Quantum X-Secure E-Eavesdropped T-Colluding Symmetric Private Information Retrieval*

Alptug Aytekin Mohamed Nomeir Sajani Vithana Sennur Ulukus

Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742
aaytekin@umd.edu *mnomeir@umd.edu* *spallego@umd.edu* *ulukus@umd.edu*

Abstract

We consider both 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, i.e., user-privacy as well as database-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 two different schemes for quantum X-secure, E-eavesdropped and T-colluding SPIR (QXSETSPIR) with secure over-the-air decoding. The first scheme achieves the highest possible super-dense coding gain, i.e., $R_Q = \min\left\{1, 2\left(1 - \frac{X+\max(T,E)}{N}\right)\right\}$, which requires additional uploads from the user. The second scheme on the other hand requires no extra uploads. However, it does not achieve the super-dense coding gain in some cases based on the relation between the number of eavesdropped links and the number of interference terms. The second scheme is based on the idea that there exist some special entanglement states that can be used to hide the contents of the user-required messages from the eavesdroppers using the interference symbols.

1 Introduction

In the private information retrieval (PIR) problem introduced in [2], a user wishes to retrieve a message out of K messages stored in N databases without revealing the index of the

*A partial result of this work has been accepted for publication in 2023 IEEE Globecom [1].

required message to any of the databases. The optimal rate for the PIR problem with N databases and K replicated messages is shown to be $C(N,K) = (1 + \frac{1}{N} + \ldots + \frac{1}{N^{K-1}})^{-1}$ in [3]. Subsequently, several variations of this problem have been studied with different requirements for the databases and the user. In [4], 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 $1 - \frac{1}{N}$ as shown in [4], which is also $C(N, \infty)$. T-colluding PIR is introduced in [5] 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 $(1 + \frac{T}{N} + \ldots + \frac{T^{K-1}}{N^{K-1}})^{-1}$, which is equivalent to $C(\frac{N}{T}, K)$. T-colluding SPIR is considered in [6] and its capacity is shown to be $1 - \frac{T}{N}$, which is $C(\frac{N}{T}, \infty)$. In [7], 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 along with T-colluding databases. The capacity for this case is shown to be $1 - \frac{\max(T,E)}{N}$ in [7]. The problem of X-secure PIR is introduced in [8], where the messages need to be hidden from the databases themselves even when X databases share their complete datasets. In [9], the X-secure T-colluding PIR capacity for an asymptotic number of messages $K \to \infty$ is shown to be $1 - \frac{X+T}{N}$. Many other variations of the PIR and SPIR problems have been studied and different applications have been introduced in [10–21].

The problem of quantum PIR (QPIR) is recently introduced in [22]. 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. It is shown in [22] that the capacity of symmetric QPIR (QSPIR) is 1 when the number of databases is $N \geq 2$. Variations of QPIR include T-colluding QPIR with and without coded storage [23–25], QPIR with noisy channels [26], and several other variations analogous to their classical counterparts [27, 28]. Most recently, [29] proposed a mathematical abstraction for the entanglement between transmitters sending information to a common receiver over separate quantum channels. The work in [29] 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, [29] 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 essentially double the corresponding classical rate given by $R_C = 1 - \frac{X+T}{N}$. It is worth mentioning that the N-sum box abstraction is suitable for other quantum settings as well, see for instance [30, 31].

In this paper, we focus on both classical and quantum variations of the SPIR problem with a passive eavesdropper that listens to the queries and answers going into and out of any of the E links of the databases. In addition, up to any T and X databases are allowed to collude and communicate, respectively. Considering the classical case as a standalone problem, this

is the first work that studies SPIR with the X-security requirement, even if eavesdropping is neglected. In this problem, we develop a scheme based on CSA that ensures symmetric privacy. This is done by manipulating the transmitted answers such that the interference symbols appear as complete random noise to the user. We show that this scheme achieves the same rate as the state-of-the-art scheme that does not satisfy the symmetric privacy requirement [9]. In the QPIR setting, this is the first work to consider the presence of eavesdroppers that can listen to communications over quantum channels. We show that the eavesdroppers have more power in quantum channels compared to the classical ones. The main reason behind this is that the retrieval scheme is globally known, and the user receives the required symbols after performing globally known projective-value measurements without further processing. This is known as *over-the-air decodability* in the quantum variation of PIR. Note that the eavesdropper listening to any E links can perform the same projective-value measurements and retrieve up to E symbols of the required message.

To combat over-the-air decodability of the eavesdropper, we develop two different schemes. In the first scheme, we introduce a mechanism that masks the answers from the databases to the user, to prevent the eavesdroppers from decoding the message contents, while also ensuring that the user is able to decode. This scheme achieves the maximum super-dense coding gain, i.e., $R_Q = \min\left\{1, 2\left(1 - \frac{X+\max(T,E)}{N}\right)\right\}$, when the databases share the entanglement state. However, this scheme requires extra uploads from the user. To alleviate this requirement, we develop another scheme that makes use of interference symbols to hide the message contents from the eavesdroppers (this is not applicable in the classical case). The main idea behind the second scheme is the use of certain entanglement states that can preserve the privacy of the user-required message against the eavesdroppers. However, the second scheme is only able to achieve the super-dense coding gain in certain special cases.

2 Problem Formulation

The system consists of N databases, T of which may collude, X of which may communicate, and E of which may be eavesdropped on. The system contains K messages, $W_1, \ldots, W_K$, of equal length L, that are independent and identically distributed. The variables N, T, X, E, and K are assumed to be globally known. The messages are generated uniformly at random from the field $\mathbb{F}_q$, with $q = p^r$, where p is any prime number. Thus, in q-ary bits,

$$H(W_k) = L, \quad k \in \{1, \ldots, K\}, \tag{1}$$

$$H(W_{[1:K]}) = \sum_{k=1}^{K} H(W_k) = KL. \tag{2}$$

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

$$I(W_{[1:K]}; S_{\mathcal{X}}) = 0, \tag{3}$$

where $S_{\mathcal{X}}$ is all stored data in any set of $\mathcal{X}$ 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 $Q_n^{[\theta]}$ to each database n. The set of queries sent to all N databases is denoted by $Q_{[1:N]}^{[\theta]}$. As the user is unaware of the messages, the queries generated are independent of the messages content, i.e.,

$$I(W_{[1:K]}; Q_{[1:N]}^{[\theta]}) = 0, \quad \theta \in [1:K]. \tag{4}$$

In addition, we require that the index of the retrieved message by the user is secure against any T colluding databases, i.e.,

$$I(\theta; Q_{\mathcal{T}}^{[\theta]}) = 0, \quad \theta \in [1:K], \tag{5}$$

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

Upon receiving the query, database n replies with a deterministic answer string $A_n^{[\theta]}$ based on its received query $Q_n^{[\theta]}$, shared common randomness between the databases $\mathcal{S}$, and stored data S_n, i.e.,

$$H(A_n^{[\theta]} | S_n, Q_n^{[\theta]}, \mathcal{S}) = 0, \quad n \in [1:N], \quad \theta \in [1:K]. \tag{6}$$

The required message must be decodable to the user based on the answer strings received $A_{[1:N]}^{[\theta]}$ and the transmitted queries, i.e.,

$$H(W_\theta | A_{[1:N]}^{[\theta]}, Q_{[1:N]}^{[\theta]}) = 0, \quad \theta \in [1:K]. \tag{7}$$

In addition, the database privacy constraint[1] requires that the user gains no information about the message set except for the required message, i.e.,

$$I(\mathcal{W}_{\theta^C}; A_{[1:N]}^{[\theta]} | Q_{[1:N]}^{[\theta]}, \theta) = 0, \quad \theta \in [1:K], \tag{8}$$

where $\mathcal{W}_{\theta^C}$ are all other messages aside from the required message W_θ.

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

$$I(\theta; Q_{\mathcal{E}_1}^{[\theta]}, A_{\mathcal{E}_2}^{[\theta]}) = 0, \quad \theta \in [1:K], \tag{9}$$

[1] In the literature of PIR, database privacy requires that no information on the *contents* of the messages beyond what is required should be revealed to the user. However, it is not related to hiding the *indices* of the unwanted messages.

and

$$I(W_{[1:K]}; A_{\mathcal{E}_1}^{[\theta]} | Q_{\mathcal{E}_2}^{[\theta]}) = 0, \quad \theta \in [1:K], \tag{10}$$

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. Thus,

$$R = \frac{L}{H(A_{[1:N]}^{[\theta]})}. \tag{11}$$

In the quantum X-secure, E-eavesdropped, T-colluding quantum symmetric PIR (QXSET-SPIR) problem, we follow the system models introduced in the literature [22–24]. 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 $\mathcal{A}_n^0 = tr_{\substack{j=[1:N] \\ j \neq n}}(\rho)$, where $tr(\cdot)$ is the trace operator, replies to the user queries over a separate quantum channel. Upon receiving the query, each database n performs the quantum operation Enc_n based on the received query, storage and $\mathcal{A}_n^0$ to produce the quantum state $\mathcal{A}_n^{[\theta]}$ as,

$$\mathcal{A}_n^{[\theta]} = Enc_n(Q_n^{[\theta]}, S_n, \mathcal{A}_n^0, \Lambda_n, \mathcal{S}), \quad \theta \in [1:K], \quad n \in [1:N], \tag{12}$$

where Enc_n is the nth database's encoder, and Λ_n is a masking random variable sent by the user to the databases.[2] The final received state at the user is given as,

$$\mathcal{A}_{[1:N]}^{[\theta]} = \mathcal{A}_1^{[\theta]} \otimes \ldots \otimes \mathcal{A}_N^{[\theta]}, \quad \theta \in [1:K], \tag{13}$$

where $\otimes$ is the tensor product. Since the storage is in the form of classical bits and the queries are sent over classical channels, constraint (3) 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.e,

$$I(\theta; Q_{\mathcal{T}}^{[\theta]}, \Lambda_{\mathcal{T}}) = 0, \quad \theta \in [1:K]. \tag{14}$$

Additionally, the Von Neumann entropy, $S(\cdot)$,[3] of the required message W_θ given the queries and the answers must be zero,

$$S(W_\theta | \mathcal{A}_{[1:N]}^{[\theta]}, Q_{[1:N]}^{[\theta]}, \Lambda_{[1:N]}) = 0, \quad \theta \in [1:K]. \tag{15}$$

[2]The main reason for the masking random variables is to combat the over-the-air decodability in quantum channels. This is examined in detail in Section 4.2. For the scheme presented in Section 4.3, the masking variables can be dropped.

[3]The Von Neumann entropy, conditional entropy, and quantum mutual information are defined in Section 6.

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

$$S(\mathcal{W}_{\theta^C}; \mathcal{A}^{[\theta]}_{[1:N]} | Q^{[\theta]}_{[1:N]}, \theta, \Lambda_{[1:N]}) = 0, \quad \theta \in [1:K]. \tag{16}$$

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

$$S(\theta; Q^{[\theta]}_{\mathcal{E}_1}, \mathcal{A}^{[\theta]}_{\mathcal{E}_2}, \Lambda_{\mathcal{E}_3}) = 0, \quad \theta \in [1:K], \tag{17}$$

and

$$S(W_{[1:K]}; \mathcal{A}^{[\theta]}_{\mathcal{E}_1} | Q^{[\theta]}_{\mathcal{E}_2}, \Lambda_{\mathcal{E}_3}) = 0, \quad \theta \in [1:K], \tag{18}$$

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

The QXSETSPIR rate R_Q for the retrieval scheme satisfying (3)-(5) and (14)-(18) is defined as,

$$R_Q = \frac{H(W_\theta)}{\log dim(\mathcal{A}^{[\theta]}_1 \otimes \ldots \otimes \mathcal{A}^{[\theta]}_N)}, \tag{19}$$

where $dim(A)$ is the dimension of the vector space spanned by A.

In this paper, we follow the encoding and decoding structure using the N-sum box abstraction introduced recently in [29]. In the encoding stage, the databases use Pauli operators $\mathsf{X}(a) = \sum_{j=0}^{q-1} |j+a\rangle \langle j|$, and $\mathsf{Z}(a) = \sum_{j=0}^{q-1} \omega^{tr(aj)} |j\rangle \langle j|$, 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,

$$\mathcal{L}(\mathcal{V}) = \{c_v \tilde{W}(v) : v \in \mathcal{V}\}, \tag{20}$$

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

$$\tilde{W}(v) = \mathsf{X}(v_1)\mathsf{Z}(v_{N+1}) \otimes \ldots \otimes \mathsf{X}(v_N)\mathsf{Z}(v_{2N}), \tag{21}$$

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,

$$HW_q^N = \{c\tilde{W}(s) : s \in \mathbb{F}_q^{2N}, c \in \mathbb{C} \setminus \{0\}\}. \tag{22}$$

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

3 Main Results

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

$$R = 1 - \frac{X + \max(T, E)}{N}, \tag{23}$$

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 [4, 6].*

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 [7].*

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

$$R_Q = \min\left\{1, 2\left(1 - \frac{X + \max(T, E)}{N}\right)\right\}, \tag{24}$$

is achievable with modified quantum CSA.

Remark 4 *When $X = 0$ and $E = 0$, the proposed scheme achieves the capacity of T-colluding quantum SPIR, $R_Q = \min\left\{1, 2\left(1 - \frac{T}{N}\right)\right\}$, found in [23].*

Theorem 3 *For quantum X-secure, E-eavesdropped, T-colluding SPIR (QXSETSPIR) with N databases which are allowed to share entanglement and have quantum channels for answer strings and no extra upload cost is allowed, the rate given by*

$$R_Q = \begin{cases} \min\left\{1, 2\left(1 - \frac{X+M}{N}\right)\right\}, & E \leq N - 2\min\{\frac{N}{2}, N - X - M\} \\ \min\left\{1 - \frac{E}{N}, 2\left(1 - \frac{X+M+\frac{\delta}{2}}{N}\right)\right\}, & E > N - 2\min\{\frac{N}{2}, N - X - M\} \end{cases}, \tag{25}$$

where $M = \max(T, E)$, and $\delta = E - \left(N - 2\min\{\frac{N}{2}, N - X - M\}\right)$, is achievable with modified quantum CSA.

Remark 5 *As noted before, due to over-the-air decoding, the scheme only achieves the super-dense coding gain when $E \leq N - 2\min\{\frac{N}{2}, N - X - M\}$.*

Remark 6 *To give a numerical example for illustration, assume that $N = 10$, $X = 2$, $T = 5$, and $E = 3$. Then, $E < 10 - 6 = 4$, which corresponds to the first case in Theorem 3. Thus, the achievable rate is $R_Q = 3/5$. However, if $E = 5$ (in general if $E > 4$), we have less number of interference symbols that can be used to hide the message symbols, which results in a reduced rate, $R_Q = 1/2$.*

4 Achievable Schemes

In this section, we first present the proposed scheme for classical XSETSPIR, which is based on a modified version of CSA. We then describe the quantum version of the scheme.

4.1 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 given by,

$$S_n = \begin{bmatrix} W_{\cdot,1} + (f_1 - \alpha_n)R_{11} + (f_1 - \alpha_n)^2 R_{12} + \ldots + (f_1 - \alpha_n)^X R_{1X} \\ W_{\cdot,2} + (f_2 - \alpha_n)R_{21} + (f_2 - \alpha_n)^2 R_{22} + \ldots + (f_2 - \alpha_n)^X R_{2X} \\ \vdots \\ W_{\cdot,L} + (f_L - \alpha_n)R_{L1} + (f_L - \alpha_n)^2 R_{L2} + \ldots + (f_L - \alpha_n)^X R_{LX} \end{bmatrix}, \tag{26}$$

where $W_{\cdot,j} = [W_{1,j}, \ldots, W_{K,j}]^t$ is a vector representing the jth bit of all K messages, with $W_{i,j}$ being the jth bit of message i, R_{ij} are uniform independent random vectors with the same dimensions as $W_{\cdot,j}$, $j \in [1 : L]$, $a_i, f_j \in \mathbb{F}_q$ are distinct where $i = 1, \ldots, N$ and $j = 1, \ldots, L$, and t denotes the transpose operator.

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 nth database as,

$$Q_n^{[\theta]} = \begin{bmatrix} \frac{1}{f_1 - \alpha_n}\left(e_\theta + (f_1 - \alpha_n)Z_{11} + \ldots + (f_1 - \alpha_n)^M Z_{1M}\right) \\ \vdots \\ \frac{1}{f_L - \alpha_n}\left(e_\theta + (f_L - \alpha_n)Z_{L1} + \ldots + (f_L - \alpha_n)^M Z_{LM}\right) \end{bmatrix}, \tag{27}$$

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

As the databases do not want any information on the messages other than what is required to be leaked to the user, they agree on $X + M - 1$ independent uniform random variables $Z'_1, \ldots, Z'_{X+M-1}$ 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,

$$A_n^{[\theta]} = S_n^t Q_n^{[\theta]} + P_n \tag{28}$$

$$= \sum_{i=1}^{L} \frac{1}{f_i - \alpha_n} W_{\theta,i} + \sum_{i=0}^{X+M-1} \alpha_n^i (I_i + Z_i'), \tag{29}$$

where $P_n = \sum_{i=0}^{X+M-1} \alpha_n^i Z_i'$, 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 answers from the N databases, the user obtains the required L symbols of W_θ, by solving the following equation, as $X + M + L = N$,

$$A^{[\theta]} = \begin{bmatrix} A_1^{[\theta]} \\ \vdots \\ A_N^{[\theta]} \end{bmatrix} = B_N(\alpha, f)[W_{\theta,1}, \ldots, W_{\theta,L}, I_0 + Z_0', \ldots, I_{X+M-1} + Z_{X+M-1}']^t, \tag{30}$$

where $\alpha = [\alpha_1, \ldots, \alpha_N]^t$, $f = [f_1, \ldots, f_L]^t$, and $B_N(\alpha, f)$ is an $N \times N$ invertible Cauchy-Vandermonde matrix is given by,

$$B_N(\alpha, f) = \begin{bmatrix} \frac{1}{f_1-\alpha_1} & \ldots & \frac{1}{f_L-\alpha_1} & 1 & \alpha_1 & \ldots & \alpha_1^{X+M-1} \\ \frac{1}{f_1-\alpha_2} & \ldots & \frac{1}{f_L-\alpha_2} & 1 & \alpha_2 & \ldots & \alpha_2^{X+M-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{f_1-\alpha_N} & \ldots & \frac{1}{f_L-\alpha_N} & 1 & \alpha_N & \ldots & \alpha_N^{X+M-1} \end{bmatrix}. \tag{31}$$

The main difference between the $N-L$ interference symbols here and the interference symbols in [9] 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. In Section 6, we prove that this scheme achieves symmetric privacy and protects against any E-eavesdroppers, T-colluding, and X-communicating databases.

Remark 7 *Compared to the CSA scheme, the proposed symmetric CSA scheme achieves the same rate with the extra benefit of symmetric privacy.*

4.2 Achievable Scheme in the Quantum Setting: QXSETSPIR

To develop the quantum scheme based on the N-sum box abstraction [29], we first recall some important definitions in [29].

Definition 1 (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,*

$D_N(\alpha, \beta, f) =$

$$\begin{bmatrix} \frac{\beta_1}{f_1-\alpha_1} & \frac{\beta_1}{f_2-\alpha_1} & \dots & \frac{\beta_1}{f_L-\alpha_1} & \beta_1 & \beta_1\alpha_1 & \dots & \beta_1\alpha_1^{\lfloor N/2\rfloor-1} & \beta_1\alpha_1^{\lceil N/2\rceil-1} & \dots & \beta_1\alpha_1^{N-L-1} \\ \frac{\beta_2}{f_1-\alpha_2} & \frac{\beta_2}{f_2-\alpha_2} & \dots & \frac{\beta_2}{f_L-\alpha_2} & \beta_2 & \beta_2\alpha_2 & \dots & \beta_2\alpha_2^{\lfloor N/2\rfloor-1} & \beta_2\alpha_2^{\lceil N/2\rceil-1} & \dots & \beta_2\alpha_2^{N-L-1} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \frac{\beta_N}{f_1-\alpha_N} & \frac{\beta_N}{f_2-\alpha_N} & \dots & \frac{\beta_N}{f_L-\alpha_N} & \beta_N & \beta_N\alpha_N & \dots & \beta_N\alpha_N^{\lfloor N/2\rfloor-1} & \beta_N\alpha_N^{\lceil N/2\rceil-1} & \dots & \beta_N\alpha_N^{N-L-1} \end{bmatrix}, \tag{32}$$

where $\alpha = [\alpha_1, \dots, \alpha_N]^t$, $f = [f_1, \dots, f_L]^t$, *with* $\{f_i\}_{i=1}^L, \{\alpha_i\}_{i=1}^N$ *being distinct elements from* $\mathbb{F}_q$, $\beta = [\beta_1, \dots, \beta_N]^t$, *with* $\beta_1, \dots, \beta_N$ *being non-zero constants from* $\mathbb{F}_q$, *and* $L \leq \frac{N}{2}$.

Definition 2 (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* $v_j,\ j \in [1:N]$,

$$v_j = \frac{1}{u_j}\left(\prod_{\substack{i=1\\ i\neq j}}^{N}(\alpha_j - \alpha_i)\right)^{-1}. \tag{33}$$

Based on these definitions, we restate the definition of a stabilizer-based N-sum box and and the feasibility theorem from [29, Thm. 1 and Thm. 6].

Theorem 4 *A stabilizer-based* N*-sum box transfer matrix is an* $N \times 2N$ *transfer matrix* M *that satisfies,*

$$M = \begin{bmatrix} 0_{N\times N} & I_N \end{bmatrix} \begin{bmatrix} G_{2N\times N} & H_{2N\times N} \end{bmatrix}^{-1}, \tag{34}$$

where I_N *is the identity matrix of size* $N \times N$, $0_{A\times B}$ *is the all zeros matrix of size* $A \times B$ *and* G *is an SSO matrix, i.e.,* $G^t J G = 0$, $\mathrm{rank}(G) = N$, *and*

$$J = \begin{bmatrix} 0 & -I_N \\ I_N & 0 \end{bmatrix}. \tag{35}$$

Theorem 5 *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 by,*

$$G(u,v) = \begin{bmatrix} I_L & 0_{L\times\lceil N/2\rceil} & 0 & 0 & 0 & 0 \\ 0 & 0 & I_{\lfloor N/2\rfloor-L} & 0 & 0 & 0 \\ 0 & 0 & 0 & I_L & 0_{L\times\lfloor N/2\rfloor} & 0 \\ 0 & 0 & 0 & 0 & 0 & I_{\lceil N/2\rceil-L} \end{bmatrix} \begin{bmatrix} H_N^u & 0 \\ 0 & H_N^v \end{bmatrix}^{-1}. \tag{36}$$

Remark 8 *To simply explain the main concept behind Theorem 5: 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 9 *A main difference between the quantum channel and the classical channel is that the decoding is done over-the-air in the former. 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 setting compared to the classical one.*

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

$$S_Q = [S_n(1)^t,\ S_n(2)^t]^t, \tag{37}$$

where $S_n(1)$ and $S_n(2)$ are as in (26), 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 (27). Each database n, $n \in [1 : N]$, then generates the noise added answers as in (28),

$$\hat{A}_n^{[\theta]}(1) = S_n(1)^t Q_n^{[\theta]} + P_n(1) \tag{38}$$

$$\hat{A}_n^{[\theta]}(2) = S_n(2)^t Q_n^{[\theta]} + P_n(2) \tag{39}$$

where $P_n(1) = \sum_{i=0}^{X+M-1} \alpha_n^i Z_i'(1)$ and $P_n(2) = \sum_{i=0}^{X+M-1} \alpha_n^i Z_i'(2)$ with all $Z_i'(j)$ being random noise symbols. To prevent the eavesdropper from decoding over-the-air, the user sends a set of masking variables as follows, to each database n,

$$\begin{aligned}\Lambda_n(\kappa) =& \frac{1}{f_1 - \alpha_n}\lambda_1(\kappa) + \ldots + \frac{1}{f_L - \alpha_n}\lambda_L(\kappa) + \lambda_{L+1}(\kappa) \\ &+ \alpha_n \lambda_{L+2}(\kappa) + \ldots + \alpha_n^{N-L-1}\lambda_N(\kappa), \quad \kappa \in [1 : 2],\end{aligned} \tag{40}$$

where $\lambda_n(\kappa),\ n \in [1 : N],\ \kappa \in [1 : 2]$ are uniform independent random variables generated by the user. For each generated answer instance $\hat{A}_n^{[\theta]}(i),\ i \in [1 : 2]$ the databases add the masking variables to protect the answers from the eavesdroppers.

$$A_n^{[\theta]}(1) = \hat{A}_n^{[\theta]}(1) + \Lambda_n(1) \tag{41}$$

$$A_n^{[\theta]}(2) = \hat{A}_n^{[\theta]}(2) + \Lambda_n(2). \tag{42}$$

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

$$A = \begin{bmatrix} A_1^{[\theta]}(1) \\ \vdots \\ A_N^{[\theta]}(1) \\ A_1^{[\theta]}(2) \\ \vdots \\ A_N^{[\theta]}(2) \end{bmatrix} = \begin{bmatrix} D_N(\alpha, 1_N, f) & 0 \\ 0 & D_N(\alpha, 1_N, f) \end{bmatrix} \begin{bmatrix} X(1) \\ X(2) \end{bmatrix} \tag{43}$$

where

$$X(1) = \begin{bmatrix} W_{\theta,1}(1) + \lambda_1(1) \\ W_{\theta,2}(1) + \lambda_2(1) \\ \vdots \\ W_{\theta,L}(1) + \lambda_L(1) \\ I_0(1) + Z_0'(1) + \lambda_{L+1}(1) \\ \vdots \\ I_{X+M-1}(1) + Z'_{X+M-1}(1) + \lambda_N(1) \end{bmatrix}, \quad X(2) = \begin{bmatrix} W_{\theta,1}(2) + \lambda_1(2) \\ W_{\theta,2}(2) + \lambda_2(2) \\ \vdots \\ W_{\theta,L}(2) + \lambda_L(2) \\ I_0(2) + Z_0'(2) + \lambda_{L+1}(2) \\ \vdots \\ I_{X+M-1}(2) + Z'_{X+M-1}(2) + \lambda_N(2) \end{bmatrix}. \tag{44}$$

Then, to make use of the entanglement and quantum channels, the answers are modified as,

$$\tilde{A} = \begin{bmatrix} \mathrm{diag}(u) & 0 \\ 0 & \mathrm{diag}(v) \end{bmatrix} A, \tag{45}$$

where $u = [u_1, \ldots, u_N]^t$, and $v = [v_1, \ldots, v_N]^t$ are chosen such that they satisfy Definition 2. These answers are sent through the quantum channels. Based on the properties of the quantum channel, the N symbols received by the user, denoted by y are given as,

$$y = G(u, v)\tilde{A} \tag{46}$$

$$= G(u, v) \begin{bmatrix} H_N^u & 0 \\ 0 & H_N^v \end{bmatrix} \begin{bmatrix} X(1) \\ X(2) \end{bmatrix} \tag{47}$$

$$= \begin{bmatrix} I_L & 0_{L\times\lceil N/2\rceil} & 0 & 0 & 0 & 0 \\ 0 & 0 & I_{\lfloor N/2\rfloor - L} & 0 & 0 & 0 \\ 0 & 0 & 0 & I_L & 0_{L\times\lfloor N/2\rfloor} & 0 \\ 0 & 0 & 0 & 0 & 0 & I_{\lceil N/2\rceil - L} \end{bmatrix} \begin{bmatrix} X(1) \\ X(2) \end{bmatrix} \tag{48}$$

$$= \begin{bmatrix} W_{\theta,1}(1) + \lambda_1(1) \\ \vdots \\ W_{\theta,L}(1) + \lambda_L(1) \\ I'(1) \\ W_{\theta,1}(2) + \lambda_1(2) \\ \vdots \\ W_{\theta,L}(2) + \lambda_L(2) \\ I'(2) \end{bmatrix} \tag{49}$$

where $I'(1)$ represents the last $\lfloor N/2 \rfloor - L$ interference symbols of $X(1)$ in (44), and $I'(2)$ represents the last $\lceil N/2 \rceil - L$ interference symbols of $X(2)$ in (44). 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 $W_{\theta,1}(1), \ldots, W_{\theta,L}(1), W_{\theta,1}(2), \ldots, W_{\theta,L}(2)$.

Remark 10 *In this scheme, we use the fact that $L \leq \frac{N}{2}$. If $L > \frac{N}{2}$, we drop the extra databases as in [29].*

Remark 11 *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 12 *Due to the over-the-air decoding, the user needs to send masking variables, $\lambda_1, \ldots, \lambda_N$, to the N databases, i.e., a total of N^2 bits. However, in our proposed scheme the user only sends one bit to each database over the non-secure channel, i.e., N bits in total, to achieve the same goal.*

Remark 13 *If $L \geq E$, we can use each masking variable more than once, thus decreasing the extra upload cost.*

4.3 Achievable Scheme in the Quantum Setting with No Extra Uploads

In this section, we propose a scheme that does not require any extra uploads by the user to combat the eavesdropper, i.e., the masking random variables, $\lambda_{[1:N]}$, introduced in Section 4.2 are not used. However, the rate is affected in some cases as stated in Theorem 3. The main idea here is the usage of certain entanglement states that are useful in hiding messages using interference symbols to provide security against eavesdroppers. In the first case of this scheme, we consider the case where the number of eavesdropped links is less than the number of interference terms, and in the second case, we consider the opposite setting.

4.3.1 Case 1: $E \leq N - 2\min\{\frac{N}{2}, N - X - M\}$

If the number of interference terms is greater than the number of eavesdropped links, the interference terms can be used to mask the required message symbols. To this end, we define the channel transition matrix as,

$$G'(u,v) = \begin{bmatrix} (V_N(b))_{\mathcal{I}_1,\mathcal{I}_1} & 0_{L\times\lceil N/2\rceil} & (V_N(b))_{\mathcal{I}_1,\mathcal{I}_3} & (V_N(b))_{\mathcal{I}_1,\mathcal{I}_2} & 0_{L\times\lfloor\frac{N}{2}\rfloor} & (V_N(b))_{\mathcal{I}_1,\mathcal{I}_4} \\ (V_N(b))_{\mathcal{I}_3,\mathcal{I}_1} & 0_{L\times\lceil N/2\rceil} & (V_N(b))_{\mathcal{I}_3,\mathcal{I}_3} & (V_N(b))_{\mathcal{I}_3,\mathcal{I}_2} & 0_{L\times\lfloor\frac{N}{2}\rfloor} & (V_N(b))_{\mathcal{I}_3,\mathcal{I}_4} \\ (V_N(b))_{\mathcal{I}_2,\mathcal{I}_1} & 0_{L\times\lceil N/2\rceil} & (V_N(b))_{\mathcal{I}_2,\mathcal{I}_3} & (V_N(b))_{\mathcal{I}_2,\mathcal{I}_2} & 0_{L\times\lfloor\frac{N}{2}\rfloor} & (V_N(b))_{\mathcal{I}_2,\mathcal{I}_4} \\ (V_N(b))_{\mathcal{I}_4,\mathcal{I}_1} & 0_{L\times\lceil N/2\rceil} & (V_N(b))_{\mathcal{I}_4,\mathcal{I}_3} & (V_N(b))_{\mathcal{I}_4,\mathcal{I}_2} & 0_{L\times\lfloor\frac{N}{2}\rfloor} & (V_N(b))_{\mathcal{I}_4,\mathcal{I}_4} \end{bmatrix} \begin{bmatrix} H_N^u & 0 \\ 0 & H_N^u \end{bmatrix} \tag{50}$$

$$= \Pi^t V_N(b) \Pi \begin{bmatrix} I_L & 0_{L\times\lceil N/2\rceil} & 0 & 0 & 0 & 0 \\ 0 & 0 & I_{\lfloor N/2\rfloor - L} & 0 & 0 & 0 \\ 0 & 0 & 0 & I_L & 0_{L\times\lfloor N/2\rfloor} & 0 \\ 0 & 0 & 0 & 0 & 0 & I_{\lceil N/2\rceil - L} \end{bmatrix} \begin{bmatrix} H_N^u & 0 \\ 0 & H_N^u \end{bmatrix}, \tag{51}$$

where

$$\Pi = \begin{bmatrix} I_L & 0_{L\times(\lfloor\frac{N}{2}\rfloor - L)} & 0_L & 0_{L\times(\lceil\frac{N}{2}\rceil - L)} \\ 0 & 0 & I_L & 0 \\ 0 & I_{\lfloor\frac{N}{2}\rfloor - L} & 0 & 0 \\ 0 & 0 & 0 & I_{\lceil\frac{N}{2}\rceil - L} \end{bmatrix}, \tag{52}$$

and $b = [b_1, \ldots, b_N]^t$ is a vector of length N with distinct elements from $\mathbb{F}_q$. $V_N(b)$ is the $N \times N$ Vandermonde matrix given by,

$$V_N(b) = \begin{bmatrix} 1 & b_1 & b_1^2 & \ldots & b_1^{N-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & b_N & b_N^2 & \ldots & b_N^{N-1} \end{bmatrix}, \tag{53}$$

and $\mathcal{I}_1 = (1, 2, \ldots, L)$, $\mathcal{I}_2 = (L+1, 2, \ldots, 2L)$, $\mathcal{I}_3 = (2L+1, 2, \ldots, \lfloor\frac{N}{2}\rfloor + L)$ and $\mathcal{I}_4 = (\lfloor\frac{N}{2}\rfloor + L + 1, \ldots, N)$. $(V_N(b))_{\mathcal{I}_i,\mathcal{I}_j}$ denotes the submatrix of $V_N(b)$ given by the rows and columns indicated by the ordered tuples $\mathcal{I}_i$ and $\mathcal{I}_j$, respectively. Lemma 9 (and its proof) in Section 6 shows that this channel transition matrix is indeed a valid N-sum box construction.

The storage and query structure are identical to the scheme defined in Section 4.2. The difference between this scheme and the scheme presented in Section 4.2 is the fact that the user does not upload any other information to the databases except the queries that convey the required message. Databases then generate the answers, $A_n^{[\theta]}(1)$, $A_n^{[\theta]}(2)$ as in (38)-(39), i.e., there is no masking required in this scheme. Similar to the scheme defined in Section 4.2, the answers from all the N databases can be written in the same form as (43), where in that

case $X(1)$ and $X(2)$ can be written as

$$X(1)=\begin{bmatrix} W_{\theta,1}(1) \\ W_{\theta,2}(1) \\ \vdots \\ W_{\theta,L}(1) \\ I_0(1)+Z'_0(1) \\ \vdots \\ I_{X+M-1}(1)+Z'_{X+M-1}(1) \end{bmatrix}, \quad X(2)=\begin{bmatrix} W_{\theta,1}(2) \\ W_{\theta,2}(2) \\ \vdots \\ W_{\theta,L}(2) \\ I_0(2)+Z'_0(2) \\ \vdots \\ I_{X+M-1}(2)+Z'_{X+M-1}(2) \end{bmatrix}. \tag{54}$$

Afterwards, the databases modify this answer vector in the same manner as (45).

Next, the databases send the new answer over the channel defined in (50) as follows,

$$\begin{aligned} y &= G'(u,v)\tilde{A} && (55)\\ &= G'(u,v)\begin{bmatrix} H_N^u & 0 \\ 0 & H_N^v \end{bmatrix}\begin{bmatrix} X(1) \\ X(2) \end{bmatrix} && (56)\\ &= \begin{bmatrix} (V_N(b))_{\mathcal{I}_1,\mathcal{I}_1} & 0_{L\times\lceil N/2\rceil} & (V_N(b))_{\mathcal{I}_1,\mathcal{I}_3} & (V_N(b))_{\mathcal{I}_1,\mathcal{I}_2} & 0_{L\times\lfloor\frac{N}{2}\rfloor} & (V_N(b))_{\mathcal{I}_1,\mathcal{I}_4} \\ (V_N(b))_{\mathcal{I}_3,\mathcal{I}_1} & 0_{L\times\lceil N/2\rceil} & (V_N(b))_{\mathcal{I}_3,\mathcal{I}_3} & (V_N(b))_{\mathcal{I}_3,\mathcal{I}_2} & 0_{L\times\lfloor\frac{N}{2}\rfloor} & (V_N(b))_{\mathcal{I}_3,\mathcal{I}_4} \\ (V_N(b))_{\mathcal{I}_2,\mathcal{I}_1} & 0_{L\times\lceil N/2\rceil} & (V_N(b))_{\mathcal{I}_2,\mathcal{I}_3} & (V_N(b))_{\mathcal{I}_2,\mathcal{I}_2} & 0_{L\times\lfloor\frac{N}{2}\rfloor} & (V_N(b))_{\mathcal{I}_2,\mathcal{I}_4} \\ (V_N(b))_{\mathcal{I}_4,\mathcal{I}_1} & 0_{L\times\lceil N/2\rceil} & (V_N(b))_{\mathcal{I}_4,\mathcal{I}_3} & (V_N(b))_{\mathcal{I}_4,\mathcal{I}_2} & 0_{L\times\lfloor\frac{N}{2}\rfloor} & (V_N(b))_{\mathcal{I}_4,\mathcal{I}_4} \end{bmatrix}\begin{bmatrix} X(1) \\ X(2) \end{bmatrix} && (57)\\ &= \Pi^t V_N(b)\Pi\left[W_{\theta,1}(1),\ldots,W_{\theta,L}(1),I'(1),W_{\theta,1}(2),\ldots,W_{\theta,L}(2),I'(2)\right]^t && (58) \end{aligned}$$

where $I'(1)$ is the last $\lfloor\frac{N}{2}\rfloor - L$ interference symbols of $X(1)$, $I'(2)$ is similarly the last $\lceil\frac{N}{2}\rceil - L$ interference symbols of $X(2)$. As both Π and $V_N(b)$ are invertible matrices, $\Pi^t V_N(b)\Pi$ could be inverted by the user to get $\{W_{\theta,1}(1),\ldots,W_{\theta,L}(1),W_{\theta,1}(2),\ldots,W_{\theta,L}(2)\}$, achieving a rate of $\frac{2L}{N}$ given in Theorem 3.

4.3.2 Case 2: $E > N - 2\min\{\frac{N}{2}, N-X-M\}$

In this case, we define a hybrid scheme that can use the interference symbols along with the noise variables introduced in the storage to prevent the eavesdropper from decoding the messages. As the number of eavesdropped links is greater than the number of interference symbols, we need to introduce new noise variables to compensate for the difference. Let $\delta = E - \big(N - 2\min\{\frac{N}{2}, N-X-M\}\big)$, $L_2 = \min\{\frac{N}{2}, N-X-M\}$, and $L_1 + \delta = L_2$. Define

$r_1, \ldots, r_\delta$ to be uniform random variables in $\mathbb{F}_q$. Then, the storage is defined as,

$$S_n(1) = \begin{bmatrix} r_1 + (f_1 - \alpha_n)R_{11} + (f_1 - \alpha_n)^2 R_{12} + \ldots + (f_1 - \alpha_n)^X R_{1X} \\ \vdots \\ r_\delta + (f_\delta - \alpha_n)R_{\delta 1} + (f_\delta - \alpha_n)^2 R_{\delta 2} + \ldots + (f_\delta - \alpha_n)^X R_{\delta X} \\ W_{\cdot,1} + (f_{\delta+1} - \alpha_n)R_{\delta+1,1} + (f_{\delta+1} - \alpha_n)^2 R_{\delta+1,2} + \ldots + (f_{\delta+1} - \alpha_n)^X R_{\delta+1,X} \\ W_{\cdot,2} + (f_{\delta+2} - \alpha_n)R_{\delta+2,1} + (f_{\delta+2} - \alpha_n)^2 R_{\delta+2,2} + \ldots + (f_{\delta+2} - \alpha_n)^X R_{\delta+2,X} \\ \vdots \\ W_{\cdot,L_1} + (f_{L_2} - \alpha_n)R_{L_1+\delta,1} + (f_{L_2} - \alpha_n)^2 R_{L_1+\delta,2} + \ldots + (f_{L_2} - \alpha_n)^X R_{L_1+\delta,X} \end{bmatrix}, \tag{59}$$

$$S_n(2) = \begin{bmatrix} W_{\cdot,L_1+1} + (f_1 - \alpha_n)R'_{11} + (f_1 - \alpha_n)^2 R'_{12} + \ldots + (f_1 - \alpha_n)^X R'_{1X} \\ W_{\cdot,L_1+2} + (f_2 - \alpha_n)R'_{21} + (f_2 - \alpha_n)^2 R'_{22} + \ldots + (f_2 - \alpha_n)^X R'_{2X} \\ \vdots \\ W_{\cdot,L_1+L_2} + (f_{L_2} - \alpha_n)R'_{L_2,1} + (f_{L_2} - \alpha_n)^2 R'_{L_2,2} + \ldots + (f_{L_2} - \alpha_n)^X R'_{L_2,X} \end{bmatrix}. \tag{60}$$

Then, we use the same scheme structure as in Section 4.3.1 with the same quantum transition matrix. The received answers at the user side are given by,

$$y = G'(u,v)\tilde{A} \tag{61}$$

$$= G'(u,v) \begin{bmatrix} H_N^u & 0 \\ 0 & H_N^v \end{bmatrix} \begin{bmatrix} X(1) \\ X(2) \end{bmatrix} \tag{62}$$

$$= \Pi^t V_N(b) \Pi \begin{bmatrix} I_L & 0_{L\times\lceil N/2\rceil} & 0 & 0 & 0 & 0 \\ 0 & 0 & I_{\lfloor N/2\rfloor - L} & 0 & 0 & 0 \\ 0 & 0 & 0 & I_L & 0_{L\times\lfloor N/2\rfloor} & 0 \\ 0 & 0 & 0 & 0 & 0 & I_{\lceil N/2\rceil - L} \end{bmatrix} \begin{bmatrix} X(1) \\ X(2) \end{bmatrix} \tag{63}$$

$$= \Pi^t V_N(b) \Pi \left[r_1, \ldots, r_\delta, W_{\theta,1}, \ldots, W_{\theta,L_1}, I'(1), W_{\theta,L_1+1}, \ldots, W_{\theta,L_1+L_2}, I'(2) \right]^t \tag{64}$$

where $I'(1)$, and $I'(2)$ are the last $\lfloor \frac{N}{2} \rfloor - L_2$ interference symbols of $X(1)$, and the last $\lceil \frac{N}{2} \rceil - L_2$ interference symbols of $X(2)$, respectively. Thus, $L_1 + L_2$ symbols are retrieved out of N symbols received by the user, yielding a rate of $R_Q = \frac{L_1+L_2}{N} = \min\{1 - \frac{E}{N}, 2(1 - \frac{X+M+\frac{\delta}{2}}{N})\}$ given in Theorem 3.

5 Conclusion

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 security at the same rate as the state-of-the-art scheme that solves the

same problem without symmetric security. In the quantum variation, we pointed to how the eavesdroppers have better access to the transmitted answer strings due to over-the-air decodability imposed by the N-sum box abstraction. To that end, we designed a scheme that represses the over-the-air decodability while maintaining the super-dense coding gain, i.e., doubling the rate compared to the classical variation. In addition, we designed another scheme that is more efficient in terms of the uploads. However, this scheme achieves the super-dense coding gain only in certain cases.

6 Proofs

In this section, we provide the required proofs to show that the schemes for XSETSPIR, QXSETSPIR presented in Sections 4.1, 4.2, 4.3 provide user-privacy against databases and eavesdroppers, database-privacy against the user, and security against communicating databases. To simplify the notations, we use Z to refer to the noise terms generated by the user for query generation, Z' to denote the noise terms generated by the databases to ensure their privacy, and R to denote the noise vectors generated to hide the messages at the databases.

Lemma 1 *The XSETSPIR scheme presented in Section 4.1 ensures privacy against any T-colluding databases.*

Proof: Recall that $T \leq \max(T, E) = M$, and let θ be the required user index. Then, for any set of T colluding servers denoted by $\mathcal{T}$, the collective observations corresponding to the ℓth bit of the received queries, i.e., rows $(\ell-1)K+1$ to $K\ell$ of each of the queries $Q_{\mathcal{T}}^{[\theta]}$, can be written as,

$$\begin{bmatrix} Q_{i_1}^t((\ell-1)K+1:K\ell) \\ \vdots \\ Q_{i_T}^t((\ell-1)K+1:K\ell) \end{bmatrix} = E_{\theta,\ell} + B_\ell \begin{bmatrix} Z_{\ell 1}^t \\ \vdots \\ Z_{\ell M}^t \end{bmatrix}, \tag{65}$$

where

$$E_{\theta,\ell} = \begin{bmatrix} \frac{1}{f_\ell - \alpha_n} e_\theta^t \\ \vdots \\ \frac{1}{f_\ell - \alpha_n} e_\theta^t \end{bmatrix}, \quad B_\ell = \begin{bmatrix} 1 & (f_\ell - \alpha_{i_1}) & \dots & (f_i - \alpha_{i_1})^{M-1} \\ \vdots & \vdots & \dots & \vdots \\ 1 & (f_\ell - \alpha_{i_T}) & \dots & (f_i - \alpha_{i_T})^{M-1} \end{bmatrix}. \tag{66}$$

Note that, all columns of B_ℓ are linearly independent. Now, we proceed as follows to ensure the privacy of the required message index θ,

$$I(\theta; Q_{\mathcal{T}}^{[\theta]}) = I(\theta; \{E_{\theta,\ell} + B_\ell Z_\ell\}_{\ell \in [L]}) \tag{67}$$

where $Z_\ell = \begin{bmatrix} Z_{\ell 1}^t \\ \vdots \\ Z_{\ell M}^t \end{bmatrix}$. As $T \leq M$ each term satisfies $I(\theta; E_{\theta,\ell} + B_\ell Z_\ell) = 0$, and all pairs of $\{E_{\theta,\ell} + B_\ell Z_\ell\}, \{E_{\theta,\ell'} + B'_\ell Z'_\ell\}$ for any $\ell \neq \ell'$ are mutually independent from Shannon's one-time-pad theorem, proving $I(\theta; Q_{\mathcal{T}}^{[\theta]}) = 0$. ■

Lemma 2 *The XSETSPIR scheme presented in Section 4.1 provides database privacy.*

Proof: The single-symbol answer received by the user from database n, $n \in [1:N]$ given in (29) is simply a random noise symbol, based on Shannon's one-time-pad theorem, as Z' terms are random noise symbols unknown to the user. Let $\hat{Z}_i$ be defined as $\hat{Z}_i = I_i + Z'_i$ in (29), and $\mathcal{W}_{\theta^C}$ is the set of all messages aside for the required message W_θ. Note that,

$$H(A_{[1:N]}^{[\theta]}) \leq H(W_\theta, \hat{Z}_0, \ldots, \hat{Z}_{X+M-1}), \tag{68}$$

as each $A_n^{[\theta]}$ is a function of W_θ, $\{\hat{Z}_i\}_{i=1}^{X+M-1}$ and the f, α constants. Therefore,

$$\begin{aligned} I(\mathcal{W}_{\theta^C}; A_{[1:N]}^{[\theta]} | W_\theta, Q_{[1:N]}^{[\theta]}, \theta) &= H(\mathcal{W}_{\theta^C} | W_\theta, Q_{[1:N]}^{[\theta]}, \theta) - H(\mathcal{W}_{\theta^C} | A_{[1:N]}^{[\theta]}, W_\theta, Q_{[1:N]}^{[\theta]}, \theta) && (69) \\ &\leq H(\mathcal{W}_{\theta^C}) - H(\mathcal{W}_{\theta^C} | W_\theta, \hat{Z}_0, \ldots, \hat{Z}_{X+M-1}, Q_{[1:N]}^{[\theta]}, \theta) && (70) \\ &= 0, && (71) \end{aligned}$$

proving the database privacy. ■

Lemma 3 *The XSETSPIR scheme presented in Section 4.1 is private against any eavesdropper with access to any E queries and E answers. In addition, it is secure against eavesdroppers with E answer strings and E queries.*

Proof: The first part of the lemma is proven as follows,

$$\begin{aligned} I(\theta; Q_{\mathcal{E}_1}^{[\theta]}, A_{\mathcal{E}_2}^{[\theta]}) &\leq I(\theta; Q_{\mathcal{E}_1}^{[\theta]}, A_{\mathcal{E}_1}^{[\theta]}) && (72) \\ &= I(\theta; Q_{\mathcal{E}_1}^{[\theta]}) + I(\theta, Q_{\mathcal{E}_1}^{[\theta]}; A_{\mathcal{E}_1}^{[\theta]}) - I(A_{\mathcal{E}_1}; Q_{\mathcal{E}_1}^{[\theta]}) && (73) \\ &= 0 + I(\theta, Q_{\mathcal{E}_1}^{[\theta]}; A_{\mathcal{E}_1}^{[\theta]}) - 0 && (74) \\ &= H(A_{\mathcal{E}_1}^{[\theta]}) - H(A_{\mathcal{E}_1}^{[\theta]} | \theta, Q_{\mathcal{E}_1}^{[\theta]}) && (75) \\ &= 0, && (76) \end{aligned}$$

where the first inequality follows from the fact that $Q_{\mathcal{E}_1}^{[\theta]}$, and $A_{\mathcal{E}_2}^{[\theta]}$ can convey the highest information of θ only when $\mathcal{E}_1 = \mathcal{E}_2$ since $A_{\mathcal{E}_1}^{[\theta]}$ is a function of $Q_{\mathcal{E}_1}^{[\theta]}$. Then, (74) follows from the proof of Lemma 1 since $|\mathcal{E}| \leq M$ and $I(A_{\mathcal{E}_1}^{[\theta]}; Q_{\mathcal{E}_1}^{[\theta]}) = 0$ since each answer depends on the uniform random variables $Z'_0, \ldots, Z'_{X+M-1}$ which are independent from the queries.

To prove the second part, we proceed as follows,

$$
\begin{aligned}
I(W_{[1:K]}; A_{\mathcal{E}_1}^{[\theta]}|Q_{\mathcal{E}_2}^{[\theta]}) &\leq I(W_{[1:K]}; A_{\mathcal{E}_1}^{[\theta]}|Q_{\mathcal{E}_1}^{[\theta]}) && (77)\\
&= H(W_{[1:K]}|Q_{\mathcal{E}_1}^{[\theta]}) - H(W_{[1:K]}|Q_{\mathcal{E}_1}^{[\theta]}, A_{\mathcal{E}_1}^{[\theta]}) && (78)\\
&= H(W_{[1:K]}) - H(W_{[1:K]}|Z_{\mathcal{E}_1}, Z'_{\mathcal{E}_1}) && (79)\\
&= H(W_{[1:K]}) - H(W_{[1:K]}) = 0, && (80)
\end{aligned}
$$

completing the proof. ■

Remark 14 *The proof of security against any X-communicating databases is the same as in [9] since we use the same storage construction.*

Next, to prove that the quantum schemes presented in Sections 4.2 and 4.3 provide symmetric privacy and security against eavesdroppers, we need to use the following definitions adopted from [32].

Definition 3 (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,*

$$\rho_A = \sum_j p_j \left|\psi_j\right\rangle \left\langle\psi_j\right|, \tag{81}$$

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

Definition 4 (Von Neumann Entropy) *For the density matrix ρ, its Von Neumann entropy is defined as,*

$$S(\rho) = -tr(\rho \log \rho) = H(\Lambda), \tag{82}$$

where 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 5 (Quantum Relative Entropy) *The relative entropy between two density matrices ρ and σ is defined to be,*

$$D(\rho\|\sigma) = tr(\rho(\log \rho - \log \sigma)). \tag{83}$$

Definition 6 (Quantum Conditional Entropy and Mutual Information) *The conditional entropy of a quantum system A with respect to a system B is defined as,*

$$S(A|B) = S(A, B) - S(B), \tag{84}$$

and the corresponding mutual information between them is defined as,

$$S(A;B) = S(A) + S(B) - S(A,B) = S(A) - S(A|B) = S(B) - S(B|A). \tag{85}$$

In the following lemmas, we prove some important relations between general quantum density matrices. Those relations are required to prove the properties of the quantum scheme in Section 4.2.

Lemma 4 *Let A, B, and C be general quantum density matrices. Then, the following hold:*

1. *If ρ_A is rank-1, then $S(A) = 0$.*
2. *$S(A;B) = D(\rho_{AB}\|\rho_A \otimes \rho_B)$, where the density matrix of joint quantum system A,B is ρ_{AB}, $\rho_A \coloneqq tr_B(\rho_{AB})$, $\rho_B \coloneqq tr_A(\rho_{AB})$, and $tr_B(\rho_A \otimes \rho_B) = \rho_A tr(\rho_B)$.*
3. *If $S(A,B) = 0$, then $S(A) = S(B)$.*
4. *$|S(A) - S(B)| \leq S(A,B) \leq S(A) + S(B)$.*
5. *$S(A;B,C) \geq S(A;B)$.*

Proof: The statements are proved as follows:

1. The proof of this statement follows from the fact that for a rank-1 density matrix (pure states), there is one eigenvalue with value 1 and the remaining eigenvalues are 0, which gives Shannon entropy of 0.
2. To prove this, we proceed as follows,

$$\begin{aligned}
D(\rho_{AB}\|\rho_A \otimes \rho_B) =& tr(\rho_{AB}(\log \rho_{AB} - \log(\rho_A \otimes \rho_B))) && (86)\\
=& tr(\rho_{AB} \log \rho_{AB}) - tr(\rho_{AB} \log(\rho_A \otimes \rho_B)) && (87)\\
=& - S(A,B) - tr(\rho_{AB}(\log(\rho_A \otimes I_B) + log(I_A \otimes \rho_B))) && (88)\\
=& - S(A,B) - tr_A(tr_B(\rho_{AB}(\log \rho_A \otimes I_B)) \\
& - tr_B(tr_A(\rho_{AB}(I_A \otimes \log \rho_B))) && (89)\\
=& - S(A,B) - tr_A(\rho_A \log \rho_A) - tr_B(\rho_B \log \rho_B) && (90)\\
=& - S(A,B) + S(A) + S(B). && (91)
\end{aligned}$$

3. For a composite quantum system A, B with a pure state density, note that

$$\begin{aligned}
|\phi\rangle &= \sum_{ij} a_{ij} |i_A\rangle |j_B\rangle && (92)\\
&= \sum_{ijk} u_{ik} \epsilon_{kk} v_{kj} |i_A\rangle |j_B\rangle && (93)
\end{aligned}$$

$$= \sum_k \epsilon_{kk} \left|\psi_{k,A}\right\rangle \left|\psi_{k,B}\right\rangle, \tag{94}$$

where $|\psi_{k,A}\rangle = \sum_i u_{ik} |i_A\rangle$ and $|\psi_{k,B}\rangle = \sum_j v_{kj} |j_B\rangle$, by using singular value decomposition for the matrix with elements a_{ij} with unitary matrices u_{ik} and v_{kj}. Then,

$$\rho_A = tr_B \left(\sum_k \epsilon_{kk} \left|\psi_{k,A}\right\rangle \left|\psi_{k,B}\right\rangle \sum_j \epsilon_{jj} \left\langle\psi_{j,A}\right| \left\langle\psi_{j,B}\right| \right) \tag{95}$$

$$= \sum_{kj} \epsilon_{kk}\epsilon_{jj} \left|\psi_{k,A}\right\rangle \left\langle\psi_{j,A}\right| tr_B \left(\left|\psi_{k,B}\right\rangle \left\langle\psi_{j,B}\right|\right) \tag{96}$$

$$= \sum_{kj} \epsilon_{kk}\epsilon_{jj} \left|\psi_{k,A}\right\rangle \left\langle\psi_{j,A}\right| \delta_{jk} \tag{97}$$

$$= \sum_k \epsilon_{kk}^2 \left|\psi_{k,A}\right\rangle \left\langle\psi_{k,A}\right|, \tag{98}$$

where $\delta_{ij} = 1$ if $i = j$ and 0, otherwise. Similarly, $\rho_B = \sum_k \epsilon_{kk}^2 |\psi_{k,B}\rangle\langle\psi_{k,B}|$. Note that ρ_A and ρ_B share the same eigenvalues, thus by definition, $S(A) = S(B)$.

4. The right hand side of the inequality follows from Klein's inequality for quantum relative entropy which states $D(\rho\|\sigma) \geq 0$ [32], Lemma 4 (item 2) and the definition of quantum mutual information. For the left hand side, note that by the purification theorem [32], there exists a system R such that $S(A, B, R) = 0$. Thus, from the right-hand side of the inequality, we can see that $S(A, R) \leq S(A) + S(R)$ and Lemma 4 (item 3) implies that $S(B) \leq S(A) + S(A, B)$. Similarly, $S(A) \leq S(B) + S(A, B)$.

5. $S(A; B, C) - S(A; B) = S(A) + S(B, C) - S(A, B, C) - S(A) - S(B) + S(A, B) = S(B, C) + S(A, B) - S(A, B, C) - S(B) \geq 0$, where the last inequality follows from the strong subadditivity of the Von Neumann entropy [32].

■

Lemma 5 *If a quantum system A has rank-1 density matrix, then $S(A; B) = 0$.*

Proof: Note from Lemma 4 (item 1) and Lemma 4 (item 4), $S(B) \leq S(A, B) \leq S(B)$, so that $S(A; B) = S(A) + S(B) - S(A, B) = S(B) - S(B) = 0$. Similarly, $S(A; B, C) = 0$ for a pure state system A. ■

Remark 15 *From Lemma 5, we can easily see that a pure state system A with $\rho_A = |\phi\rangle\langle\phi|$ acts as a deterministic classical source sending ϕ.*

Now, we prove the properties of the QXSETSPIR scheme presented in Section 4.2.

Lemma 6 *The QXSETSPIR scheme presented in Section 4.2 ensures privacy against any T-colluding databases.*

Proof: We note that,

$$I(Q_{\mathcal{T}}^{[\theta]}, \Lambda_{\mathcal{T}}; \theta) = I(Q_{\mathcal{T}}^{[\theta]}; \theta) + I(\Lambda_{\mathcal{T}}; \theta | Q_{\mathcal{T}}^{[\theta]}) = 0 \tag{99}$$

where the first term equals 0 from the classical scheme for any $\mathcal{T}$ that is a subset of $[1:N]$ with $|\mathcal{T}| \leq T$, and the second term equals 0 from the fact that by construction, the masking variables for both instances, $\Lambda_{\mathcal{T}} = [\Lambda_{\mathcal{T}}(1), \Lambda_{\mathcal{T}}(2)]$, are independent of the queries $Q_{\mathcal{T}}^{[\theta]}$, and the index θ. ■

Lemma 7 *The QXSETSPIR scheme presented in Section 4.2 provides database privacy.*

Proof: We start with,

$$\begin{aligned} S(\mathcal{A}_{[1:N]}^{[\theta]}, Q_{[1:N]}^{[\theta]}, \theta, \Lambda_{[1:N]}; \mathcal{W}_{\theta^C}) =& S(Q_{[1:N]}^{[\theta]}, \theta, \Lambda_{[1:N]}; \mathcal{W}_{\theta^C}) + S(\mathcal{A}_{[1:N]}^{[\theta]}; Q_{[1:N]}^{[\theta]}, \theta, \Lambda_{[1:N]}, \mathcal{W}_{\theta^C}) \\ &- S(\mathcal{A}_{[1:N]}^{[\theta]}; Q_{[1:N]}^{[\theta]}, \theta, \Lambda_{[1:N]}) \qquad (100) \\ =& I(Q_{[1:N]}^{[\theta]}, \theta, \Lambda_{[1:N]}; \mathcal{W}_{\theta^C}) + S(\mathcal{A}_{[1:N]}^{[\theta]}; Q_{[1:N]}^{[\theta]}, \theta, \Lambda_{[1:N]}, \mathcal{W}_{\theta^C}) \\ &- S(\mathcal{A}_{[1:N]}^{[\theta]}; Q_{[1:N]}^{[\theta]}, \theta, \Lambda_{[1:N]}), \qquad (101) \end{aligned}$$

where $\Lambda_{[1:N]} = [\Lambda_{[1:N]}(1), \Lambda_{[1:N]}(2)]$. In the last equality, the fact that the answers $\mathcal{A}_n^{[\theta]},\ n \in [1:N]$ are the only quantum system in the scheme is used to reduce the Von Neumann entropy to Shannon entropy.

Now, note that from the N-sum box abstraction [29], at the end of the transmission the received answers form a pure state quantum system, thus using Lemma 5,

$$S(\mathcal{A}_{[1:N]}^{[\theta]}; Q_{[1:N]}^{[\theta]}, \theta, \Lambda_{[1:N]}, \mathcal{W}_{\theta^C}) = S(\mathcal{A}_{[1:N]}^{[\theta]}; Q_{[1:N]}^{[\theta]}, \theta, \Lambda_{[1:N]}) = 0, \tag{102}$$

and from (49), it is known that the quantum measurement of the answers will give $\{W_{\theta,[L]}(1), W_{\theta,[L]}(2), I'(1), I'(2)\}$, all of which are independent of $\mathcal{W}_{\theta^C}$.

Moreover, since the contents of the messages, $\mathcal{W} = W_{[1:K]}$, and the random vectors generated by the user are independent, we have $I(Q_{[1:N]}^{[\theta]}, \theta, \Lambda_{[1:N]}; \mathcal{W}_{\theta^C}) = 0$. Thus, the scheme achieves symmetric privacy. ■

Remark 16 *In Lemma 6, a stronger guarantee of privacy can be proven as*

$$I(Q_{\mathcal{T}}^{[\theta]}, \Lambda_{[1:N]}; \theta) = I(Q_{\mathcal{T}}^{[\theta]}; \theta) + I(\Lambda_{[1:N]}; \theta | Q_{\mathcal{T}}^{[\theta]}) = 0, \tag{103}$$

since $\Lambda_{[1:N]}$ is generated independent of the queries and the required message index.

Remark 17 *By the construction of the protocol, it is clear that Lemma 6 holds for M colluding databases as well. In addition, using the same steps we note that, the eavesdroppers with access to $E \leq M$ queries cannot deduce the index of the required message from the queries, i.e., $I(\theta; Q_{\mathcal{E}}^{[\theta]}) = 0,\ \mathcal{E} \subset [1:N],\ |\mathcal{E}| \leq M$.*

Lemma 8 *The QXSETSPIR scheme presented in Section 4.2 is private and secure against any eavesdropper with access to E queries and answer strings.*

Proof: To prove the privacy part, i.e., privacy, note that

$$S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}, Q_{\mathcal{E}_2}^{[\theta]}, \Lambda_{\mathcal{E}_3}; \theta) = S(Q_{\mathcal{E}_2}^{[\theta]}, \Lambda_{\mathcal{E}_3}; \theta) + S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}; \theta, Q_{\mathcal{E}_2}^{[\theta]}, \Lambda_{\mathcal{E}_3}) - S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}; Q_{\mathcal{E}_2}^{[\theta]}, \Lambda_{\mathcal{E}_3}). \tag{104}$$

Then, from Remark 17, Lemma 6 and the independence between the masking variables $\Lambda_{[1:N]}$ and queries $Q_{[1:N]}^{[\theta]}$, the first term is 0. Moreover, using Lemma 5, the last two terms are equal to 0 as well. Hence, eavesdroppers cannot learn the index, concluding the first part of the proof.

To prove the second part, i.e., security, note that from Lemma 4 (item 5) and Lemma 7, we have $S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}, Q_{\mathcal{E}_2}^{[\theta]}, \Lambda_{\mathcal{E}_3}; \mathcal{W}_{\theta^C}) = 0$. In addition, from Remark 15 and the structure of N-sum box abstraction [29], it is inferred that $S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}, Q_{\mathcal{E}_2}^{[\theta]}, \Lambda_{\mathcal{E}_3}; W_\theta | \mathcal{W}_{\theta^C}) = 0$ since the user cannot decode the indexed message without knowing L out of N masking variables $\lambda_1, \ldots, \lambda_N$, even if all other messages are to be known, by the construction of the scheme, and since $\Lambda_{\mathcal{E}_3}$ is generated by Cauchy-Vandermonde coding of $\lambda_1, \ldots, \lambda_N$, no $\lambda_{j_1}, \ldots, \lambda_{j_L}$ can be leaked since $E \leq N - L = X + M$ by using arguments similar to 1. Moreover,

$$S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}, Q_{\mathcal{E}_2}^{[\theta]}, \Lambda_{\mathcal{E}_3}; W_{[1:K]}) = S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}, Q_{\mathcal{E}_2}^{[\theta]}, \Lambda_{\mathcal{E}_3}; \mathcal{W}_{\theta^C}) + S(\mathcal{A}_{\mathcal{E}_1}, Q_{\mathcal{E}_2}, \Lambda_{\mathcal{E}_3}; W_\theta | \mathcal{W}_{\theta^C}), \tag{105}$$

where $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3 \subseteq [1:N]$ with $|\mathcal{E}_1| = |\mathcal{E}_2| = |\mathcal{E}_3| = E$. Then, using Lemma 4.5, it we see that $S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}, Q_{\mathcal{E}_2}^{[\theta]}, \Lambda_{\mathcal{E}_3}; W_{[1:K]}) = 0$. ■

Remark 18 *It is clear, based on the previous results, that the main structure of the schemes presented in Section 4.3 are private against any T-colluding databases. It is also secure against any X communicating databases and provides symmetric privacy.*

Next, we prove that the schemes in Section 4.3 are private and secure from the eavesdroppers. For this, we first show that the channel transition matrix defined in Section 4.3 is a feasible N-sum box transition matrix.

Lemma 9 *The channel transition matrix described for the QXSETSPIR schemes in Section 4.3 is a stabilizer-based N-sum box transfer matrix as described in Theorem 4.*

Proof: Note that for any $2N \times N$ SSO matrix G and an $N \times N$ invertible matrix Υ, $G\Upsilon$ is an SSO matrix as well, since by the invertibility of Υ, $G\Upsilon$ is rank-N, and $\Upsilon^t G^t J G \Upsilon = \Upsilon^t (G^t J G) \Upsilon = 0$. Note also that if $\begin{bmatrix} G & H \end{bmatrix}$ is invertible, then $\begin{bmatrix} G\Upsilon & H\Upsilon \end{bmatrix}$ is invertible since

$$\begin{bmatrix} G\Upsilon & H\Upsilon \end{bmatrix} = \begin{bmatrix} G & H \end{bmatrix} \begin{bmatrix} \Upsilon & 0 \\ 0 & \Upsilon \end{bmatrix}, \tag{106}$$

and $\begin{bmatrix} \Upsilon & 0 \\ 0 & \Upsilon \end{bmatrix}$ is invertible as Υ is invertible. Then, if $M_1 = \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} G & H \end{bmatrix}^{-1}$ is a stabilizer-based N-sum box transfer matrix, so is

$$M_2 = \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} G\Upsilon & H\Upsilon \end{bmatrix}^{-1} \tag{107}$$

$$= \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} \Upsilon & 0 \\ 0 & \Upsilon \end{bmatrix}^{-1} \begin{bmatrix} G & H \end{bmatrix}^{-1} \tag{108}$$

$$= \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} \Upsilon^{-1} & 0 \\ 0 & \Upsilon^{-1} \end{bmatrix} \begin{bmatrix} G & H \end{bmatrix}^{-1} \tag{109}$$

$$= \begin{bmatrix} 0 & \Upsilon^{-1} \end{bmatrix} \begin{bmatrix} G & H \end{bmatrix}^{-1} \tag{110}$$

$$= \Upsilon^{-1} \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} G & H \end{bmatrix}^{-1} = \Upsilon^{-1} M_1. \tag{111}$$

The channel transition matrix of Section 4.3 is the channel transition matrix of Section 4.2 multiplied on the left by an $N \times N$ matrix by description, and as shown above, if that matrix is invertible, the proof follows. Note that $N \times N$ matrix is written also as $\Pi^t V_N(b)\Pi$. Note that Π is a permutation matrix that changes the positions between $\{L+1, \ldots, L+\lfloor \frac{N}{2} \rfloor\}$ elements and $\{L+\lfloor \frac{N}{2} \rfloor+1, \ldots, 2L+\lfloor \frac{N}{2} \rfloor\}$ elements, thus it is invertible by definition, and $V_N(b)$ is a Vandermonde matrix with distinct elements in b, and has nonzero determinant. Thus, the $N \times N$ matrix of Section 4.3.1 is invertible, concluding the proof. ■

Lemma 10 *The QXSETSPIR scheme in Section 4.3.1 achieves privacy and security against an eavesdropper that has access to any E queries and answers.*

Proof: To prove the privacy part, we need to show that,

$$S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}, Q_{\mathcal{E}_2}^{[\theta]}; \theta) = 0. \tag{112}$$

which follows from steps similar to those in the proof of Lemma 8. To prove the second part, i.e., security, note that, $S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}, Q_{\mathcal{E}_2}^{[\theta]}; \mathcal{W}_{\theta^C})$ which follows from the database privacy of the scheme and Lemma 4.5. Afterwards, note that $S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}, Q_{\mathcal{E}_2}^{[\theta]}; W_\theta | \mathcal{W}_{\theta^C}) = S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}, Q_{\mathcal{E}_2}^{[\theta]}, \mathcal{W}_{\theta^C}; W_\theta)$ as the messages are independent from each other. Moreover,

$$S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}, Q_{\mathcal{E}_2}^{[\theta]}, \mathcal{W}_{\theta^C}; W_\theta) = S(Q_{\mathcal{E}_2}^{[\theta]}, \mathcal{W}_{\theta^C}; W_{[1:K]}) + S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}; W_{[1:K]} | Q_{\mathcal{E}_2}^{[\theta]}, \mathcal{W}_{\theta^C}), \tag{113}$$

where $\mathcal{E}_1, \mathcal{E}_2 \subseteq [1:N]$ with $|\mathcal{E}_1| = |\mathcal{E}_2| = E$. Note that the first term on the right- and side is 0 by construction. From Lemma 5, the second term is 0 as well. Moreover, note that $I(I'(i); \mathcal{W}_{\theta^C}) = 0$, $i = [1:2]$, by the database privacy, that is knowing the non-indexed messages does not help the eavesdropper in knowing the interference terms, i.e., $S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}; \mathcal{W}_{\theta^C}) = 0$. Also, $S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}; Q_{\mathcal{E}_2}^{[\theta]}) = 0$ as the answers are generated by multiplying the

queries with the storage, which is a uniform random variable and as bove, the queries are independent of the messages as well. Secondly, note from the scheme that

$$\mathcal{A}_{\mathcal{E}_1}^{[\theta]} = \begin{bmatrix} 1 & b_{\mathcal{E}_1(1)} & \dots & b_{\mathcal{E}_1(1)}^{N-1} \\ \vdots & \vdots & \dots & \vdots \\ 1 & b_{\mathcal{E}_1(E)} & \dots & b_{\mathcal{E}_1(E)}^{N-1} \end{bmatrix} \begin{bmatrix} W_{\theta,1}(1) \\ \vdots \\ W_{\theta,L}(1) \\ W_{\theta,1}(2) \\ \vdots \\ W_{\theta,L}(2) \\ I'(1) \\ I'(2) \end{bmatrix}, \tag{114}$$

thus, for any leakage of the indexed message from the answers to happen, there has to be $\begin{bmatrix} c_1 & \dots & c_E \end{bmatrix}$ such that the last $N-2L$ columns should be all 0, i.e., there should not be any noise term that is masking the messages. That is,

$$\begin{bmatrix} 0 & \dots & 0 \end{bmatrix} = \begin{bmatrix} c_1 & \dots & c_E \end{bmatrix} \begin{bmatrix} b_{\mathcal{E}_1(1)}^{2L} & \dots & b_{\mathcal{E}_1(1)}^{N-1} \\ \vdots & \dots & \vdots \\ b_{\mathcal{E}_1(E)}^{2L} & \dots & b_{\mathcal{E}_1(E)}^{N-1} \end{bmatrix} = \prod_{k=1}^{E} b_{\mathcal{E}_1(k)}^{2L} \begin{bmatrix} c_1 & \dots & c_E \end{bmatrix} \begin{bmatrix} 1 & \dots & b_{\mathcal{E}_1(1)}^{N-2L-1} \\ \vdots & \dots & \vdots \\ 1 & \dots & b_{\mathcal{E}_1(E)}^{N-2L-1} \end{bmatrix}. \tag{115}$$

As $N-2L \geq E$, the matrix is guaranteed to be rank-E by the Vandermonde structure, so that the equation is satisfied only if $c_1 = \dots = c_E = 0$. That is, there is no leakage to the eavesdropper about the indexed message. Then, we see that the answers do not disclose any information about the messages and queries, thus using Remark 15, we see that $S(\mathcal{A}_{\mathcal{E}_1}^{[\theta]}, Q_{\mathcal{E}_2}^{[\theta]}; W_{[1:K]}) = 0$. ■

Remark 19 *The only difference between the schemes presented in Sections 4.3.1, and 4.3.2 is that the number of interference symbols is less than the number of the eavesdropped links in the latter. That is why extra noise symbols are introduced in the storage to compensate for the difference. The proof of privacy and security against the eavesdroppers of the scheme in Section 4.3.2 follow similar steps to those in the proof of Lemma 10.*

References

[1] A. Aytekin, M. Nomeir, S. Vithana, and S. Ulukus. Quantum symmetric private information retrieval with secure storage and eavesdroppers. In *IEEE Globecom*, December 2023. Also available online at arXiv:2308.10883.

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

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

[4] H. Sun and S. A. Jafar. The capacity of symmetric private information retrieval. *IEEE Trans. Info. Theory*, 65(1):322–329, June 2018.

[5] 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.

[6] 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.

[7] Q. Wang, H. Sun, and M. Skoglund. The capacity of private information retrieval with eavesdroppers. *IEEE Transactions on Information Theory*, 65(5):3198–3214, December 2018.

[8] 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.

[9] 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.

[10] 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.

[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] 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.

[15] 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.

[16] 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*, June 2022.

[17] 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.

[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. Commun.*, 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] S. Song and M. Hayashi. Capacity of quantum private information retrieval with multiple servers. *IEEE Trans. Info. Theory*, 67(1):452–463, September 2021.

[23] S. Song and M. Hayashi. Capacity of quantum private information retrieval with colluding servers. *IEEE Transactions on Information Theory*, 67(8):5491–5508, May 2021.

[24] S. Song and M. Hayashi. Capacity of quantum symmetric private information retrieval with collusion of all but one of servers. *IEEE Jour. Sel. Areas Info. Theory*, 2(1):380–390, January 2021.

[25] 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. Sel. Areas Commun.*, 40(3):885–898, January 2022.

[26] 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.

[27] S. Song and M. Hayashi. Quantum private information retrieval for quantum messages. In *IEEE ISIT*, July 2021.

[28] 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.

[29] M. Allaix, Y. Lu, Y. Yao, T. Pllaha, C. Hollanti, and S. A. Jafar. N-sum box: An abstraction for linear computation over many-to-one quantum networks. Available online at arXiv: 2304.07561.

[30] Y. Yao and S. A. Jafar. The capacity of classical summation over a quantum MAC with arbitrarily distributed inputs and entanglements. Available online at arXiv: 2305.03122.

[31] Y. Lu, Y. Yao, and S. A. Jafar. On the capacity of secure K-user product computation over a quantum MAC. Available online at arXiv: 2305.20073.

[32] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.