1}^N (1 - \mu_n)\tau_n$ is the sum of the unobserved traffic ratios by the eavesdropper from databases $[\ell + 1 : N]$.

The proof of this upper bound is given in Section IV.

Theorem 2 (Lower bound) *For PIR-WTC-II and monotone non-decreasing sequence $\mathbf{n} = \{n_i\}_{i=0}^{M-1} \subset \{1, \cdots, N\}^M$, let $n_{-1} = 0$, and $\mathcal{S} = \{i \geq 0 : n_i - n_{i-1} > 0\}$. Denote $y_\ell[k]$ as the number of stages of the achievable scheme that downloads k-sums from the nth database in one repetition of the scheme, such that $n_{\ell-1} \leq n \leq n_\ell$, and $\ell \in \mathcal{S}$. Let $\xi_\ell = \prod_{s \in \mathcal{S} \setminus \{\ell\}} \binom{M-2}{s-1}$. The number of stages $y_\ell[k]$ is characterized by the following system of difference equations:*

$$\begin{aligned}
y_0[k] &= (n_0 - 1) y_0[k-1] + \sum_{j \in \mathcal{S} \setminus \{0\}} (n_j - n_{j-1}) y_j[k-1] \\
y_1[k] &= (n_1 - n_0 - 1) y_1[k-1] + \sum_{j \in \mathcal{S} \setminus \{1\}} (n_j - n_{j-1}) y_j[k-1] \\
y_\ell[k] &= n_0 \xi_\ell \delta[k - \ell - 1] + (n_\ell - n_{\ell-1} - 1) y_\ell[k-1] \\
&\quad + \sum_{j \in \mathcal{S} \setminus \{\ell\}} (n_j - n_{j-1}) y_j[k-1], \quad \ell \geq 2
\end{aligned} \tag{10}$$

where $\delta[\cdot]$ denotes the Kronecker delta function. The initial conditions of (10) are $y_0[1] = \prod_{s \in \mathcal{S}} \binom{M-2}{s-1}$, and $y_j[k] = 0$ for $k \leq j$. The achievable rate corresponding to $\mathbf{n}$ is:

$$R(\mathbf{n}, \boldsymbol{\mu}) = \frac{\sum_{\ell \in \mathcal{S}} \sum_{k=1}^M \binom{M-1}{k-1} y_\ell[k] (n_\ell - n_{\ell-1})}{\sum_{\ell \in \mathcal{S}} \sum_{n=n_{\ell-1}+1}^{n_\ell} \frac{\sum_{k=1}^M \binom{M}{k} y_\ell[k]}{1 - \mu_n}} \tag{11}$$

Consequently, the capacity $C(\boldsymbol{\mu})$ is lower bounded by:

$$C(\boldsymbol{\mu}) \geq R(\boldsymbol{\mu}) = \max_{n_0 \leq \cdots \leq n_{M-1} \in \{1, \cdots, N\}} R(\mathbf{n}, \boldsymbol{\mu}) \tag{12}$$

The achievable scheme can be found in Section V.

Corollary 1 (Capacity of $M = 2, 3$ messages) *The capacity of PIR-WTC-II, $C(\boldsymbol{\mu})$, for $M = 2$ and arbitrary N is:*

$$C(\boldsymbol{\mu}) = \max_{n_i \in \{1, \cdots, N\}} \frac{n_0 n_1}{\sum_{n=1}^{n_0} \frac{n_0+1}{1-\mu_n} + \sum_{n=n_0+1}^{n_1} \frac{n_0}{1-\mu_n}} \tag{13}$$

and for $M = 3$ and arbitrary N is:

$$C(\boldsymbol{\mu}) = \max_{n_i \in \{1, \cdots, N\}} \frac{n_0 n_1 n_2}{\sum\limits_{n=1}^{n_0} \frac{n_0 n_1 + n_0 + 1}{1-\mu_n} + \sum\limits_{n=n_0+1}^{n_1} \frac{n_0 n_1 + n_0}{1-\mu_n} + \sum\limits_{n=n_1+1}^{n_2} \frac{n_0 n_1}{1-\mu_n}} \tag{14}$$

The proof of Corollary 1 can be found in [38].

IV. Converse Proof

We derive a general upper bound for PIR-WTC-II. Since the eavesdropper observes a different fraction of the traffic from each database (different μ_n), the answer strings (hence the traffic ratios) from databases are asymmetric in length. Thus, we extend [37] to account for the imposed security constraint. The proofs of the following lemmas can be found in [38].

Lemma 1 (Interference lower bound) *For PIR-WTC-II, the interference from undesired messages within the unobserved portion of the answer strings by the eavesdropper $\sum_{n=1}^{N}(1 - \mu_n)t_n - L$ is lower bounded by,*

$$\sum_{n=1}^{N}(1-\mu_n)t_n - L + o(L) \geq I\left(W_{2:M}; Q_{1:N}^{[1]}, Y_{1:N}^{[1]} | W_1, Z_{1:N}^{[1]}\right) \tag{15}$$

Lemma 2 (Induction lemma) *For all $m \in \{2, \ldots, M\}$ and for an arbitrary $n_{m-1} \in \{1, \cdots, N\}$, the mutual information term in Lemma 1 can be inductively lower bounded as,*

$$\begin{aligned} I&\left(W_{m:M}; Q_{1:N}^{[m-1]}, Y_{1:N}^{[m-1]} | W_{1:m-1}, Z_{1:N}^{[m-1]}\right) \\ &\geq \frac{1}{n_{m-1}} I\left(W_{m+1:M}; Q_{1:N}^{[m]}, Y_{1:N}^{[m]} | W_{1:m}, Z_{1:N}^{[m]}\right) \\ &\quad + \frac{1}{n_{m-1}}\left(L - \sum_{n=n_{m-1}+1}^{N}(1-\mu_n)t_n\right) - \frac{o(L)}{n_{m-1}} \end{aligned} \tag{16}$$

Now, we are ready to prove an explicit upper bound for the retrieval rate in the PIR-WTC-II problem $R(\boldsymbol{\mu})$ by applying Lemma 1 and Lemma 2 successively. For a pre-specified answer string lengths $\{t_n\}_{n=1}^{N}$, and an arbitrary sequence $\{n_i\}_{i=1}^{M-1}$, we can write

$$\sum_{n=1}^{N}(1-\mu_n)t_n - L + \tilde{o}(L)$$

$$\overset{(15)}{\geq} I\left(W_{2:M}; Q_{1:N}^{[1]}, Y_{1:N}^{[1]} | W_1, Z_{1:N}^{[1]}\right) \tag{17}$$

$$\overset{(16)}{\geq} \frac{1}{n_1}\left(L - \sum_{n=n_1+1}^{N}(1-\mu_n)t_n\right) + \frac{1}{n_1 n_2}\left(L - \sum_{n=n_2+1}^{N}(1-\mu_n)t_n\right)$$

$$+ \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}\left(L - \sum_{n=n_{M-1}+1}^{N}(1-\mu_n)t_n\right) \tag{18}$$

where $\tilde{o}(L) = \left(1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}\right) o(L)$, (17) follows from Lemma 1, and the remaining bounding steps follow from successive application of Lemma 2.

Ordering terms and letting $\tau_n = \frac{t_n}{\sum_{i=1}^{N} t_i}$, we have,

$$\begin{aligned} &\left(1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}\right) L \\ &\quad \leq \left(\phi(0) + \frac{\phi(n_1)}{n_1} + \cdots + \frac{\phi(n_{M-1})}{\prod_{i=1}^{M-1} n_i}\right) \sum_{n=1}^{N} t_n + \tilde{o}(L) \end{aligned} \tag{19}$$

We conclude the proof by taking $L \to \infty$. Thus, for an arbitrary sequence $\{n_i\}_{i=1}^{M-1}$, the retrieval rate when the traffic ratio vector is constrained to $\boldsymbol{\tau}$, $R(\boldsymbol{\tau}, \boldsymbol{\mu})$ is upper bounded by,

$$\frac{L}{\sum_{n=1}^{N} t_n} \leq \frac{\phi(0) + \frac{\phi(n_1)}{n_1} + \frac{\phi(n_2)}{n_1 n_2} + \cdots + \frac{\phi(n_{M-1})}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{20}$$

We obtain the tightest upper bound for $R(\boldsymbol{\tau}, \boldsymbol{\mu})$ by minimizing over the sequence $\{n_i\}_{i=1}^{M-1}$ over the set $\{1, \cdots, N\}$ to get

$$R(\boldsymbol{\tau}, \boldsymbol{\mu}) \leq \min_{n_i \in \{1, \cdots, N\}} \frac{\phi(0) + \frac{\phi(n_1)}{n_1} + \cdots + \frac{\phi(n_{M-1})}{\prod_{i=1}^{M-1} n_i}}{1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{\prod_{i=1}^{M-1} n_i}} \tag{21}$$

Since the user and the databases can choose any suitable traffic ratio vector $\boldsymbol{\tau}$ in the set $\mathbb{T}$, by maximizing over $\boldsymbol{\tau}$ in the set $\mathbb{T}$, we obtain the upper bound in (9).

V. Achievable Scheme

We illustrate the main ingredients of the achievable scheme by presenting the following motivating example.

A. Motivating Example: $M = 3$, $N = 2$, $\boldsymbol{\mu} = (\frac{1}{4}, \frac{1}{2})$

1) Explicit Upper Bound $\bar{C}(\boldsymbol{\mu})$: By observing that $\tau_1 = 1 - \tau_2$, the upper bound in Theorem 1 can be explicitly written as the following linear program:

$$\begin{aligned} \max_{\tau_2, R} \quad & R \\ \text{s.t.} \quad & R \leq \frac{1}{3}(1-\mu_1) + \left[(1-\mu_2) - \frac{1}{3}(1-\mu_1)\right]\tau_2 \\ & R \leq \frac{2}{5}(1-\mu_1) + \left[\frac{4}{5}(1-\mu_2) - \frac{2}{5}(1-\mu_1)\right]\tau_2 \\ & R \leq \frac{4}{7}(1-\mu_1) + \left[\frac{4}{7}(1-\mu_2) - \frac{4}{7}(1-\mu_1)\right]\tau_2 \\ & 0 \leq \tau_2 \leq 1 \end{aligned} \tag{22}$$

The optimal solution of (22) is attained at one of the corner points of the feasible region. Thus, the upper bound $\bar{C}(\boldsymbol{\mu})$ is,

$$\max\left\{\frac{1-\mu_1}{3}, \frac{2}{\frac{3}{(1-\mu_1)} + \frac{1}{(1-\mu_2)}}, \frac{4}{\frac{4}{(1-\mu_1)} + \frac{3}{(1-\mu_2)}}\right\} \tag{23}$$

2) Capacity-Achieving Scheme for $\boldsymbol{\mu} = (\frac{1}{4}, \frac{1}{2})$: (See Table I.) The user permutes the indices of the symbols of W_1, W_2, W_3 independently, uniformly, and privately. Assume that W_1 is the desired message. Let a_i, b_i, c_i denote the permuted symbols of W_1, W_2, W_3, respectively. In the case of $\boldsymbol{\mu} = (\frac{1}{4}, \frac{1}{2})$, the upper bound (23) is $\bar{C}(\boldsymbol{\mu}) = \frac{6}{17}$. To achieve this bound, we focus first on the *meaningful* queries, i.e., the queries without the randomness that is added to satisfy the security constraint. From database 1, the user asks for an individual symbol from every message, thus, asks for a_1, b_1, c_1. From database 2, the user does not ask for new individual symbols but rather exploits the side information that is generated from database 1 to query for 2-sums, i.e., the user asks for $a_2 + b_1$, $a_3 + c_1$, $b_2 + c_2$ from database 2. Then, the user exploits b_2+c_2 as side information to ask for $a_4+b_2+c_2$ from database 1. To get an integer number for the meaningful queries, which is $(1-\mu_n)t_n$ symbols from database n, we repeat this scheme ν times. Since this scheme gets 4 symbols from database 1 and 3 symbols from database 2, we have

$$(1-\mu_1)t_1 = 4\nu \;\Rightarrow\; t_1 = \frac{16\nu}{3} \tag{24}$$

$$(1-\mu_2)t_2 = 3\nu \;\Rightarrow\; t_2 = 6\nu \tag{25}$$

Then, the minimal $\nu = 3$. Database 1 generates the independent keys $K_1 = \left(k_1^{(1)}, \cdots, k_4^{(1)}\right)$, such that K_1 is picked uniformly from $\mathbb{F}_q^4$. Database 1 encodes these random keys using a $(16, 4)$ MDS code, to get $u_{[1:16]}$, i.e., $u_{[1:16]} = \mathbf{MDS}_{16\times 4} K_1$. Similarly, database 2 generates $K_2 = \left(k_1^{(2)}, \cdots, k_9^{(2)}\right)$ uniformly from $\mathbb{F}_q^9$. Database 2 encodes the keys using an $(18, 9)$ MDS code, to get $v_{[1:18]}$, hence, $v_{[1:18]} = \mathbf{MDS}_{18\times 9} K_2$. All the meaningful downloads are *encrypted* by the coded keys. Furthermore, the user downloads $u_{[13:16]}$ individually from database 1, and $v_{[10:18]}$ from database 2.

TABLE I
THE QUERY TABLE FOR $M = 3$, $N = 2$, $\mu_1 = \frac{1}{4}$, $\mu_2 = \frac{1}{2}$.

Database 1	Database 2
$a_1 + u_1$ $b_1 + u_2$ $c_1 + u_3$	$a_2 + b_1 + v_1$ $a_3 + c_1 + v_2$ $b_2 + c_2 + v_3$
$a_4 + b_2 + c_2 + u_4$	
$a_5 + u_5$ $b_3 + u_6$ $c_3 + u_7$	$a_6 + b_3 + v_4$ $a_7 + c_3 + v_5$ $b_4 + c_4 + v_6$
$a_8 + b_4 + c_4 + u_8$	
$a_9 + u_9$ $b_5 + u_{10}$ $c_5 + u_{11}$	$a_{10} + b_5 + v_7$ $a_{11} + c_5 + v_8$ $b_6 + c_6 + v_9$
$a_{12} + b_6 + c_6 + u_{12}$	
$u_{13}, u_{14}, u_{15}, u_{16}$	$v_{10}, v_{11}, u_{12}, u_{13}, v_{14}$ $v_{15}, u_{16}, u_{17}, u_{18}$

For the decodability, since database 1 encodes its keys K_1 using a $(16, 4)$ MDS code, by the MDS property, any 4 symbols suffice to reconstruct $u_{[1:16]}$. The user downloads $u_{[13:16]}$ separately, hence $u_{[1:12]}$ can be canceled from the downloads to get the meaningful information only; and similarly for database 2. Furthermore, since the side information at any database is obtained from the undesired symbols downloaded from the other database, all undesired symbols can be canceled and the user is left only with the desired $a_{[1:12]}$.

For the security, the eavesdropper can obtain any 4 symbols from database 1, and any 9 symbols from database 2. Since K_1, K_2 are generated uniformly and independently from $\mathbb{F}_q^4$, $\mathbb{F}_q^9$, respectively, any 4 symbols $(u_{i_1}, \cdots, u_{i_4})$ from $u_{[1:16]}$ are independent and uniformly distributed over $\mathbb{F}_q$, and similarly for any 9 symbols $(v_{j_1}, \cdots, v_{j_9})$ from $v_{[1:18]}$. The leakage at the eavesdropper is upper bounded as,

$$I(W_{1:3}; Z_{1:2}^{[1]}) = H(Z_{1:2}) - H(Z_{1:2}|W_{1:3}) \tag{26}$$

$$\leq \log_q 13 - \log_q 13 = 0 \tag{27}$$

The privacy constraint is satisfied, as the queries and the indices of the message symbols are uniformly and independently permuted. Hence, the user downloads $t_1 = 16$ symbols from database 1, and $t_2 = 18$ symbols from database 2. From these downloads, the user can decode $L = 12$ symbols from W_1. Hence $R = \frac{12}{34} = \frac{6}{17}$, which matches the upper bound.

B. General Achievable Scheme

We present the general achievable scheme for PIR-WTC-II that achieves the retrieval rate in Theorem 2. The core of the achievable scheme is the achievable scheme of the corner points in the PIR problem under asymmetric traffic constraints in [37]. One new ingredient is needed to satisfy the security constraint, namely, encrypting the answer strings by random keys. The nth database uses random key K_n of length $\mu_n t_n$ that is sufficient to span the space of the eavesdropper's observations. The nth database encodes K_n using a $(t_n, \mu_n t_n)$ MDS code and uses the resulting codeword to *encrypt* each downloaded symbol from the meaningful downloads.

We use the terminology as in [37]. Let s_n denotes the number of side information symbols that are used simultaneously in the initial round of download at the nth database. For a non-decreasing sequence $\{n_i\}_{i=0}^{M-1} \subset \{1, \cdots, N\}^M$, the databases are divided into groups, such that group 0 contains databases $1 : n_0$, group 1 contains databases $n_0 + 1 : n_1$, etc. Let $s_n = i$ for all $n_{i-1}+1 \leq n \leq n_i$ with $n_{-1} = 0$ by convention. Denote $\mathcal{S} = \{i : s_n = i \text{ for some } n \in \{1, \cdots, N\}\}$. Denote $y_\ell[k]$ to be the number of stages in round k downloaded from the nth database, such that $n_{\ell-1}+1 \leq n \leq n_\ell$. First, the user permutes each message independently and uniformly. The details of the scheme are as follows:

1) *Calculation of the repetitions:* The scheme associated with $\mathbf{n} = \{n_i\}_{i=0}^{M-1}$ is repeated ν times such that the answer string length $t_n(\mathbf{n}, \boldsymbol{\mu})$ satisfies:

$$t_n(\mathbf{n}, \boldsymbol{\mu}) = \frac{\nu D_n(\mathbf{n})}{1-\mu_n} \in \mathbb{N}, \quad \forall n \in \{1, \cdots, N\} \tag{28}$$

where $D_n(\mathbf{n})$ is the number of meaningful downloads of one repetition of the scheme associated with $\{n_i\}_{i=0}^{M-1}$.

2) *Preparation of the keys:* The nth database generates a random key K_n. The random key K_n is of length $\mu_n t_n$, whose elements are independent and uniformly distributed over $\mathbb{F}_q$. The nth database encodes K_n to an *artificial noise* vector $u_{[1:t_n]}^{(n)}$ using a $(t_n, \mu_n t_n)$ MDS code, i.e., $u_{[1:t_n]}^{(n)} = \mathbf{MDS}_{t_n \times \mu_n t_n} K_n$.
3) *Initial download:* From the nth database where $1 \leq n \leq n_0$, the user downloads $\prod_{s \in \mathcal{S}} \binom{M-2}{s-1}$ symbols from the desired message. The user sets the round index $k = 1$.
4) *Message symmetry:* To satisfy the privacy constraint, for each stage initiated in the previous step, the user completes the stage by downloading the remaining $\binom{M-1}{k-1}$ k-sums that do not include the desired symbols.
5) *Database symmetry:* We divide the databases into groups. Group $\ell \in \mathcal{S}$ contains databases $n_{\ell-1}+1$ to n_ℓ. Database symmetry is applied within each group only.
6) *Exploitation of side information:* The user downloads $(k+1)$-sum consisting of 1 desired symbol and a k-sum of undesired symbols that were generated in the kth round. If $s_n > k$, the nth database does not exploit the side information generated in the kth round. For $s_n = k$, extra side information can be used in the nth database. The user forms $n_0 \prod_{s \in \mathcal{S} \setminus \{s_n\}} \binom{M-2}{s-1}$ stages of side information by constructing k-sums of the undesired symbols in round 1 from the databases in group 0.
7) *Repeat* steps 5, 6, 7 after setting $k = k+1$ until $k = M$.
8) *Repeat* steps $3, \cdots, 7$ for a total of ν repetitions.
9) *Shuffling the order of the queries:* The user shuffles the order of queries as in [8] to guarantee the privacy.
10) *Encryption of the downloads:* The database encrypts each meaningful download by adding one symbol from $u_{[1:(1-\mu_n)t_n]}^{(n)}$. Furthermore, the user downloads $u_{[(1-\mu_n)t_n+1:t_n]}^{(n)}$ coded key symbols individually.

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 symmetric private information retrieval. In *IEEE Globecom*, Dec 2016.

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

[9] 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.

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

[11] 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.

[12] 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.

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

[14] H. Sun and S. Jafar. Multiround private information retrieval: Capacity and storage overhead. 2016. Available at arXiv:1611.02257.

[15] 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.

[16] H. Sun and S. Jafar. Private information retrieval from MDS coded data with colluding servers: Settling a conjecture by Freij-Hollanti et al. 2017. Available at arXiv: 1701.07807.

[17] 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.

[18] 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.

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

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

[21] K. Banawan and S. Ulukus. The capacity of private information retrieval from Byzantine and colluding databases. *IEEE Trans. on Info. Theory*. Submitted June 2017. Also available at arXiv:1706.01442.

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

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

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

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

[26] 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.

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

[28] 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.

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

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

[31] 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.

[32] 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.

[33] J. Garay, R. Gennaro, C. Jutla, and T. Rabin. Secure distributed storage and retrieval. *Theoretical Computer Science*, 243(1):363 – 389, 2000.

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

[35] Q. Wang and M. Skoglund. Secure private information retrieval from colluding databases with eavesdroppers. 2017. Available at arXiv: 1710.01190.

[36] L. H. Ozarow and A. D. Wyner. Wire-tap channel II. *AT&T Bell Laboratories Technical Journal*, 63(10):2135–2157, December 1984.

[37] 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.

[38] 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.