

Quantum Symmetric Private Information Retrieval with Secure Storage and Eavesdroppers

Alptug Aytekin Mohamed Nomeir Sajani Vithana Sennur Ulukus

Department of Electrical and Computer Engineering

University of Maryland, College Park, MD 20742

aaaytekin@umd.edu mnomeir@umd.edu spallego@umd.edu ulukus@umd.edu

Abstract—We consider both the classical and quantum variations of X -secure, E -eavesdropped and T -colluding symmetric private information retrieval (SPIR). This is the first work to study SPIR with X -security in classical or quantum variations. We first develop a scheme for classical X -secure, E -eavesdropped and T -colluding SPIR (XSETSPIR) based on a modified version of cross subspace alignment (CSA), which achieves a rate of $R = 1 - \frac{X + \max(T, E)}{N}$. The modified scheme achieves the same rate as the scheme used for X -secure PIR with the extra benefit of symmetric privacy. Next, we extend this scheme to its quantum counterpart based on the N -sum box abstraction. This is the first work to consider the presence of eavesdroppers in quantum private information retrieval (QPIR). In the quantum variation, the eavesdroppers have better access to information over the quantum channel compared to the classical channel due to the over-the-air decodability. To that end, we develop another scheme specialized to combat eavesdroppers over quantum channels. The scheme proposed for X -secure, E -eavesdropped and T -colluding quantum SPIR (XSETQSPIR) in this work maintains the super-dense coding gain from the shared entanglement between the databases, i.e., achieves a rate of $R_Q = \min \left\{ 1, 2 \left(1 - \frac{X + \max(T, E)}{N} \right) \right\}$.

I. INTRODUCTION

In the private information retrieval (PIR) problem introduced in [1], a user wishes to retrieve a message out of K messages stored in N databases without revealing the index of the required message to any of the databases. The optimal rate of PIR with N databases and K replicated messages is shown to be $C(N, K) = \left(1 + \frac{1}{N} + \dots + \frac{1}{N^{K-1}} \right)^{-1}$ in [2]. Subsequently, several variations of this problem have been studied with different requirements for the databases and the user. In [3], symmetric PIR (SPIR) is introduced, where the user is not allowed to obtain any information about the message set other than the required message. The capacity of SPIR is shown to be $1 - \frac{1}{N}$ in [3], which is also $C(N, \infty)$. In [4], T -colluding PIR is introduced where any T databases can share the queries received from the user to learn the required message index. The capacity of T -colluding PIR is shown to be $\left(1 + \frac{T}{N} + \dots + \frac{T^{K-1}}{N^{K-1}} \right)^{-1}$ in [4], which is also $C\left(\frac{N}{T}, K\right)$. T -colluding SPIR is considered in [5] and its capacity is shown to be $1 - \frac{T}{N}$, which is $C\left(\frac{N}{T}, \infty\right)$. In [6], the E -eavesdropped, T -colluding SPIR is introduced. In this setting, there is an eavesdropper that can listen to all answers from any E databases to the user. The capacity for this case is shown to be $1 - \frac{\max(T, E)}{N}$ in [6]. The problem of X -secure PIR is introduced in [7], where the messages need to be hidden

from the databases themselves even when X databases share their complete datasets. In [8], the asymptotic capacity of X -secure T -colluding PIR, i.e., the capacity when $K \rightarrow \infty$, is shown to be $1 - \frac{X+T}{N}$. Some other variations of the PIR and SPIR problems have been studied and different applications have been introduced in [9]–[23]; see also [24].

The problem of quantum PIR (QPIR) is recently introduced in [25]. In this model, the message bits are sent over a quantum channel from the databases to the user, and the databases can share entanglement between them. [25] shows that the capacity of symmetric QPIR (SQPIR) is 1 when the number of databases is $N \geq 2$. Variations of QPIR include T -colluding QPIR with and without coded storage [26]–[28], QPIR with noisy channels [29], and several other variations analogous to their classical counterparts [30]–[32]. Most recently, [33] has proposed a mathematical abstraction for the entanglement between transmitters sending information to a common receiver over separate quantum channels. The work in [33] shows that the entanglement between N transmitters that use Pauli operators to encode classical messages to quantum states can be represented mathematically as a multiple input multiple output (MIMO) multiple access channel (MAC) with $2N$ inputs and N outputs, i.e., a matrix with $N \times 2N$ dimensions. In addition, this matrix must have elements from a finite field, and must satisfy the strongly self-orthogonal (SSO) property. Using these properties, [33] shows that the rate of X -secure T -colluding QPIR for their proposed scheme is $R_Q = \min \left\{ 1, 2 \left(1 - \frac{X+T}{N} \right) \right\}$. This is a doubling of the classical rate $R_C = 1 - \frac{X+T}{N}$ in the regime of interest.

In this paper, we focus on both classical and quantum variations of the SPIR problem with a passive eavesdropper that listens to queries and answers going into and out of any of the E databases. In addition, up to T databases collude, and up to X databases communicate. This is the first work that considers X -secure SPIR in general, even in the classical domain, and even without E -eavesdropped and T -colluding databases. We show that the rate of the modified CSA scheme (modified for the symmetric privacy) R is the same as the rate of the CSA scheme proposed in [8] with the extra benefit of symmetric privacy, i.e., $R = 1 - \frac{X + \max(T, E)}{N}$. In addition, this is the first work to consider the presence of eavesdroppers in QPIR. We develop a QPIR scheme that maintains privacy and security against eavesdroppers, colluding databases and communicating databases. Our proposed

quantum scheme achieves the maximum super-dense coding gain when an entangled state is shared between the databases, i.e., $R_Q = \min \left\{ 1, 2 \left(1 - \frac{X + \max(T, E)}{N} \right) \right\}$. The QPIR problem with eavesdroppers is more complex compared to the classical PIR problem with eavesdroppers due to the over-the-air decodability imposed by the N -sum box abstraction. The quantum scheme we propose in this work achieves double the rate of its classical counterpart.

II. PRELIMINARIES

In this section, we state some important definitions related to quantum physics and quantum information theory [34]. We use these quantities subsequently to formulate the X -secure E -eavesdropped T -colluding QSPIR.

Definition 1 (Quantum density matrices) For a general quantum system A , that can be in the state $|\psi_j\rangle$ with probability p_j , the quantum density matrix ρ_A is defined as,

$$\rho_A = \sum_j p_j |\psi_j\rangle \langle \psi_j|, \quad (1)$$

with $p_j \geq 0$, $\sum_j p_j = 1$.

Definition 2 (Von Neumann entropy) For the density matrix ρ , Von Neumann entropy is defined as,

$$S(\rho) = -\text{tr}(\rho \log \rho) = H(\Lambda), \quad (2)$$

where $\text{tr}(\cdot)$ is the trace operator, Λ are the eigenvalues of ρ , and $H(\cdot)$ is the Shannon entropy. For a quantum system A with density matrix ρ_A , we define $S(A) = S(\rho_A)$.

Definition 3 (Quantum relative entropy) The relative entropy between two density matrices ρ and σ is defined as,

$$D(\rho \parallel \sigma) = \text{tr}(\rho(\log \rho - \log \sigma)). \quad (3)$$

Definition 4 (Quantum conditional entropy) The conditional entropy of a quantum system A with respect to a system B is defined as,

$$S(A|B) = S(A, B) - S(B). \quad (4)$$

Definition 5 (Quantum mutual information) The quantum mutual information between two quantum systems A and B is defined as,

$$S(A; B) = S(A) + S(B) - S(A, B) \quad (5)$$

$$= S(A) - S(A|B). \quad (6)$$

In the next section, we formulate the problem in both classical and quantum variations.

III. PROBLEM FORMULATION

The system consists of N databases and a user who wants to retrieve a message. Out of the N databases, T are allowed to collude, i.e., share the user's queries, and X are allowed to communicate, i.e., share their storage to decode the messages.

In addition, any E links are accessible to the eavesdroppers that can listen to E of the user's queries and databases' answers. The system contains K messages, W_1, \dots, W_K , of equal length L , that are independent and identically distributed (i.i.d.). The messages are generated uniformly at random from the field \mathbb{F}_q , with $q = p^r$, where p is any prime number. Thus,

$$H(W_k) = L, \quad k \in [1 : K], \quad (7)$$

$$H(W_{[1:K]}) = \sum_{k=1}^K H(W_k) = KL. \quad (8)$$

The messages $W_{[1:K]}$ need to be secure against any X communicating databases,

$$I(W_{[1:K]}; S_{\mathcal{X}}) = 0, \quad (9)$$

where $S_{\mathcal{X}}$ denotes all of the stored data in subset \mathcal{X} databases satisfying $|\mathcal{X}| \leq X$. The user wants to retrieve a message W_θ , where θ is chosen uniformly at random from $[1 : K]$, and sends a query to each database ($Q_1^{[\theta]}, \dots, Q_N^{[\theta]}$) denoted by $Q_{[1:N]}^{[\theta]}$. As the user does not know the messages, the queries are generated independent of the message content,

$$I(W_{[1:K]}; Q_{[1:N]}^{[\theta]}) = 0, \quad \theta \in [1 : K]. \quad (10)$$

In addition, we require that the index of the retrieved message by the user is private against any T colluding databases,

$$I(\theta; Q_{\mathcal{T}}^{[\theta]}) = 0, \quad \theta \in [1 : K], \quad (11)$$

where $\mathcal{T} \subset [1 : N]$, $|\mathcal{T}| \leq T$.

Upon receiving the queries, the n th database replies with a deterministic answer string $A_n^{[\theta]}$ based on its received query $Q_n^{[\theta]}$, shared common randomness between the databases S , and stored data, S_n , $n \in [1 : N]$,

$$H(A_n^{[\theta]} | S_n, Q_n^{[\theta]}, S) = 0, \quad \theta \in [1 : K]. \quad (12)$$

When the user receives all answer strings $A_{[1:N]}^{[\theta]}$, the required message must be decodable based on the answer strings and the sent queries,

$$H(W_\theta | A_{[1:N]}^{[\theta]}, Q_{[1:N]}^{[\theta]}) = 0, \quad \theta \in [1 : K]. \quad (13)$$

In addition, the symmetric privacy constraint requires that the user gains no information about the message set except for the required message,

$$I(W_{\theta^c}; A_{[1:N]}^{[\theta]} | Q_{[1:N]}^{[\theta]}, \theta) = 0, \quad \theta \in [1 : K], \quad (14)$$

where W_{θ^c} denotes all other messages aside from the required message W_θ .

Finally, the scheme must be private and secure against an eavesdropper who can listen to any set of E queries and E answers,

$$I(\theta; Q_{\mathcal{E}_1}^{[\theta]}, A_{\mathcal{E}_2}^{[\theta]}) = 0, \quad \theta \in [1 : K], \quad (15)$$

and

$$I(W_{[1:K]}; A_{\mathcal{E}_1}^{[\theta]} | Q_{\mathcal{E}_2}^{[\theta]}) = 0, \quad \theta \in [1 : K], \quad (16)$$

where $\mathcal{E}_1, \mathcal{E}_2 \subset [1 : N]$, $|\mathcal{E}_1|, |\mathcal{E}_2| \leq E$.

The rate R of any scheme satisfying the above requirements is defined as the ratio between the length of the required message and the average length of the answer strings,

$$R = \frac{L}{H(\mathcal{A}_{[1:N]}^{[\theta]})}. \quad (17)$$

In the X -secure, E -eavesdropped, T -colluding quantum symmetric PIR (XSETQSPIR) problem, we follow the system models introduced in the literature [25]–[27], [31]. The databases store S_n , $n \in [1 : N]$, as classical bits and share an entangled state of N quantum bits denoted by ρ . The user sends the queries $Q_{[1:N]}^{[\theta]}$ over a classical channel to each of the N databases, and each database n , $n \in [1 : N]$, with the quantum system $\mathcal{A}_n^0 = \text{tr}_{j \neq n}^{[1:N]}(\rho)$, where $\text{tr}(\cdot)$ is the trace operator, replies to the user queries over a separate quantum channel. Upon receiving the query, the n th database performs the quantum operation Enc_n based on the received query, storage and \mathcal{A}_n^0 to produce the quantum state $\mathcal{A}_n^{[\theta]}$, $n \in [1 : N]$, as follows,

$$\mathcal{A}_n^{[\theta]} = \text{Enc}_n(Q_n^{[\theta]}, S_n, \mathcal{A}_n^0, \Lambda_n, \mathcal{S}), \quad \theta \in [1 : K], \quad (18)$$

where Enc_n is the n th database's encoder, and Λ_n is a masking random variable sent by the user to the databases.¹ The final received state at the user is given as,

$$\mathcal{A}_{[1:N]}^{[\theta]} = \mathcal{A}_1^{[\theta]} \otimes \dots \otimes \mathcal{A}_N^{[\theta]}, \quad \theta \in [1 : K], \quad (19)$$

where \otimes is the tensor product. Since the storage is in the form of classical bits and the queries are sent over classical channels, constraints (9)-(11) must hold. It is also required that the index of the required message be secure against the received queries and masking random variables $\Lambda_{[1:N]}$ for any $\mathcal{T} \subset [1 : N]$, $|\mathcal{T}| \leq T$ colluding databases,

$$I(\theta; Q_{\mathcal{T}}^{[\theta]}, \Lambda_{\mathcal{T}}) = 0, \quad \theta \in [1 : K]. \quad (20)$$

Additionally, the Von Neumann entropy of the required message W_θ given the queries and the answers must be zero,

$$S(W_\theta | \mathcal{A}_{[1:N]}^{[\theta]}, Q_{[1:N]}^{[\theta]}, \Lambda_{[1:N]}) = 0, \quad \theta \in [1 : K], \quad (21)$$

and for symmetric privacy, the quantum mutual information between the other messages \mathcal{W}_{θ^c} and the received quantum densities $\mathcal{A}_{[1:N]}^{[\theta]}$ must satisfy,

$$S(\mathcal{W}_{\theta^c}; \mathcal{A}_{[1:N]}^{[\theta]} | Q_{[1:N]}^{[\theta]}, \theta, \Lambda_{[1:N]}) = 0, \quad \theta \in [1 : K]. \quad (22)$$

In addition, for the eavesdroppers who listen to any E classical and quantum channels, the privacy and security requirements must be satisfied,

$$S(\theta; Q_{\mathcal{E}_1}^{[\theta]}, \mathcal{A}_{\mathcal{E}_2}^{[\theta]}, \Lambda_{\mathcal{E}_3}) = 0, \quad \theta \in [1 : K], \quad (23)$$

$$S(W_{[1:K]}; \mathcal{A}_{\mathcal{E}_1}^{[\theta]} | Q_{\mathcal{E}_2}^{[\theta]}, \Lambda_{\mathcal{E}_3}) = 0, \quad \theta \in [1 : K], \quad (24)$$

where $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3 \subset [1 : N]$ and $|\mathcal{E}_1|, |\mathcal{E}_2|, |\mathcal{E}_3| \leq E$. Then,

¹The main reason for the masking random variables is to fight over-the-air decodability in quantum channels. It is discussed in detail in Section V-B.

the XSETQSPIR rate R_Q for the retrieval scheme satisfying (9)-(11) and (20)-(24) is defined as

$$R_Q = \frac{H(W_\theta)}{\dim(\mathcal{A}_1^{[\theta]} \otimes \dots \otimes \mathcal{A}_N^{[\theta]})}. \quad (25)$$

In this paper, we follow the encoding and decoding structure using the N -sum box abstraction introduced recently in [33]. In the encoding stage, the databases use Pauli operators $X(a) = \sum_{j=0}^{q-1} |j+a\rangle \langle j|$, and $Z(a) = \sum_{j=0}^{q-1} \omega^{tr(aj)} |j\rangle \langle j|$, where $q = p^r$ with p as any prime number, $a \in \mathbb{F}_q$ and $\omega = \exp(2\pi i/p)$. In the decoding stage, the user applies projective value measurement (PVM) defined on the quotient space of the stabilizer group $\mathcal{L}(\mathcal{V})$ defined by

$$\mathcal{L}(\mathcal{V}) = \{c_v \tilde{W}(v) : v \in \mathcal{V}\}, \quad (26)$$

where \mathcal{V} is a self orthogonal subspace in \mathbb{F}_q^{2N} ,

$$\tilde{W}(v) = X(v_1)Z(v_{N+1}) \otimes \dots \otimes X(v_N)Z(v_{2N}), \quad (27)$$

and $c_v \in \mathbb{C}$ is chosen such that $\mathcal{L}(\mathcal{V})$ is an Abelian subgroup of HW_q^N with $c_v I_q^N$ being an element of the stabilizer group if $c_v = 1$, where HW_q^N is the Heisenberg-Weyl group defined as,

$$HW_q^N = \{c \tilde{W}(s) : s \in \mathbb{F}_q^{2N}, c \in \mathbb{C} \setminus \{0\}\}. \quad (28)$$

In the next section, we state our main results for this problem, both for the classical and the quantum variations.

IV. MAIN RESULTS

Theorem 1 For classical X -secure, E -eavesdropped, T -colluding SPIR (XSETSPIR) with N databases, the rate given by

$$R = 1 - \frac{X + \max(T, E)}{N}, \quad (29)$$

is achievable, using modified cross subspace alignment (CSA) with message length $L = N - \max(T, E) - X$.

Remark 1 When $X = 0$ and $E = 0$, the proposed scheme achieves the optimal rate for T -colluding SPIR, $R = 1 - \frac{T}{N}$, found in [3], [5].

Remark 2 When $X = 0$, the proposed scheme achieves the optimal rate for E -eavesdropped, T -colluding SPIR, $R = 1 - \frac{\max(T, E)}{N}$, found in [6].

Remark 3 For $X \geq 1$ the exact capacity of X -secure PIR with a fixed number of messages K is still an open problem.

Theorem 2 For X -secure, E -eavesdropped, T -colluding quantum SPIR (XSETQSPIR) with N databases which are allowed to share entanglement and have quantum channels for answer strings, the rate given by

$$R_Q = \min \left\{ 1, 2 \left(1 - \frac{X + \max(T, E)}{N} \right) \right\}, \quad (30)$$

is achievable with modified quantum CSA.

Remark 4 When $X = 0$ and $E = 0$, the proposed scheme achieves the capacity of T -colluding $QSPIR$, $R_Q = \min\{1, 2(1 - \frac{T}{N})\}$, found in [26].

V. ACHIEVABLE SCHEME

Before describing the achievable scheme for the quantum X -secure, E -eavesdropped, T -colluding $SPiR$, we first introduce the classical scheme which uses the modified classical CSA to solve the classical version of the problem.

A. Achievable Scheme in the Classical Setting: $XSETSPiR$

Consider a total of N databases with the T -colluding, E -eavesdropped and X -secure setting. Let the message length L be $L = N - X - M$, where $M = \max(E, T)$. The storage at each database n denoted by S_n is,

$$S_n = \begin{bmatrix} W_{\cdot,1} + \sum_{i=1}^X (f_1 - \alpha_n)^i R_{1i} \\ W_{\cdot,2} + \sum_{i=1}^X (f_2 - \alpha_n)^i R_{2i} \\ \vdots \\ W_{\cdot,L} + \sum_{i=1}^X (f_L - \alpha_n)^i R_{Li} \end{bmatrix}, \quad (31)$$

where $W_{\cdot,j} = [W_{1,j}, \dots, W_{K,j}]^T$ is a vector representing the j th bit of all K messages, with $W_{i,j}$ being the j th bit of message i , R_{ij} are uniform independent random vectors with the same dimensions as $W_{\cdot,j}$, and $\{f_i\}_{i=1}^L, \{\alpha_n\}_{n=1}^N$ are globally known distinct constants from \mathbb{F}_q .

The user wishes to retrieve W_θ while protecting its privacy from any T colluding databases and E eavesdroppers. The user sends the query $Q_n^{[\theta]}$ to the n th database as,

$$Q_n^{[\theta]} = \begin{bmatrix} \frac{\prod_{i=1}^L (f_i - \alpha_n)}{f_1 - \alpha_n} \left(e_\theta + \sum_{i=1}^M (f_1 - \alpha_n)^i Z_{1i} \right) \\ \vdots \\ \frac{\prod_{i=1}^L (f_i - \alpha_n)}{f_L - \alpha_n} \left(e_\theta + \sum_{i=1}^M (f_L - \alpha_n)^i Z_{Li} \right) \end{bmatrix}, \quad (32)$$

where e_θ is a vector of length K with a 1 in the θ th index and zero otherwise, and Z_{ij} are uniform independent random vectors of length K each, chosen by the user.

Since the databases want to hide any information about the messages other than the user-required message, they agree on $X + M - 1$ independent uniform random variables Z'_1, \dots, Z'_{X+M-1} before the retrieval process starts, i.e., they share common randomness, where all $X + M - 1$ common randomness variables Z'_i are random noise symbols from \mathbb{F}_q . Each database n , $n \in [1 : N]$, then computes the answer to be sent to the user as,

$$\begin{aligned} A_n^{[\theta]} &= S_n^t Q_n^{[\theta]} + P_n \\ &= \gamma_n \left(\sum_{i=1}^L \frac{1}{f_i - \alpha_n} W_{\theta,i} + \sum_{i=0}^{X+M-1} \alpha_n^i (I_i + Z'_i) \right), \end{aligned} \quad (33)$$

$$(34)$$

where $\gamma_n = \prod_{i=1}^L (f_i - \alpha_n)$, $P_n = \sum_{i=0}^{X+M-1} \alpha_n^i Z'_i$, and I_i is the coefficient of α_n^i in the polynomial resulting from the product $S_n^t Q_n^{[\theta]}$. After receiving all the answer strings from the N databases, the user has the following answer vector,

from which the required L symbols of W_θ can be obtained, as $X + M + L = N$,

$$\begin{aligned} A^{[\theta]} &= [A_1^{[\theta]}, \dots, A_N^{[\theta]}]^t \\ &= B_N(\alpha, f) \\ &\quad \times [W_{\theta,1}, \dots, W_{\theta,L}, I_0 + Z'_0, \dots, I_{X+M-1} + Z'_{X+M-1}]^t, \end{aligned} \quad (35)$$

$$(36)$$

where t represents the transpose operation, $\alpha = [\alpha_1, \dots, \alpha_N]^t$, $f = [f_1, \dots, f_L]^t$, and $B_N(\alpha, f)$ is an $N \times N$ invertible matrix given by,

$$B_N(\alpha, f) = \text{diag}(\gamma) \begin{bmatrix} \frac{1}{f_1 - \alpha_1} & \cdots & \frac{1}{f_L - \alpha_1} & 1 & \alpha_1 & \cdots & \alpha_1^{X+M-1} \\ \frac{1}{f_1 - \alpha_2} & \cdots & \frac{1}{f_L - \alpha_2} & 1 & \alpha_2 & \cdots & \alpha_2^{X+M-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{f_1 - \alpha_N} & \cdots & \frac{1}{f_L - \alpha_N} & 1 & \alpha_N & \cdots & \alpha_N^{X+M-1} \end{bmatrix}, \quad (37)$$

where $\gamma = [\gamma_1, \dots, \gamma_N]^t$. The main difference between the $N - L$ interference symbols here and the interference symbols in [8] is that they are contaminated with random noise unknown to the user, i.e., Z'_i terms, which leak no information to the user except for the required L bits.

Remark 5 Compared to the CSA scheme, the proposed symmetric CSA scheme achieves the same rate with the extra benefit of symmetric privacy.

B. Achievable Scheme in the Quantum Setting: $XSETQSPiR$

To develop the quantum scheme based on the N -sum box abstraction [33], we first recall some important definitions in [33].

Definition 6 (QCSA matrix) The quantum CSA (QCSA) matrix of size $N \times N$ and elements from \mathbb{F}_q designed to retrieve $2L$ symbols in the quantum PIR scheme is defined as follows

$$D_N(\alpha, \beta, f)[i, j] = \begin{cases} \frac{\beta_i}{f_j - \alpha_i}, & j \leq L, \\ \beta_i \alpha_i^{j-L-1}, & L < j \leq N, \end{cases} \quad (38)$$

where $\alpha = [\alpha_1, \dots, \alpha_N]^t$, $\beta = [\beta_1, \dots, \beta_N]^t$, $f = [f_1, \dots, f_N]^t$, $\alpha_1, \dots, \alpha_N, f_1, \dots, f_L$ are distinct, β_1, \dots, β_N are non-zero and $L \leq \frac{N}{2}$.

Definition 7 (Dual QCSA matrices) The matrices H_N^u and H_N^v are defined as $H_N^u = D_N(\alpha, u, f)$ and $H_N^v = D_N(\alpha, v, f)$. Then, H_N^u and H_N^v are dual QCSA matrices if:

- 1) u_1, \dots, u_N are non-zero,
- 2) u_1, \dots, u_N are distinct,
- 3) for each v_j , $j \in [1 : N]$,

$$v_j = \frac{1}{u_j} \left(\prod_{\substack{i=1 \\ i \neq j}}^N (\alpha_j - \alpha_i) \right)^{-1}. \quad (39)$$

Using these definitions, we restate the N -sum box feasibility theorem from [33, Thm. 6].

Theorem 3 For any dual QCSA matrices H_N^u and H_N^v , there exists a feasible N -sum box transfer matrix $G(u, v)$ of size $N \times 2N$ given as follows,

$$G(u, v) = G_N \begin{bmatrix} H_N^u & 0 \\ 0 & H_N^v \end{bmatrix}^{-1}, \quad (40)$$

where

$$G_N = \begin{bmatrix} I_L & 0_{L \times \nu} & 0 & 0 & 0 & 0 \\ 0 & 0 & I_{\mu-L} & 0 & 0 & 0 \\ 0 & 0 & 0 & I_L & 0_{L \times \mu} & 0 \\ 0 & 0 & 0 & 0 & 0 & I_{\nu-L} \end{bmatrix},$$

$\nu = \lceil N/2 \rceil$, $\mu = \lfloor N/2 \rfloor$, I_L is the identity matrix of size $L \times L$, and $0_{A,B}$ is the all zeros matrix of size $A \times B$.

Remark 6 We explain the main concept behind Theorem 3 as follows: If u and v are chosen such that H_N^u and H_N^v are dual QCSA matrices, then there exists an N -entangled qubit shared between the N databases such that the quantum channels between the databases and the user can be represented by $G(u, v)$.

Remark 7 A main difference between the quantum channel and the classical channel is that the decoding is done over-the-air. This implies that if the eavesdropper listens to E answers, there is a possibility that it can get up to E out of the L symbols. This means that the eavesdropper is more powerful in the quantum variation compared to the classical variation.

Now, we are ready to describe the XSETQSPIR scheme. The storage at each database is slightly modified compared to the classical case. The storage in the quantum scheme S_Q is given as,

$$S_Q = [S_n(1)^t, S_n(2)^t]^t, \quad (41)$$

where $S_n(1)$ and $S_n(2)$ are as in (31), i.e., each containing $L = N - X - M \leq \frac{N}{2}$ new symbols of the K messages, along with new random noise vectors. In other words, the length of the messages considered in the quantum scheme is twice of what was considered in the classical case. To retrieve the required message, the user sends the query $Q_n^{[\theta]}$ to database n , which is of the same form as in the classical scheme in (32). Each database n , $n \in [1 : N]$, then generates the noise added answers as in (33),

$$\hat{A}_n^{[\theta]}(1) = S_n(1)^t Q_n^{[\theta]} + P_n(1) \quad (42)$$

$$\hat{A}_n^{[\theta]}(2) = S_n(2)^t Q_n^{[\theta]} + P_n(2) \quad (43)$$

where

$$P_n(1) = \sum_{i=0}^{X+M-1} \alpha_n^i Z'_i(1), \quad P_n(2) = \sum_{i=0}^{X+M-1} \alpha_n^i Z'_i(2) \quad (44)$$

with all $Z'_i(j)$ being random noise symbols. To prevent the eavesdropper from decoding over-the-air, the user sends two masking variables to each database n , given by,

$$\begin{aligned} \Lambda_n(\kappa) &= \gamma_n \left(\frac{1}{f_1 - \alpha_n} \lambda_1(\kappa) + \dots + \frac{1}{f_L - \alpha_n} \lambda_L(\kappa) \right. \\ &\quad \left. + \lambda_{L+1}(\kappa) + \alpha_n \lambda_{L+2}(\kappa) + \dots + \alpha_n^{N-L-1} \lambda_N(\kappa) \right), \end{aligned} \quad (45)$$

for $\kappa \in [1, 2]$, where $\lambda_n(\kappa)$, $n \in [1 : N]$, $\kappa \in [1 : 2]$ are uniform independent random variables generated by the user. Then, each database generates two answer instances $A_n^{[\theta]}(1)$, $A_n^{[\theta]}(2)$,

$$A_n^{[\theta]}(1) = \hat{A}_n^{[\theta]}(1) + \Lambda_n(1) \quad (46)$$

$$A_n^{[\theta]}(2) = \hat{A}_n^{[\theta]}(2) + \Lambda_n(2). \quad (47)$$

The N initial answers from the N databases are written compactly as,

$$A = [A_1^{[\theta]}(1), \dots, A_N^{[\theta]}(1), A_1^{[\theta]}(2), \dots, A_N^{[\theta]}(2)]^t \quad (48)$$

$$= \begin{bmatrix} \text{diag}(\gamma) & 0 \\ 0 & \text{diag}(\gamma) \end{bmatrix} \begin{bmatrix} D_N & 0 \\ 0 & D_N \end{bmatrix} \begin{bmatrix} X(1) \\ X(2) \end{bmatrix} \quad (49)$$

where $D_N = D_N(\alpha, 1_N, f)$, $\gamma = [\gamma_1, \dots, \gamma_N]^t$, and

$$X(i) = \begin{bmatrix} W_{\theta,1}(i) + \lambda_1(i) \\ W_{\theta,2}(i) + \lambda_2(i) \\ \vdots \\ W_{\theta,L}(i) + \lambda_L(i) \\ I_0(i) + Z'_0(i) + \lambda_{L+1}(i) \\ \vdots \\ I_{X+M-1}(i) + Z'_{X+M-1}(i) + \lambda_N(i) \end{bmatrix}. \quad (50)$$

for $i \in [1 : 2]$. Then, to make use of the entanglement and quantum channels, the answers are modified as,

$$\tilde{A} = \begin{bmatrix} \text{diag}(u) & 0 \\ 0 & \text{diag}(v) \end{bmatrix} \begin{bmatrix} \text{diag}(\gamma) & 0 \\ 0 & \text{diag}(\gamma) \end{bmatrix}^{-1} A, \quad (51)$$

where $u = [u_1, \dots, u_N]^t$ and $v = [v_1, \dots, v_N]^t$ are chosen such that they satisfy Definition 7. These answers are sent through the quantum channels using the encoder defined by the Pauli operators, i.e., each database sends its answer instances $\tilde{A}_n(1)$, $\tilde{A}_n(2)$, $n \in [1 : N]$ as follows,

$$\mathcal{A}_n^{[\theta]} = Z(\tilde{A}_n(2))X(\tilde{A}_n(1))\mathcal{A}_n^0. \quad (52)$$

Based on the properties of the quantum channel, the N symbols received by the user, denoted by y are given as,

$$y = G(u, v)\tilde{A} \quad (53)$$

$$= G(u, v) \begin{bmatrix} H_N^u & 0 \\ 0 & H_N^v \end{bmatrix} \begin{bmatrix} X(1) \\ X(2) \end{bmatrix} \quad (54)$$

$$= G_N \begin{bmatrix} X(1) \\ X(2) \end{bmatrix} \quad (55)$$

$$= [W_{\theta,1}(1) + \lambda_1(1), \dots, W_{\theta,L}(1) + \lambda_L(1), I'(1), \\ W_{\theta,1}(2) + \lambda_1(2), \dots, W_{\theta,L}(2) + \lambda_L(2), I'(2)]^t, \quad (56)$$

where $I'(1)$ represents the last $\lfloor N/2 \rfloor - L$ interference symbols of $X(1)$ in (50), and $I'(2)$ represent the last $\lfloor N/2 \rfloor - L$ interference symbols of $X(2)$ in (50). As the user already knows the values of $\lambda_\ell(\kappa)$ for $\ell \in [1 : L]$ and $\kappa \in [2]$, the user obtains the $2L$ symbols of the required message W_θ , denoted by $W_{\theta,1}(1), \dots, W_{\theta,L}(1), W_{\theta,1}(2), \dots, W_{\theta,L}(2)$.

Remark 8 *In this scheme, we use the fact that the length of each message sub-packet L must satisfy both $L = N - X - \max(T, E)$, and $L \leq \frac{N}{2}$. If $L > \frac{N}{2}$, we drop the extra databases as in [33].*

Remark 9 *Note that since u and v can be globally known, the no-cloning theorem cannot be invoked, thus the eavesdropper can listen to quantum channels.*

Remark 10 *Due to the over-the-air decoding, the user needs to send masking variables, $\lambda_1, \dots, \lambda_N$, to the N databases, thus N^2 bits in total. However, in our proposed scheme the user needs only to send 1 bit to each database over the non-secure channel, i.e., N bits in total, to achieve the same goal.*

VI. CONCLUSIONS

In this paper, we studied the classical and quantum variations of the X -secure, E -eavesdropped, and T -colluding symmetric PIR. In the classical variation, we developed a scheme that achieves symmetric privacy at the same rate as the state-of-the-art scheme that solves the same problem without symmetric privacy. In the quantum variation, we uncovered how the eavesdroppers have better access to the transmitted answer strings due to the over-the-air decodability imposed by the N -sum box abstraction. To that end, we designed a scheme that represses over-the-air decodability while maintaining the super-dense coding gain, i.e., doubling the rate compared to the classical variation.

REFERENCES

- [1] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. Private information retrieval. *Jour. of the ACM*, 45(6):965–981, November 1998.
- [2] H. Sun and S. A. Jafar. The capacity of private information retrieval. *IEEE Trans. Info. Theory*, 63(7):4075–4088, July 2017.
- [3] H. Sun and S. A. Jafar. The capacity of symmetric private information retrieval. *IEEE Trans. Info. Theory*, 65(1):322–329, June 2018.
- [4] 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. Info. Theory*, 64(2):1000–1022, December 2017.
- [5] Q. Wang and M. Skoglund. Symmetric private information retrieval from MDS coded distributed storage with non-colluding and colluding servers. *IEEE Trans. Info. Theory*, 65(8):5160–5175, March 2019.
- [6] Q. Wang, H. Sun, and M. Skoglund. The capacity of private information retrieval with eavesdroppers. *IEEE Trans. Info. Theory*, 65(5):3198–3214, December 2018.
- [7] H. Yang, W. Shin, and J. Lee. Private information retrieval for secure distributed storage systems. *IEEE Trans. Info. Foren. Security*, 13(12):2953–2964, May 2018.
- [8] Z. Jia, H. Sun, and S. A. Jafar. Cross subspace alignment and the asymptotic capacity of X -secure T -private information retrieval. *IEEE Trans. Info. Theory*, 65(9):5783–5798, May 2019.
- [9] X. Yao, N. Liu, and W. Kang. The capacity of private information retrieval under arbitrary collusion patterns for replicated databases. *IEEE Trans. Info. Theory*, 67(10):6841–6855, July 2021.

- [10] K. Banawan and S. Ulukus. The capacity of private information retrieval from Byzantine and colluding databases. *IEEE Trans. Info. Theory*, 65(2):1206–1219, September 2018.
- [11] K. Banawan and S. Ulukus. Private information retrieval through wiretap channel II: Privacy meets security. *IEEE Trans. Info. Theory*, 66(7):4129–4149, February 2020.
- [12] K. Banawan and S. Ulukus. Multi-message private information retrieval: Capacity results and near-optimal schemes. *IEEE Trans. Info. Theory*, 64(10):6842–6862, April 2018.
- [13] K. Banawan and S. Ulukus. The capacity of private information retrieval from coded databases. *IEEE Trans. Info. Theory*, 64(3):1945–1956, January 2018.
- [14] M. A. Attia, D. Kumar, and R. Tandon. The capacity of private information retrieval from uncoded storage constrained databases. *IEEE Trans. Info. Theory*, 66(11):6617–6634, September 2020.
- [15] K. Banawan, B. Arasli, Y.-P. Wei, and S. Ulukus. The capacity of private information retrieval from heterogeneous uncoded caching databases. *IEEE Trans. Info. Theory*, 66(6):3407–3416, June 2020.
- [16] N. Raviv, I. Tamo, and E. Yaakobi. Private information retrieval in graph-based replication systems. *IEEE Trans. Info. Theory*, 66(6):3590–3602, November 2019.
- [17] K. Banawan, B. Arasli, and S. Ulukus. Improved storage for efficient private information retrieval. In *IEEE ITW*, August 2019.
- [18] M. J. Siavoshani, S. P. Shariatpanahi, and M. Ali Maddah-Ali. Private information retrieval for a multi-message scenario with private side information. *IEEE Trans. on Comm.*, 69(5):3235–3244, January 2021.
- [19] Y.-P. Wei, K. Banawan, and S. Ulukus. Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching. *IEEE Trans. Info. Theory*, 65(5):3215–3232, November 2018.
- [20] Z. Wang and S. Ulukus. Symmetric private information retrieval at the private information retrieval rate. *IEEE Jour. on Selected Areas in Info. Theory*, 3(2):350–361, June 2022.
- [21] S. Vithana, K. Banawan, and S. Ulukus. Semantic private information retrieval. *IEEE Trans. Info. Theory*, 68(4):2635–2652, December 2021.
- [22] A. Heidarzadeh, S. Kadhe, S. El Rouayheb, and A. Sprintson. Single-server multi-message individually-private information retrieval with side information. In *IEEE ISIT*, July 2019.
- [23] R. Tajeddine, O. Gnilke, and S. El Rouayheb. Private information retrieval from MDS coded data in distributed storage systems. *IEEE Trans. Info. Theory*, 64(11):7081–7093, March 2018.
- [24] S. Ulukus, S. Avestimehr, M. Gastpar, S. A. Jafar, R. Tandon, and C. Tian. Private retrieval, computing and learning: Recent progress and future challenges. *IEEE Jour. on Selected Areas in Commun.*, 40(3):729–748, March 2022.
- [25] S. Song and M. Hayashi. Capacity of quantum private information retrieval with multiple servers. *IEEE Trans. Info. Theory*, 67(1):452–463, September 2021.
- [26] S. Song and M. Hayashi. Capacity of quantum private information retrieval with colluding servers. *IEEE Trans. Info. Theory*, 67(8):5491–5508, May 2021.
- [27] S. Song and M. Hayashi. Capacity of quantum symmetric private information retrieval with collusion of all but one of servers. *IEEE Jour. on Selected Areas in Info. Theory*, 2(1):380–390, January 2021.
- [28] M. Allaix, S. Song, L. Holzbaur, T. Pllaha, M. Hayashi, and C. Hollanti. On the capacity of quantum private information retrieval from mds-coded and colluding servers. *IEEE Jour. on Selected Areas in Commun.*, 40(3):885–898, January 2022.
- [29] Y. Yang, P. Yang, G. Xu, Y. Zhou, and W. Shi. Quantum private information retrieval over a collective noisy channel. *Modern Physics Letters A*, 38(01):2350001, 2023.
- [30] S. Song and M. Hayashi. Quantum private information retrieval for quantum messages. In *IEEE ISIT*, September 2021.
- [31] P. Saarela, M. Allaix, R. Freij-Hollanti, and C. Hollanti. Private information retrieval from colluding and byzantine servers with binary reed-muller codes. In *IEEE ISIT*, August 2022.
- [32] M. Allaix, L. Holzbaur, T. Pllaha, and C. Hollanti. High-rate quantum private information retrieval with weakly self-dual star product codes. In *IEEE ISIT*, July 2021.
- [33] M. Allaix, Y. Lu, Y. Yao, T. Pllaha, C. Hollanti, and S. Jafar. N -sum box: An abstraction for linear computation over many-to-one quantum networks. 2023. Available online at arXiv:2304.07561.
- [34] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.