Dual-Source Symmetric PIR without Data Replication or Shared Randomness

Rémi A. Chou
Department of Computer Science and Engineering
University of Texas at Arlington
Arlington, TX 76019
remi.chou@uta.edu

2024 IEEE International Symposium on Information Theory (ISIT) | 979-8-3503-8284-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISIT57864.2024.10619444

***Abstract*—Information-theoretically secure Symmetric Private Information Retrieval (SPIR) is known to be infeasible over noiseless channels with a single server. Previous solutions involved additional resources such as database replication, shared randomness, or noisy channels. This paper demonstrates that, using a noiseless multiple access channel, SPIR with information-theoretic security guarantees is feasible without shared randomness, a noisy channel, or data replication. Specifically, we leverage a noiseless binary adder channel and employ two non-colluding servers with independent content. Furthermore, we characterize the optimal file rates, i.e., the file lengths normalized by the number of channel uses, that can be transferred.**

I. Introduction

Consider a client who wishes to download a file from a server such that (i) their file selection is kept private from the server and (ii) the client does not learn any other information beyond the selected file. This setting is referred to as Symmetric Private Information Retrieval (SPIR). SPIR is also known as oblivious transfer [1], which is anterior to SPIR and a fundamental cryptographic building block that is sufficient to implement secure multiparty computation [2], [3].

Under information-theoretic security guarantees, which is the focus of this paper, it is well known, e.g., [4], [5], that SPIR between a client and a single server is infeasible over a noiseless communication channel. To overcome this impossibility result, two approaches have previously been considered. The first approach consists in considering additional resources at the client and server in the form of correlated randomness, which could, for instance, be obtained through a noisy channel between the client and the server. Specifically, for some classes of noisy channels, SPIR is known to be feasible under information-theoretic security, e.g., [5]–[8]. The second approach consists in replicating data in multiple servers and assuming that the servers share randomness, e.g., [9]–[13]. Note that with this second approach, a necessary assumption is that only a strict subset of servers can collude against the client, otherwise SPIR is infeasible as the setting reduces to the case of SPIR between a client and a single server over a noiseless channel.

This work was supported in part by NSF grant CCF-2047913 and CCF-2401373.

In this paper, we propose to perform SPIR under information-theoretic security guarantees without shared randomness, a noisy channel, or data replication. Instead, we leverage a noiseless binary adder channel and employ two non-colluding servers. Specifically, we consider SPIR between one client and two non-colluding servers. We assume the client wishes to obtain one file from each server such that (i) the file selection remains private from the servers, (ii) the unselected files remain unknown to the receiver, and (iii) one server does not learn anything about the content of the other server. As formally described in Section II, in our setting, the servers and the client can communicate over a noiseless channel and have access to a noiseless adder multiple-access channel, but no noisy channels nor pre-shared correlated randomness are available at the parties. While an information-theoretically secure SPIR is impossible if the client engages in independent protocols with each of the servers, we show that a multiuser protocol between the client and the two servers can enable information-theoretically secure SPIR with positive rates for both servers simultaneously. Additionally, we fully characterize the capacity region for this setting.

The remainder of the paper is organized as follows. We formally introduce the setting in Section II and state our main results in Section III. We prove our converse and achievability results in Section IV and V, respectively. Finally, we provide concluding remarks in Section VI.

II. Problem Statement

Notation: For $a, b \in \mathbb{R}$, define $[\![a, b]\!] \triangleq [\lfloor a \rfloor, \lceil b \rceil] \cap \mathbb{N}$. For any $z \in \{0, 1\}$, define $\bar{z} \triangleq 1 - z$. For a vector X^n of length n, for any set $\mathcal{S} \subseteq [\![1, n]\!]$, let $X^n[\mathcal{S}]$ denote the components of X^n whose indices are in $\mathcal{S}$.

Consider a binary adder multiple access channel (MAC) $(\mathcal{Y}, p_{Y|X_1X_2}, \mathcal{X}_1 \times \mathcal{X}_2)$ defined by

$$Y \triangleq X_1 + X_2, \tag{1}$$

where $Y \in \mathcal{Y} \triangleq \{0, 1, 2\}$ is the output and X_1, $X_2 \in \mathcal{X}_1 \triangleq \mathcal{X}_2 \triangleq \{0, 1\}$ are the inputs. In the following, we assume that all the participants in the protocol are honest-but-curious, i.e., strictly follow the protocol. The setting is summarized in Figure 1 and formally described in the following definitions.

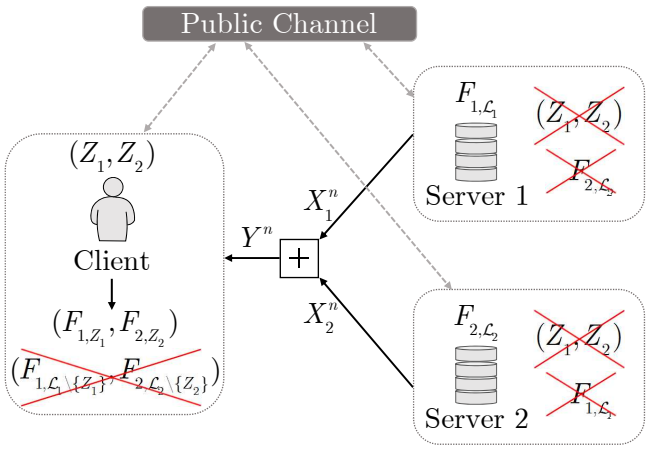

Fig. 1. Dual-source symmetric private information retrieval between one client and two servers who have access to a public channel and a binary adder MAC as described in (1). The file selection of the client is $(Z_1, Z_2) \in \mathcal{L}_1 \times \mathcal{L}_2$, i.e., the client wish to obtain (F_{1,Z_1}, F_{2,Z_2}) from the servers with the constraints that the client must not learn information about $(F_{1,\mathcal{L}_1\setminus\{Z_1\}}, F_{2,\mathcal{L}_2\setminus\{Z_2\}})$, (Z_1, Z_2) must remain private from both servers, and Server 1 must not learn information about Server 2's content, and vice versa for Server 2.

Definition 1. *An $(n, L_1, L_2, 2^{nR_1}, 2^{nR_2})$ dual-source SPIR protocol consists of*

- *Two non-colluding servers and one client;*
- *For $i \in \{1,2\}$, L_i independent random variables $(F_{i,l})_{l\in\mathcal{L}_i}$ uniformly distributed over $\{0,1\}^{nR_i}$, where $\mathcal{L}_i \triangleq [\![1, L_i]\!]$, which represent L_i files stored at Server i;*
- *U_0, U_1, U_2, three independent random variables, which represent local randomness available at the client, Server 1, and Server 2, respectively;*
- *Z_1, Z_2, two independent random variables uniformly distributed over $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively, which represent the client file choice for Servers 1, and 2, i.e., $(Z_1, Z_2) = (i, j)$ means that the client is requesting the files $(F_{1,i}, F_{2,j})$, where $(i, j) \in \mathcal{L}_1 \times \mathcal{L}_2$;*

and operates as follows from time $t = 1$ to time $t = n$,

- *The servers send $((X_1)_t, (X_2)_t) \in \mathcal{X}_1 \times \mathcal{X}_2$ over the binary adder MAC (1) and the client observes $Y_t \triangleq (X_1)_t + (X_2)_t$, where $(X_1)_t$ and $(X_2)_t$ are functions of $V_{1,t} \triangleq (F_{1,\mathcal{L}_1}, U_1, (A_{1,i})_{i\in[\![1,t-1]\!]})$ and $V_{2,t} \triangleq (F_{2,\mathcal{L}_2}, U_2, (A_{2,i})_{i\in[\![1,t-1]\!]})$, respectively;*
- *Next, the servers and the client are allowed to communicate, possibly interactively in r_t rounds of communication, over a noiseless channel. Specifically, for $j \in [\![1, r_t]\!]$, Servers 1 and 2 form and publicly send the messages $M_{1,t}(j)$ and $M_{2,t}(j)$, respectively, which are functions of $(V_{1,t}, (M_{0,1,t}(m))_{m\in[\![1,j-1]\!]})$ and $(V_{2,t}, (M_{0,2,t}(m))_{m\in[\![1,j-1]\!]})$, respectively. The client forms and publicly sends the messages $M_{0,1,t}(j)$ and $M_{0,2,t}(j)$, which are functions of $(Z_1, U_0, (A_{1,i})_{i\in[\![1,t-1]\!]}, Y^t, (M_{1,t}(m))_{m\in[\![1,j]\!]})$ and $(Z_2, U_0, (A_{2,i})_{i\in[\![1,t-1]\!]}, Y^t, (M_{2,t}(m))_{m\in[\![1,j]\!]})$, respectively. Let $A_{1,t} \triangleq (M_{0,1,t}(j), M_{1,t}(j))_{j\in[\![1,r_t]\!]}$ and $A_{2,t} \triangleq (M_{0,2,t}(j), M_{2,t}(j))_{j\in[\![1,r_t]\!]}$ represent all the messages publicly exchanged between the t-th and $t+1$-th channel use.*

Define the entire public communication by $\mathbf{A} \triangleq (\mathbf{A}_1, \mathbf{A}_2)$ with $\mathbf{A}_1 \triangleq (A_{1,t})_{t\in[\![1,n]\!]}$ and $\mathbf{A}_2 \triangleq (A_{2,t})_{t\in[\![1,n]\!]}$. Finally, the client forms $\hat{F}_{1,Z_1}$, an estimate of F_{1,Z_1}, from $(Y^n, U_0, \mathbf{A}_1)$, and $\hat{F}_{2,Z_2}$, an estimate of F_{2,Z_2}, from $(Y^n, U_0, \mathbf{A}_2)$.

Definition 2. *A rate pair (R_1, R_2) is achievable if there exists a sequence of $(n, L_1, L_2, 2^{nR_1}, 2^{nR_2})$ dual-source SPIR protocols such that*

$$\lim_{n\to\infty} \mathbb{P}\left[(\widehat{F}_{1,Z_1}, \widehat{F}_{2,Z_2}) \neq (F_{1,Z_1}, F_{2,Z_2})\right] = 0, \quad (2)$$

$$\lim_{n\to\infty} I\left(F_{1,\mathcal{L}_1} X_1^n U_1 \mathbf{A}; Z_1 Z_2\right) = 0, \quad (3)$$

$$\lim_{n\to\infty} I\left(F_{2,\mathcal{L}_2} X_2^n U_2 \mathbf{A}; Z_1 Z_2\right) = 0, \quad (4)$$

$$\lim_{n\to\infty} I(F_{1,\mathcal{L}_1} X_1^n U_1 \mathbf{A}; F_{2,\mathcal{L}_2}) = 0, \quad (5)$$

$$\lim_{n\to\infty} I(F_{2,\mathcal{L}_2} X_2^n U_2 \mathbf{A}; F_{1,\mathcal{L}_1}) = 0, \quad (6)$$

$$\lim_{n\to\infty} I\left(Z_1 Z_2 Y^n U_0 \mathbf{A}; F_{1,\mathcal{L}_1\setminus\{Z_1\}} F_{2,\mathcal{L}_2\setminus\{Z_2\}}\right) = 0. \quad (7)$$

The set of all achievable rate pairs is called the dual-source SPIR capacity region and is denoted by $\mathcal{C}_{\text{SPIR}^2}(L_1, L_2)$.

Equation (2) ensures that the client obtains the selected files. Equations (3) and (4) ensure the client's privacy by keeping the file selection (Z_1, Z_2) private from Server 1 and Server 2, respectively. Note that $(F_{i,\mathcal{L}_i}, X_i^n, U_i, \mathbf{A})$ represents all the information available at Server $i \in \{1,2\}$ at the end of the protocol. Equation (5) (respectively Equation (6)) ensures Server 2's (respectively Server 1's) privacy with respect to Server 1's (respectively Server 2's). Equation (7) ensures the servers' privacy by keeping all the non-selected files private from the client. Note that $(Z_1, Z_2, Y^n, U_0, \mathbf{A})$ represents all the information available at the client at the end of the protocol

Note that, if the servers are colluding, then no positive rates are achievable, as the setting reduces to an SPIR setting between one client and one server over a noiseless channel, for which it is known that information-theoretic security cannot be achieved, e.g., [10].

III. Main Results

Our main results provide a full characterization of the dual-source SPIR capacity region.

Theorem 1. *The dual-source SPIR capacity region is*

$$\mathcal{C}_{\text{SPIR}^2}(L_1, L_2) = \left\{(R_1, R_2) : (L_1 - 1)R_1 + (L_2 - 1)R_2 \leqslant \frac{1}{2}\right\}.$$

Moreover, any rate pair in $\mathcal{C}_{\text{SPIR}^2}(L_1, L_2)$ is achievable without time-sharing.

Proof. The converse and achievability are proved in Sections IV and V, respectively. ∎

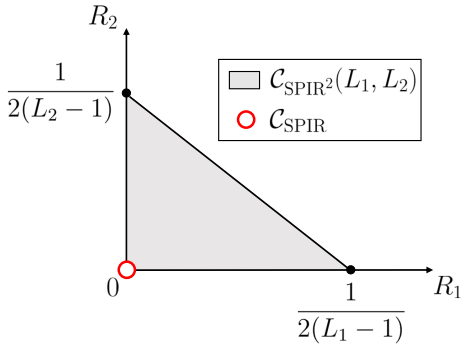

Fig. 2. The x-coordinate (resp. y-coordinate) of $\mathcal{C}_{\text{SPIR}}$ is defined as the SPIR capacity between Server 1 (resp. Server 2) and client in the absence of Server 2 (resp. Server 1), i.e., $\mathcal{C}_{\text{SPIR}} = (0,0)$. $\mathcal{C}_{\text{SPIR}^2}(L_1,L_2)$ is the dual-source SPIR capacity region.

Corollary 1. *If Servers* 1 *and* 2 *have the same number of files, i.e.,* $L_1 = L = L_2$*, then*

$$\mathcal{C}_{\text{SPIR}^2}(L) = \left\{(R_1,R_2) : R_1 + R_2 \leqslant \frac{1}{2(L-1)}\right\}.$$

Moreover, if $L_1 = L = L_2$ *and the files on both servers have the same size, i.e,* $nR_1 = nR = nR_2$*, then*

$$\mathcal{C}_{\text{SPIR}^2}(L) = \left\{(R,R) : R \leqslant \frac{1}{4(L-1)}\right\}.$$

The results contrast with the case where the client performs independent SPIR with each of the two servers, since in this case it is known that information-theoretic SPIR is impossible in the absence of additional resources such as shared randomness, a noisy channel, or data replication. We illustrate this point in Figure 2.

Note that when $L_1 = L_2 = 2$ and when the constraints (5) and (6) are ignored, the capacity region of the associated problem had been determined in our previous work [14].

IV. Converse part of Theorem 1

We first establish the following outer bound on the dual-source SPIR capacity region.

Proposition 1. *The dual-source SPIR capacity region is such that*

$$\mathcal{C}_{\text{SPIR}^2}(L_1,L_2) \subseteq \left\{(R_1,R_2) : (L_1-1)R_1 + (L_2-1)R_2 \leqslant \max_{p_{X_1}p_{X_2}} H(X_1X_2|Y)\right\}.$$

Proof. Consider a sequence of $(n, L_1, L_2, 2^{nR_1}, 2^{nR_2})$ dual-source SPIR protocols that achieve the rate pair (R_1,R_2). We prove the outer bound through a series of lemma.

Lemma 1. *For any* $z_1, z_2 \in \mathcal{L}_1 \times \mathcal{L}_2$*, we have*

$$I(F_{1,z_1}; Y^n U_0 | X_1^n \mathbf{A}_1, Z_1 = z_1, Z_2 = z_2) = 0, \quad (8)$$
$$I(F_{2,z_2}; Y^n U_0 | X_2^n \mathbf{A}_2, Z_1 = z_1, Z_2 = z_2) = 0. \quad (9)$$

Proof. It is sufficient to prove (8), as the proof of (9) can be obtained by exchanging the roles of the servers. Define for $t \in [\![1,n]\!]$, $i \in [\![1,r_t]\!]$, $A_{1,t,1:i} \triangleq (M_{0,1,t}(j), M_{1,t}(j))_{j\in[\![1,i]\!]}$ the messages sent and received by Server 1 between the first and the i-th communication exchanges with the client that happen after the t-th channel use. Let $A_1^t \triangleq (A_{1,j,1:r_j})_{j\in[\![1,t]\!]}$ be all the messages sent and received by Server 1 before the $t+1$-th channel use. Let $t \in [\![1,n]\!]$ and $j \in [\![1,r_t]\!]$. For convenience, we also write $\mathbf{Z} \triangleq (Z_1, Z_2)$. Then, we have

$$\begin{aligned}
&I(F_{1,\mathcal{L}_1}U_1; Y^tU_0 | X_1^t A_1^{t-1} A_{1,t,1:j}\mathbf{Z}) \quad (10)\\
&= I(F_{1,\mathcal{L}_1}U_1; Y^tU_0 | X_1^t A_1^{t-1} A_{1,t,1:(j-1)} M_{0,1,t}(j) M_{1,t}(j)\mathbf{Z})\\
&\leqslant I(F_{1,\mathcal{L}_1}U_1; Y^tU_0 M_{0,1,t}(j) | X_1^t A_1^{t-1} A_{1,t,1:(j-1)} M_{1,t}(j)\mathbf{Z})\\
&\overset{(a)}{=} I(F_{1,\mathcal{L}_1}U_1; Y^tU_0 | X_1^t A_1^{t-1} A_{1,t,1:(j-1)} M_{1,t}(j)\mathbf{Z})\\
&\leqslant I(F_{1,\mathcal{L}_1}U_1 M_{1,t}(j); Y^tU_0 | X_1^t A_1^{t-1} A_{1,t,1:(j-1)}\mathbf{Z})\\
&\overset{(b)}{=} I(F_{1,\mathcal{L}_1}U_1; Y^tU_0 | X_1^t A_1^{t-1} A_{1,t,1:(j-1)}\mathbf{Z}) \quad (11)\\
&\overset{(c)}{\leqslant} I(F_{1,\mathcal{L}_1}U_1; Y^tU_0 | X_1^t A_1^{t-1} Z_1 Z_2)\\
&\overset{(d)}{=} I(F_{1,\mathcal{L}_1}U_1; Y^{t-1}U_0 | X_1^t A_1^{t-1}\mathbf{Z})\\
&\leqslant I(F_{1,\mathcal{L}_1}U_1 (X_1)_t; Y^{t-1}U_0 | X_1^{t-1} A_1^{t-1}\mathbf{Z})\\
&\overset{(e)}{=} I(F_{1,\mathcal{L}_1}U_1; Y^{t-1}U_0 | X_1^{t-1} A_1^{t-1}\mathbf{Z}), \quad (12)
\end{aligned}$$

where (a) holds by the chain rule and because by definition $M_{0,1,t}(j)$ is a function of $(Z_1, Z_2, U_0, Y^t, A_1^{t-1}, A_{1,t,1:(j-1)}, M_{1,t}(j))$, (b) holds by the chain rule and because $M_{1,t}(j)$ is a function of $(F_{1,\mathcal{L}_1}, U_1, A_1^{t-1}, A_{1,t,1:(j-1)})$, (c) holds by repeating $j-1$ times the steps between Equations (10) and (11), (d) holds because $(F_{1,\mathcal{L}_1}, U_1) - (Y^{t-1}, U_0, X_1^t, A_1^{t-1}, \mathbf{Z}) - Y_t$ forms a Markov chain, (e) holds by the chain rule and because $(X_1)_t$ is a function of $(F_{1,\mathcal{L}_1}, U_1, A_1^{t-1})$. Then, for any $t \in [\![1,n]\!]$, we have

$$\begin{aligned}
&I(F_{1,\mathcal{L}_1}U_1; Y^tU_0 | X_1^t A_1^t \mathbf{Z})\\
&\overset{(a)}{\leqslant} I(F_{1,\mathcal{L}_1}U_1; Y^{t-1}U_0 | X_1^{t-1} A_1^{t-1}\mathbf{Z}) \quad (13)\\
&\overset{(b)}{\leqslant} I(F_{1,\mathcal{L}_1}U_1; U_0 | \mathbf{Z})\\
&= 0, \quad (14)
\end{aligned}$$

where (a) holds by taking $j = r_t$ in (12), (b) holds by using $t-1$ times (13). Next, for any $z_1, z_2 \in \mathcal{L}_1 \times \mathcal{L}_2$, we have

$$\begin{aligned}
&I(F_{1,z_1}; Y^nU_0 | X_1^n \mathbf{A}_1, Z_1 = z_1, Z_2 = z_2)\\
&\leqslant I(F_{1,\mathcal{L}_1}U_1; Y^nU_0 | X_1^n \mathbf{A}_1, Z_1 = z_1, Z_2 = z_2)\\
&\leqslant (\mathbb{P}[(Z_1,Z_2) = (z_1,z_2)])^{-1} I(F_{1,\mathcal{L}_1}U_1; Y^nU_0 | X_1^n \mathbf{A}_1 \mathbf{Z})\\
&= 0,
\end{aligned}$$

where the last equality holds by taking $t = n$ in (14). ∎

Next, using Lemma 1, we prove Lemma 2.

Lemma 2. *We have*

$$H(F_{1,\mathcal{L}_1\setminus\{Z_1\}} | X_1^n \mathbf{A} Z_1 Z_2) = o(n), \quad (15)$$
$$H(F_{2,\mathcal{L}_1\setminus\{Z_2\}} | X_2^n \mathbf{A} Z_1 Z_2) = o(n). \quad (16)$$

Proof. It is sufficient to prove (15), as the proof of (16) can be obtained by exchanging the roles of the servers. For any $z_1, z_2 \in \mathcal{L}_1 \times \mathcal{L}_2$, we write $\mathcal{L}_1 \backslash \{z_1\} = \{\gamma_i : i \in [\![1, L_1 - 2]\!]\}$, and we have

$$\begin{aligned}
&H(F_{1,\mathcal{L}_1\backslash\{z_1\}}|X_1^n\mathbf{A}, Z_1 = z_1, Z_2 = z_2)\\
&\leqslant H(F_{1,\mathcal{L}_1\backslash\{z_1\}}|X_1^n\mathbf{A}_1, Z_1 = z_1, Z_2 = z_2)\\
&\overset{(a)}{\leqslant} H(F_{1,\mathcal{L}_1\backslash\{z_1\}}|X_1^n\mathbf{A}_1, Z_1 = \gamma_1, Z_2 = z_2)\\
&\quad + 6nR_1\sqrt{\ln 2}\sqrt{I(F_{1,\mathcal{L}_1\backslash\{z_1\}}X_1^n\mathbf{A}_1; Z_1Z_2)} + 1\\
&\overset{(b)}{\leqslant} H(F_{1,\mathcal{L}_1\backslash\{z_1\}}|X_1^n\mathbf{A}_1, Z_1 = \gamma_1, Z_2 = z_2) + o(n)\\
&\overset{(c)}{\leqslant} H(F_{1,\mathcal{L}_1\backslash\{z_1,\gamma_1\}}|X_1^n\mathbf{A}_1, Z_1 = \gamma_1, Z_2 = z_2)\\
&\quad + H(F_{1,\gamma_1}|X_1^n\mathbf{A}_1, Z_1 = \gamma_1, Z_2 = z_2) + o(n)\\
&\overset{(d)}{\leqslant} \sum_{i=1}^{L_1-1} H(F_{1,\gamma_i}|X_1^n\mathbf{A}_1, Z_1 = \gamma_i, Z_2 = z_2) + o(n)\\
&\overset{(e)}{=} \sum_{i=1}^{L_1-1} H(F_{1,\gamma_i}|Y^nU_0X_1^n\mathbf{A}_1, Z_1 = \gamma_i, Z_2 = z_2) + o(n)\\
&\overset{(f)}{\leqslant} \sum_{i=1}^{L_1-1} H(F_{1,\gamma_i}|Y^nU_0\mathbf{A}_1\widehat{F}_{1,\gamma_i}, Z_1 = \gamma_i, Z_2 = z_2) + o(n)\\
&\overset{(g)}{\leqslant} o(n),
\end{aligned} \tag{17}$$

where (a) holds by [8, Lemma 3], (b) holds by (3), (c) holds by the chain rule and because conditioning reduces entropy, (d) holds by repeating $L_1 - 2$ times the steps between (a) and (c), (e) holds by Lemma 1, (f) holds because for any $i \in [\![1, L_1 - 2]\!]$, $\widehat{F}_{1,\gamma_i}$ is a function of $(Y^n, U_0, \mathbf{A}_1)$, (g) holds by Fano's inequality and (2). Finally, we have

$$\begin{aligned}
&H(F_{1,\mathcal{L}_1\backslash\{Z_1\}}|X_1^n\mathbf{A}_1Z_1Z_2)\\
&= \sum_{z_1,z_2} \mathbb{P}[(Z_1, Z_2) = (z_1, z_2)]\\
&\qquad \times H(F_{1,\mathcal{L}_1\backslash\{Z_1\}}|X_1^n\mathbf{A}_1, Z_1 = z_1, Z_2 = z_2)\\
&= o(n),
\end{aligned}$$

where the last equality holds by (17).

∎

Next, using Lemma 2 we obtain the following lemma.

Lemma 3. *We have*

$$H(F_{1,\mathcal{L}_1\backslash\{Z_1\}}F_{2,\mathcal{L}_1\backslash\{Z_2\}}|Z_1Z_2) \leqslant H(X_1^nX_2^n|Y^n) + o(n).$$

Proof. We have

$$\begin{aligned}
&H(F_{1,\mathcal{L}_1\backslash\{Z_1\}}F_{2,\mathcal{L}_1\backslash\{Z_2\}}|Z_1Z_2)\\
&= H(F_{1,\mathcal{L}_1\backslash\{Z_1\}}F_{2,\mathcal{L}_1\backslash\{Z_2\}}|Y^n\mathbf{A}Z_1Z_2)\\
&\quad + I(F_{1,\mathcal{L}_1\backslash\{Z_1\}}F_{2,\mathcal{L}_1\backslash\{Z_2\}}; Y^n\mathbf{A}|Z_1Z_2)\\
&\leqslant H(F_{1,\mathcal{L}_1\backslash\{Z_1\}}F_{2,\mathcal{L}_1\backslash\{Z_2\}}|Y^n\mathbf{A}Z_1Z_2)\\
&\quad + I(F_{1,\mathcal{L}_1\backslash\{Z_1\}}F_{2,\mathcal{L}_1\backslash\{Z_2\}}; Y^n\mathbf{A}Z_1Z_2)\\
&\overset{(a)}{\leqslant} H(F_{1,\mathcal{L}_1\backslash\{Z_1\}}F_{2,\mathcal{L}_1\backslash\{Z_2\}}|Y^n\mathbf{A}Z_1Z_2) + o(n)\\
&\leqslant H(F_{1,\mathcal{L}_1\backslash\{Z_1\}}F_{2,\mathcal{L}_1\backslash\{Z_2\}}X_1^nX_2^n|Y^n\mathbf{A}Z_1Z_2) + o(n)\\
&= H(X_1^nX_2^n|Y^n\mathbf{A}Z_1Z_2)\\
&\quad + H(F_{1,\mathcal{L}_1\backslash\{Z_1\}}F_{2,\mathcal{L}_1\backslash\{Z_2\}}|X_1^nX_2^nY^n\mathbf{A}Z_1Z_2) + o(n)\\
&\overset{(b)}{\leqslant} H(X_1^nX_2^n|Y^n) + H(F_{1,\mathcal{L}_1\backslash\{Z_1\}}|X_1^n\mathbf{A}Z_1Z_2)\\
&\quad + H(F_{2,\mathcal{L}_1\backslash\{Z_2\}}|X_2^n\mathbf{A}Z_1Z_2) + o(n)\\
&\overset{(c)}{\leqslant} H(X_1^nX_2^n|Y^n) + o(n),
\end{aligned}$$

where (a) holds by (7), (b) holds by the chain rule and because conditioning reduces entropy, (c) holds by Lemma 2. ∎

Finally, we have

$$\begin{aligned}
&(L_1 - 1)nR_1 + (L_2 - 1)nR_2\\
&\overset{(a)}{=} H(F_{1,\mathcal{L}_1\backslash\{Z_1\}}F_{2,\mathcal{L}_1\backslash\{Z_2\}}|Z_1Z_2)\\
&\overset{(b)}{\leqslant} H(X_1^nX_2^n|Y^n) + o(n)\\
&\overset{(c)}{\leqslant} \sum_{t=1}^{n} H((X_1)_t(X_2)_t|Y_t) + o(n)\\
&\overset{(d)}{=} nH((X_1)_T(X_2)_T|Y_TT) + o(n)\\
&\leqslant nH((X_1)_T(X_2)_T|Y_T) + o(n)\\
&\overset{(e)}{\leqslant} n \max_{p_{X_1}p_{X_2}} H(X_1X_2|Y) + o(n),
\end{aligned}$$

where (a) holds by independence and uniformity of the files, (b) holds by Lemma 3, (c) holds by the chain rule and because conditioning reduces entropy, (d) holds by defining T as the uniform random variable over $[\![1, n]\!]$, and in (e) we have defined $Y \triangleq X_1 + X_2$. ∎

By Proposition 1, it is sufficient to [14, Lemma 5] to obtain the converse part of Theorem 1.

V. Achievability part of Theorem 1

For clarity of presentation, we focus on the case $L_1 = L_2$. Specifically, we present our coding scheme in Section V-B, which shows that the case $L_1 = L_2$ can be reduced to the special case $L_1 = L_2 = 2$, which we treat in Section V-A.

A. Special case $L_1 = L_2 = 2$

Consider the coding scheme in Algorithm 1. Since (2), (3), (4), (7) can be proved as in [14], we only have to prove (5) and (6). Next, we have

$$\begin{aligned}
&I\left(F_{1,1}F_{1,2}X_1^n\mathbf{A}; F_{2,1}F_{2,2}\right)\\
&\overset{(a)}{=} I(F_{1,1}F_{1,2}X_1^nM_{20}M_{21}S_0^{(1)}S_1^{(1)}S_0^{(2)}S_1^{(2)}; F_{2,1}F_{2,2})\\
&\overset{(b)}{=} 0,
\end{aligned} \tag{18}$$

where (a) holds because (M_{10}, M_{11}) is a function of $(X_1^n, F_{1,1}, F_{1,2}, S_0^{(1)}, S_1^{(1)})$, (b) holds by a modified version of the one-time pad lemma. By exchanging the roles of the servers we also have

$$I\left(F_{2,1}F_{2,2}X_2^n\mathbf{A}; F_{1,1}F_{1,2}\right) = 0.$$

Algorithm 1 Dual-source SPIR when $L_1 = L_2 = 2$

Require: $t < 1/2$ and $\alpha \in [0,1]$.

1: The servers use the channel (1) as follows:

a: Consider (X_1^n, X_2^n) distributed according to the uniform distribution over $\{0,1\}^{2n}$

b: Servers 1 and 2 send X_1^n and X_2^n, respectively, over the channel (1)

c: The client observes $Y^n \triangleq X_1^n + X_2^n$

2: Upon observation of Y^n, the client

a: Defines
$\mathcal{G} \triangleq \{i \in [\![1,n]\!] : Y_i \in \{0,2\}\}, \mathcal{B} \triangleq \{i \in [\![1,n]\!] : Y_i = 1\}$;

b: Defines $M \triangleq \min(|\mathcal{G}|, |\mathcal{B}|)$;

c: Constructs $\mathcal{G}_1, \mathcal{G}_2$ such that $\mathcal{G}_1 \cup \mathcal{G}_2 \subset \mathcal{G}$, $\mathcal{G}_1 \cap \mathcal{G}_2 = \emptyset$, and $(|\mathcal{G}_1|, |\mathcal{G}_2|) = (\alpha M, \bar{\alpha} M)$;

d: Constructs $\mathcal{B}_1, \mathcal{B}_2$ such that $\mathcal{B}_1 \cup \mathcal{B}_2 \subset \mathcal{B}$, $\mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset$, and $(|\mathcal{B}_1|, |\mathcal{B}_2|) = (\alpha M, \bar{\alpha} M)$;

e: Defines for $i \in \{1,2\}$

$$S_0^{(i)} \triangleq \begin{cases} \mathcal{G}_i & \text{if } Z_i = 0 \\ \mathcal{B}_i & \text{if } Z_i = 1 \end{cases}, \quad S_1^{(i)} \triangleq \begin{cases} \mathcal{B}_i & \text{if } Z_i = 0 \\ \mathcal{G}_i & \text{if } Z_i = 1 \end{cases}.$$

Note that when $Z_j = i$, $i \in \{0,1\}$, $j \in \{1,2\}$, the client can determine $X_j^n[S_i^{(j)}]$.

3: If

$$\left| \frac{|\mathcal{G}|}{n} - \frac{1}{2} \right| \leqslant n^{-t}, \tag{19}$$

then the client sends to Servers 1 and 2 the sets $S_0^{(1)}, S_1^{(1)}, S_0^{(2)}, S_1^{(2)}$, otherwise the client aborts the protocol.

4: The public communication of the servers is as follows:

a: Server 1 sends to the client (M_{10}, M_{11}) where

$$(M_{10}, M_{11}) \triangleq (X_1^n[S_0^{(1)}] \oplus F_{1,1}, X_1^n[S_1^{(1)}] \oplus F_{1,2}).$$

b: Server 2 sends to the client (M_{20}, M_{21}) where

$$(M_{20}, M_{21}) \triangleq (X_2^n[S_0^{(2)}] \oplus F_{2,1}, X_2^n[S_1^{(2)}] \oplus F_{2,2}).$$

5: The client obtains its file selection as follows:

a: If $Z_1 = i \in \{0,1\}$, then the client determines $X_1^n[S_i^{(1)}]$ and computes $M_{1i} \oplus X_1^n[S_i^{(1)}] = F_{1,i}$.

b: If $Z_2 = i \in \{0,1\}$, then the client determines $X_2^n[S_i^{(2)}]$ and computes $M_{2i} \oplus X_2^n[S_i^{(2)}] = F_{2,i}$.

B. Case $L_1 = L_2$

The idea is to construct a coding scheme for the general case by utilizing multiple times Algorithm 2 developed for the case $L_1 = L_2 = 2$. This reduction idea is well known in the context of oblivious transfer, e.g., [15]. Our coding scheme is described in Algorithm 2. The analysis of Algorithm 2 is omitted due to space constraints.

Algorithm 2 Dual-source SPIR when $L_1 = L_2$

Require: $L-2$ sequences $(S_{1,t})_{t \in [\![1,L-2]\!]}$ uniformly distributed over $\{0,1\}^{nR_1}$, $L-2$ sequences $(S_{2,t})_{t \in [\![1,L-2]\!]}$ uniformly distributed over $\{0,1\}^{nR_2}$, the file selection $(Z_1, Z_2) \in \mathcal{L}_1 \times \mathcal{L}_2$

1: Server $j \in [\![1,2]\!]$ forms $(C_{j,t})_{t \in [\![1,L-1]\!]}$ as follows:

$$\begin{aligned} (C_{j,1}[1], C_{j,1}[2]) &\triangleq (F_{j,1}, S_{j,1}) \\ (C_{j,t}[1], C_{j,t}[2]) &\triangleq (F_{j,t} \oplus S_{j,t-1}, S_{j,t-1} \oplus S_{j,t}) \\ (C_{j,L-1}[1], C_{j,L-1}[2]) &\triangleq (F_{j,L-1} \oplus S_{j,L-2}, S_{j,L-2} \oplus F_{j,L}) \end{aligned}$$

where $t \in [\![2, L-2]\!]$. Then, For $t \in [\![1, L-1]\!]$, define $C_{j,t} \triangleq (C_{j,t}[1], C_{j,t}[2])$.

2: The client forms $(Z_{j,t}, Z_{j,t})_{j \in [\![1,2]\!], t \in [\![1,L-1]\!]}$ as follows:

$$Z_{j,t} \triangleq 1 + \mathbb{1}\{t < Z_j\}, \forall t \in [\![1, L-1]\!], \forall j \in [\![1,2]\!]$$

3: **for** $t \in [\![1, L-1]\!]$ **do**

4: The client and the two servers perform the SPIR protocol in Algorithm 1 with the two sequences $(C_{j,t}[1], C_{j,t}[2])$ at Server $j \in [\![1,2]\!]$ and the selection $(Z_{1,t}, Z_{2,t})$ for the client.

The subscript t is used in the notation of the random variables $(X_{1,t}^n, X_{2,t}^n, Y_t^n, \mathbf{A}_t, Z_{1,t}, Z_{2,t})$ involved in this SPIR protocol.

5: **end for**

6: By Lines 2-5, the client can form for $j \in [\![1,2]\!]$

$$F_{j,Z_j} = \begin{cases} C_{j,Z_j}[1] \oplus \bigoplus_{t=1}^{Z_j - 1} C_{j,t}[2] & \text{if } Z_j < L \\ \bigoplus_{t=1}^{Z_j} C_{j,t}[2] & \text{if } Z_j = L \end{cases}$$

VI. Concluding remarks

We studied information-theoretically secure SPIR in the absence of shared randomness, a noisy channel, and data replication. Instead, we leveraged a noiseless binary adder channel and two non-colluding servers with independent content and characterized the capacity region for this setting. While we considered honest-but-curious parties, an open problem is to address malicious parties who might attempt to cheat.

References

[1] M. Rabin, "How to exchange secrets by oblivious transfer," *Technical Memo TR-81, Aiken Computation Laboratory, Harvard University*, 1981.

[2] O. Goldreich, S. Micali, and A. Wigderson, "How to play any mental game," in *Proceedings of the nineteenth annual ACM symposium on Theory of computing*. ACM, 1987, pp. 218–229.

[3] J. Kilian, "Founding crytpography on oblivious transfer," in *Proceedings of the twentieth annual ACM symposium on Theory of computing*, 1988, pp. 20–31.
[4] S. Wolf and J. Wullschleger, "Unconditionally secure multiparty computation from noisy resources," in *Security with Noisy Data*. Springer, 2007, pp. 127–139.
[5] C. Crépeau, "Efficient cryptographic protocols based on noisy channels," in *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 1997, pp. 306–317.
[6] C. Crépeau and J. Kilian, "Achieving oblivious transfer using weakened security assumptions," in *29th Annual Symposium on Foundations of Computer Science*. IEEE Computer Society, 1988, pp. 42–52.
[7] A. C. Nascimento and A. Winter, "On the oblivious-transfer capacity of noisy resources," *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2572–2581, 2008.
[8] R. Ahlswede and I. Csiszár, "On oblivious transfer capacity," in *Information Theory, Combinatorics, and Search Theory*. Springer, 2013, pp. 145–166.
[9] Y. Gertner, Y. Ishai, E. Kushilevitz, and T. Malkin, "Protecting data privacy in private information retrieval schemes," *Journal of Computer and System Sciences*, vol. 60, no. 3, pp. 592–629, 2000.
[10] H. Sun and S. A. Jafar, "The capacity of symmetric private information retrieval," *IEEE Transactions on Information Theory*, vol. 65, no. 1, pp. 322–329, 2018.
[11] Q. Wang and M. Skoglund, "Symmetric private information retrieval from MDS coded distributed storage with non-colluding and colluding servers," *IEEE Transactions on Information Theory*, vol. 65, no. 8, pp. 5160–5175, 2019.
[12] Q. Wang, H. Sun, and M. Skoglund, "The capacity of private information retrieval with eavesdroppers," *IEEE Transactions on Information Theory*, vol. 65, no. 5, pp. 3198–3214, 2018.
[13] Q. Wang and M. Skoglund, "On PIR and symmetric PIR from colluding databases with adversaries and eavesdroppers," *IEEE Transactions on Information Theory*, vol. 65, no. 5, pp. 3183–3197, 2018.
[14] R. A. Chou, "Pairwise oblivious transfer," *IEEE Information Theory Workshop*, 2020.
[15] G. Brassard, C. Crépeau, and M. Santha, "Oblivious transfers and intersecting codes," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 1769–1780, 1996.