On the Linear Capacity of Conditional Disclosure of Secrets

Zhou Li, *Member, IEEE*, and Hua Sun, *Member, IEEE*

***Abstract*— Conditional disclosure of secrets (CDS) is the problem of disclosing as efficiently as possible, one secret from Alice and Bob to Carol if and only if the inputs at Alice and Bob satisfy some function. The information theoretic capacity of CDS is the maximum number of bits of the secret that can be securely disclosed per bit of total communication from Alice and Bob to Carol. All CDS instances, where the capacity is the highest and is equal to $1/2$, are recently characterized through a noise and signal alignment approach and are described using a graph representation of the function. In this work, we go beyond the best case scenarios and further develop the alignment approach to characterize the linear capacity of a class of CDS instances to be $(\rho - 1)/(2\rho)$, where ρ is a newly introduced and highly specific covering parameter of the graph representation of the function.**

***Index Terms*— Conditional disclosure of secrets, linear capacity, noise and signal alignment.**

I. Introduction

THE conditional disclosure of secrets (CDS) problem is a classical cryptographic primitive with rich connections to many other primitives such as symmetric private information retrieval [2] and secret sharing [3], [4]. The goal of the CDS problem is to find the most efficient way for Alice and Bob to disclose a common secret to Carol if and only if the inputs at Alice and Bob satisfy some function f (see Fig. 1). The CDS problem was initially studied in the setting where the secret is one bit long, and the cost of a CDS scheme is measured by the worst case total amount of communication over all functions f, typically as order functions of the input size [2], [5], [6], [7], [8], [9]. That is, the focus is on the scaling law of the communication complexity as the input size grows to infinity. What is pursued in this work is the traditional Shannon theoretic formulation, where the secret size is allowed to be arbitrarily large, and the communication rate is the number of bits of the secret that can be securely disclosed per bit of total communication. The aim is to characterize the maximum rate, termed the capacity of CDS, for a fixed function f.

Manuscript received 24 May 2022; revised 9 February 2023 and 11 July 2023; accepted 14 September 2023. Date of publication 22 September 2023; date of current version 19 December 2023. This work was supported in part by funding from NSF grants CCF-2007108 and CCF-2045656. An earlier version of this paper was presented in part at the 2021 IEEE International Symposium on Information Theory (ISIT) [1]. The associate editor coordinating the review of this article and approving it for publication was S. Rini. *(Corresponding author: Zhou Li.)*

The authors are with the Department of Electrical Engineering, University of North Texas, Denton, TX 76203 USA (e-mail: zhouli@my.unt.edu; hua.sun@unt.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2023.3317914.

Digital Object Identifier 10.1109/TCOMM.2023.3317914

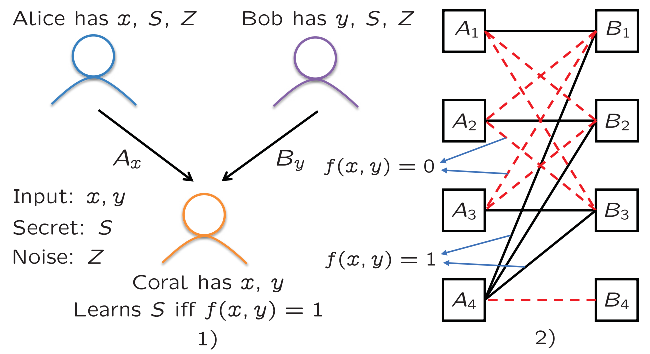


Fig. 1. 1). Alice and Bob (with secret S, noise variable Z, respective inputs x, y) wish to disclose the secret S to Carol if and only if $f(x, y) = 1$ for a binary function f, through signals A_x, B_y. 2) An example of $f(x, y)$ in graph representation. From pair of nodes connected by a solid black edge (i.e., $f(x, y) = 1$), Carol can decode S; from pair of nodes connected by a dashed red edge (i.e., $f(x, y) = 0$), Carol learns nothing about S in the information theoretic sense.

A. Motivation and Related Work

The CDS problem is introduced in the first paper that initiates the study of the classical symmetric private information retrieval problem [2] and is used as an auxiliary primitive to thwart unauthorized queries (which can be seen by viewing x, y as the queries sent to Alice and Bob, and if and only if the query is legal, i.e., $f(x, y) = 1$, the answers from Alice and Bob will reveal the secret S). Ever since, CDS has been widely studied in computer science and cryptography, as an interesting problem by itself [3], [4], [5], [6], [7], [8], [9]. Furthermore, the CDS problem is now recognized as a prominent open problem in information theoretic cryptography [10].

The broad interest in CDS also comes from its application to practical systems. One may equivalently think of CDS as a secure storage problem over graphs [11], [12]. For example, the CDS instance in Fig. 1.2 may represent a storage system with 8 servers (where each node is a server) that is used to store the data S securely (accessible by authorized pairs of servers) and only from a pair of servers connected by a solid back edge, one may decode S. As the graph (i.e., the function f) may have arbitrary topology, it can be used to model a wide class of data access patterns (from here we can also see the connection to access structure in secret sharing).

Shannon Theoretic Formulation: What differentiates our Shannon theoretic formulation from the information theoretic cryptography formulation is mainly the following two aspects. First, we are interested in the regime where the secret size may approach infinity (which is typical and classical for rate definition in information theory) while previous cryptography mostly fixes the secret to be 1 bit. The large data regime is also practical in current information age and complements existing studies. Second, we consider each fixed function f as we are looking for the exact capacity (so that solving all functions at once is formidable) while previous cryptography work aims at universal approximation results (i.e., targeting at most challenging f for worst case analysis) so that there protocols are strong as they apply to all functions while converses results (optimality) are hard to obtain.

B. Our Approach and Results

In [13], we obtain a complete characterization for all functions f where the CDS capacity is the highest, and is equal to $1/2$. In describing this result, we find it convenient to represent the function f by a bipartite graph, where each node denotes a possible signal for certain input and two types (colors) of edges are used to denote whether f is 1 or 0 (see Fig. 1.2 for a concrete example). We will use this graph representation of functions f throughout this work. The feasibility condition for capacity $1/2$ is then stated in terms of the graphic properties of f. Furthermore, this result is obtained using a novel noise and signal alignment approach, which guides the proof of both (information theoretic) impossibility claims and (linear) protocol designs.

Beyond the best rate scenarios, the simplest uncovered case is also considered in [13] (see Theorem 2), where the linear capacity[1] has been found and this is our starting point. Our goal in this work is to further develop the alignment approach to characterize the linear capacity of a larger class of CDS instances. As our first main result (see Theorem 1), we obtain a general converse bound for linear CDS schemes, which applies to any CDS instance, is parameterized by a covering parameter ρ of the graph representation of f, and is equal to $(\rho-1)/(2\rho)$. As our second main result (see Theorem 2), we show that the above converse bound is achievable for a class of graphs, i.e., CDS instances, through a vector linear code based achievable scheme with matching rate. While we find that the converse bound appears to be achievable for more graphs (by verifying a number of examples), an explicit condition of a larger class and a universal code design that applies generally remain elusive. As our final result, we show through an example that the above converse bound is not tight in general and we establish the linear capacity for that example (see Theorem 3). Interestingly, all results are obtained through a more refined view of the alignment approach.

Notation: The notation $|\mathcal{A}|$ is used to denote the cardinality of a set $\mathcal{A}$. The space spanned of the rows of a matrix A is denoted as $\mathcal{R}(A)$. For linear spaces $\mathcal{A}, \mathcal{B}$, $d(\mathcal{A})$ represents its dimension and $\mathcal{A}\cap\mathcal{B}$ denotes the intersection space. The rank of a matrix A is denoted as $r(A)$, which is equal to $d(\mathcal{R}(A))$. The notation $(\mathbf{A};\mathbf{B})$ denotes the row stack of matrices $\mathbf{A}$ and $\mathbf{B}$. For integers i, j where $i<j$, a vector $(a_i; a_{i+1}; \cdots; a_j)$ is abbreviated as $a_{[i:j]}$.

[1] It turns out that the linear capacity, i.e., the highest rate achievable by linear schemes (so non-linear schemes are excluded) [14], [15], does not match the best converse bound produced by only Shannon information inequalities, i.e., sub-modularity of entropy functions [13], which points to the possible gap between linear capacity and capacity and is an interesting open problem.

II. Problem Statement and Preliminaries

Consider a binary function $f(x, y)$, where (x, y) is from some set $\mathcal{I} \subset \{1, 2, \cdots, x_{max}\} \times \{1, 2, \cdots, y_{max}\}$ and its characteristic undirected bipartite graph $G_f = (V, E)$, where the node set $V = \{A_1, \cdots, A_{x_{max}}, B_1, \cdots, B_{y_{max}}\}$ and the edge set E is comprised of the unordered pairs $\{A_x, B_y\}$ such that $(x, y) \in \mathcal{I}$. The edges have two types: if $f(x, y) = 1$, $\{A_x, B_y\}$ is a solid black edge and is referred to as a *qualified edge*; if $f(x, y) = 0$, $\{A_x, B_y\}$ is a dashed red edge and is referred to as an *unqualified edge* (see Fig. 1.2 for an example).

The scalar x (y) denotes the input available only to Alice (Bob) and A_x (B_y) denotes the signal sent from Alice (Bob) to Carol for securely disclosing the secret S, which is comprised of L independent and identically distributed (i.i.d.) uniform symbols from a finite field $\mathbb{F}_p$. In addition to the secret S, Alice and Bob also hold an independent common noise variable Z (to assist with the secure disclosure task) that is comprised of L_Z i.i.d. uniform symbols from $\mathbb{F}_p$. In p-ary units,

$$H(S) = L,\ H(Z) = L_Z,$$
$$H(S, Z) = H(S) + H(Z) = L + L_Z. \tag{1}$$

Each signal A_x (B_y) is assumed to be comprised of N symbols from $\mathbb{F}_p$ and must be determined by information available to Alice (Bob).

$$H(A_x, B_y|S, Z) = 0. \tag{2}$$

The disclosure task is said to be successful if the following conditions are satisfied. From a qualified edge, Carol can recover S with no error; from an unqualified edge, Carol must learn nothing about S.

Definition 1 (Correctness and Security): For all $(x, y) \in \mathcal{I}$,

$$[\textit{Correctness}]\quad H(S|A_x, B_y) = 0,\ \textit{if } f(x, y) = 1; \tag{3}$$
$$[\textit{Security}]\ H(S|A_x, B_y) = H(S),\ \textit{otherwise } f(x, y) = 0. \tag{4}$$

The collection of the mappings from x, y, S, Z to A_x, B_y as specified above is called a CDS scheme.

The CDS rate[2] R characterizes how many symbols of the secret are securely disclosed per symbol of total communication and is defined as follows.

$$R = \frac{L}{2N}. \tag{5}$$

A rate R is said to be achievable if there exists a CDS scheme, for which the correctness and security constraints (3), (4) are satisfied and the rate is greater than or equal to R. The supremum of (asymptotic) achievable rates is called the capacity of CDS, C. Note that L is allowed to approach infinity, although our code construction will have finite block-length, i.e., L is a finite constant.

In this work, we focus mainly on the metric of capacity C and allow as much noise as needed, i.e., the randomness

[2] In this work, we focus exclusively on the symmetric rate where the signals from Alice and Bob are assumed to have the same size while leaving the generalizations to asymmetric rate and rate region as an interesting open research direction for future study.

size L_Z is unconstrained (as an elemental setting while noting that the randomness cost of CDS is an interesting problem by itself).

A. Graph Definitions

We will use some graphic notions of $G_f = (V, E)$ to state our results, which are defined as follows. Without loss of generality, we assume that for any node $v \in V$, there exists some node $u \in V$ such that $\{u, v\} \in E$ is an unqualified edge (otherwise, for any v that is connected to only qualified edges, we can set v to be the secret S and then eliminate v and its edges).

Definition 2 (Qualified/Unqualified Path/Component): A sequence of distinct connecting qualified (unqualified) edges is called a *qualified (unqualified) path*. A *qualified (unqualified) connected component* is a maximal induced subgraph of G_f such that any two nodes in the subgraph are connected by a qualified (unqualified) path.

For an example of an unqualified path, see $P = \{\{A_1, B_2\}, \{B_2, A_3\}, \{A_3, B_1\}\}$ in Fig. 1.2. As the graph in Fig. 1.2 except node B_4 is connected by qualified edges, it (i.e., the whole graph excluding node B_4 and edge $\{A_4, B_4\}$) is a qualified component.

Definition 3 (Internal Qualified Edge): A qualified edge that connects two nodes in an unqualified path is called an internal qualified edge*.*

For example, in Fig. 1.2, the edge $e = \{A_1, B_1\}$ is an internal qualified edge that connects the two nodes A_1, B_1 in the unqualified path $P = \{\{A_1, B_2\}, \{B_2, A_3\}, \{A_3, B_1\}\}$.

Definition 4 (Connected Edge Cover): Consider an internal qualified edge e in an unqualified path P and the node set of P is denoted as $V_P \subset V$. A connected edge cover *of V_P is a set of connected*[3] *qualified edges $M \subset E$ such that each node in V_P is covered by at least one qualified edge in M and $e \in M$. The size of a connected edge cover for (e, P) is the number of edges in M and is denoted as $\rho(e, P)$. If no such M exists, then $\rho(e, P)$ is defined as $+\infty$. Further, $\rho \triangleq \min_{e,P} \rho(e, P)$.*

For example, in Fig. 1.2, consider the internal qualified edge $e = \{A_1, B_1\}$ in the unqualified path $P = \{\{A_1, B_2\}, \{B_2, A_3\}, \{A_3, B_1\}\}$, then the nodes in P are $V_P = \{A_1, B_2, A_3, B_1\}$ and a connected edge cover of V_P is $M = \{\{A_4, B_1\}, \{A_4, B_2\}, \{A_4, B_3\}, \{A_1, B_1\}, \{A_3, B_3\}\}$. In this case, $\rho(e, P) = 5$ as M contains 5 edges and we can verify that the minimum value of $\rho(e, P)$ over all internal qualified edges and their associated unqualified path pairs (e, P) is $\rho = 5$.

It can be verified that in general, ρ can be any integer that is at least 5. Also note that as ρ is defined to be the minimum over all e, P, so the connected edge cover M that attains the value of ρ corresponds to one that has the minimal cardinality.

B. Linear Feasibility

We characterize the feasibility condition of a linear CDS scheme.

[3]That is, any two nodes in M are connected by a qualified path.

Linear Scheme: *For a feasible linear CDS scheme, each signal (equivalently, each node $v \in V$)*

$$v = \mathbf{F}_v S + \mathbf{H}_v Z, \ \mathbf{F}_v \in \mathbb{F}_p^{N \times L}, \mathbf{H}_v \in \mathbb{F}_p^{N \times L_Z} \quad (6)$$

is specified by two precoding matrices, $\mathbf{F}_v$ for the secret $S \in \mathbb{F}_p^{L \times 1}$ and $\mathbf{H}_v$ for[4] *the noise $Z \in \mathbb{F}_p^{L_Z \times 1}$ such that the following properties are satisfied.*

- *Consider any edge $\{v, u\}$ and identify the overlap of their noise spaces, i.e., the row space of $\mathbf{H}_v$ and $\mathbf{H}_u$. That is, find matrices $\mathbf{P}_v$ and $\mathbf{P}_u$ such that*[5]

$$\begin{aligned} &\mathbf{P}_v \mathbf{H}_v = \mathbf{P}_u \mathbf{H}_u, \\ &r(\mathbf{P}_v) = r(\mathbf{P}_u) = d(\mathcal{R}(\mathbf{H}_v) \cap \mathcal{R}(\mathbf{H}_u)), \end{aligned} \quad (7)$$

then the secret spaces satisfy

$$\begin{aligned} &[\textit{Correctness}] \quad r(\mathbf{P}_v \mathbf{F}_v - \mathbf{P}_u \mathbf{F}_u) = L, \\ &\qquad\qquad\qquad \textit{if } \{u, v\} \textit{ is qualified}; \quad (8) \\ &[\textit{Security}]^{6} \quad \mathbf{P}_v \mathbf{F}_v = \mathbf{P}_u \mathbf{F}_u, \\ &\qquad\qquad\qquad \textit{otherwise } \{u, v\} \textit{ is unqualified}. \quad (9) \end{aligned}$$

The correctness constraint (8) and the security constraint (9) for linear schemes imply the entropic versions (3), (4). For correctness, note that $\mathbf{P}_v v - \mathbf{P}_u u = (\mathbf{P}_v \mathbf{F}_v - \mathbf{P}_u \mathbf{F}_u) S$, so S can be decoded with no error if $\mathbf{P}_v \mathbf{F}_v - \mathbf{P}_u \mathbf{F}_u$ has full rank. For security, note that $\mathbf{P}_v \mathbf{H}_v, \mathbf{P}_u \mathbf{H}_u$ contains all the overlaps so that the remaining vectors are orthogonal. That is,

$$(\mathbf{H}_v; \mathbf{H}_u) \overset{\text{invertible}}{\longleftrightarrow} (\mathbf{P}_v \mathbf{H}_v; \mathbf{Q}_v \mathbf{H}_v; \mathbf{Q}_u \mathbf{H}_u) \quad (10)$$

where $\mathbf{Q}_v, \mathbf{Q}_u$ are chosen so that $\mathcal{R}(\mathbf{P}_v \mathbf{H}_v)$, $\mathcal{R}(\mathbf{Q}_v \mathbf{H}_v)$, $\mathcal{R}(\mathbf{Q}_u \mathbf{H}_u)$ are linearly independent (e.g., $\mathcal{R}(\mathbf{Q}_v \mathbf{H}_v)$ may be set as the orthogonal complement of $\mathcal{R}(\mathbf{P}_v \mathbf{H}_v)$ within $\mathcal{R}(\mathbf{H}_v)$ and $\mathbf{Q}_u$ is similar). Then we have

$$\begin{aligned} I(S; v, u) &\overset{(7)(9)}{=} I(S; \mathbf{P}_v \mathbf{F}_v S + \mathbf{P}_v \mathbf{H}_v Z, \mathbf{Q}_v \mathbf{F}_v S \\ &\quad + \mathbf{Q}_v \mathbf{H}_v Z, \mathbf{Q}_u \mathbf{F}_u S + \mathbf{Q}_u \mathbf{H}_u Z) \quad (11a) \\ &\overset{(1)}{=} H(\mathbf{P}_v \mathbf{F}_v S + \mathbf{P}_v \mathbf{H}_v Z, \\ &\quad \mathbf{Q}_v \mathbf{F}_v S + \mathbf{Q}_v \mathbf{H}_v Z, \mathbf{Q}_u \mathbf{F}_u S + \mathbf{Q}_u \mathbf{H}_u Z) \\ &\quad - H(\mathbf{P}_v \mathbf{H}_v Z, \mathbf{Q}_v \mathbf{H}_v Z, \mathbf{Q}_u \mathbf{H}_u Z) \quad (11b) \\ &\le r(\mathbf{P}_v \mathbf{H}_v; \mathbf{Q}_v \mathbf{H}_v; \mathbf{Q}_u \mathbf{H}_u) \\ &\quad - r(\mathbf{P}_v \mathbf{H}_v; \mathbf{Q}_v \mathbf{H}_v; \mathbf{Q}_u \mathbf{H}_u) = 0 \quad (11c) \end{aligned}$$

where (11a) follows from the fact that $\mathbf{P}_v \mathbf{F}_v S + \mathbf{P}_v \mathbf{H}_v Z = \mathbf{P}_u \mathbf{F}_u S + \mathbf{P}_u \mathbf{H}_u Z$ (see (7), (9)) and linear transformation to identify the overlap is invertible (see (10)). In (11c), the first term follows from counting the number of variables and the property that uniform distribution maximizes entropy, and the second term follows from the fact the symbols in Z

[4]Without loss of generality, we assume that $\mathbf{H}_v$ has full row rank, i.e., $\text{rank}(\mathbf{H}_v) = N$, because each v is assumed to connect to at least an unqualified edge so that $I(v; S) = 0$, then the linearly dependent rows of $\mathbf{H}_v$ in v must be linearly dependent as well (thus redundant).

[5]The overlap is found through matrices $\mathbf{P}_v, \mathbf{P}_u$ as multiplication from the left preserves the row space and the size of $\mathbf{P}_v, \mathbf{P}_u$ is set to match the dimension of the overlap.

[6]As a straightforward corollary, note that the security constraint applies to subspaces of the overlapping noise spaces. That is, when $\{u, v\}$ is unqualified, for all matrices $\bar{\mathbf{P}}_v, \bar{\mathbf{P}}_u$ where $\bar{\mathbf{P}}_v \mathbf{H}_v = \bar{\mathbf{P}}_u \mathbf{H}_u$, we have $\bar{\mathbf{P}}_v \mathbf{F}_v = \bar{\mathbf{P}}_u \mathbf{F}_u$.

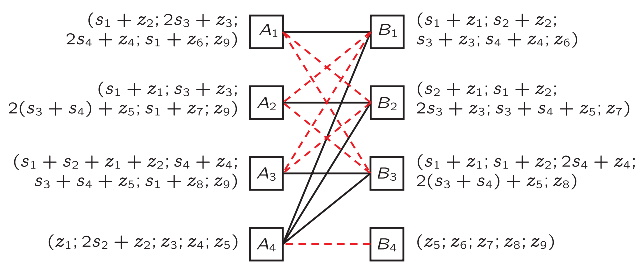


Fig. 2. A CDS instance and the vector linear achievable scheme of rate $R = 2/5$.

are i.i.d. and uniform. Conversely, any feasible linear scheme must satisfy (8), (9). Such a linear feasibility framework has appeared in related problems, e.g., index coding [16] and secure groupcast [17].

To facilitate later use, we summarize some useful properties of feasible linear schemes in the following lemma. A detailed proof can be found in Lemma 6 and Lemma 7 of [13].

Lemma 1: For any linear scheme as defined above and any edge $\{v, u\}$, we have

$$[\textit{Noise Align}] \quad \textit{dim}(\textit{rowspan}(\mathbf{H}_v) \cap \textit{rowspan}(\mathbf{H}_u)) \geq L, \quad \textit{if } \{u, v\} \textit{ is qualified}; \tag{12}$$

$$[\textit{Signal Align}] \ \mathbf{P}_v \mathbf{F}_v = \mathbf{P}_u \mathbf{F}_u, \ \textit{if } \{u, v\} \textit{ is unqualified}. \tag{13}$$

The intuition of the lemma is as follows. (12) follows from the correctness constraint (8), which requires the overlap of the noise spaces to have at least L dimensions as decoding is only possible over the overlapping space (so referred to as 'noise alignment') and other spaces are covered by independent noise variables. (13) follows from the security constraint (9), which says that over the overlapping noise space, the secret space must also be fully overlapping (so referred to as 'signal alignment' since both noise and secret fully align in this space) as otherwise the unqualified edge can reveal some linear combination of the secret symbols, violating the security constraint.

In the remainder of this paper, we use (8) and (9) to verify the correctness and security of a linear scheme. To illustrate how it works, let us consider again the CDS instance in Fig. 1.2 (reproduced in Fig. 2). We show that rate $R = 2/5$ is achievable, through presenting a vector linear scheme with $L = 4, N = 5$. That is, the secret has $L = 4$ symbols over $\mathbb{F}_3$ ($S = s_{[1:4]}$), and each signal has $N = 5$ symbols over $\mathbb{F}_3$. The assignment of the signals is given in Fig. 2. Suppose $Z = z_{[1:9]}$, where each z_i is uniform and i.i.d. over $\mathbb{F}_3$.

Let us verify that the above scheme is correct and secure. For simplicity, we do not write out explicitly the precoding matrices $\mathbf{F}_v$ and $\mathbf{H}_v$ for a signal v as the dimension is relatively large. Instead, we will directly find the overlap by inspection. Consider qualified edge $\{A_3, B_3\}$. A_3, B_3 both contain $(z_1 + z_2; z_4; z_5; z_8)$ (noise overlaps) and can then obtain 4 equations of the secret symbols, $(-s_1 + s_2; s_4; s_3 + s_4; s_1)$, which can recover $S = s_{[1:4]}$. Other cases of qualified edges can be verified similarly. Consider the unqualified edge $\{A_3, B_2\}$. $(z_1 + z_2; z_5)$ lies in the overlap of the noise spaces and the secret symbols projecting to this space are both $(s_1 + s_2; s_3 + s_4)$, thus no information is leaked. Other unqualified edges follow similarly. The rate achieved is thus $L/(2N) = 4/10 = 2/5$.

III. Results

A. Linear CDS Converse

Our first result is a general converse bound of linear CDS schemes, parameterized by the minimum connected edge cover number of internal qualified edges, ρ and stated in Theorem 1.

Theorem 1: For any CDS instance, the following converse bound holds for all linear schemes.

$$R_{\ linear} \leq \frac{\rho - 1}{2\rho}. \tag{14}$$

Remark 1: When $\rho = +\infty$, we have that for any internal qualified edge e, there is no set of connected edges that can cover all nodes in the unqualified path containing e (refer to Definition 4). This is equivalent to that there is no internal qualified edge within any qualified component, which reduces to the feasibility condition of capacity $1/2$ from Theorem 1 of [13].

The proof of Theorem 1 is presented in Section V-A.

To illustrate the idea, let us consider the CDS instance in Fig. 2. We may verify that $\rho = 5$ (see the explanation below Definition 3). Then Theorem 1 indicates that $R_{\ linear} \leq (\rho - 1)/(2\rho) = 2/5$. As rate $2/5$ is linearly achievable (see Fig. 2), the linear capacity of this CDS instance is $2/5$.

The intuition of the linear converse is as follows. (12) in Lemma 1 gives a lower bound on the dimension of the pairwise noise overlap of the two nodes in a qualified edge. We will start from this pairwise overlap to obtain a lower bound on the dimension of the overlap of all nodes in M, i.e., $\mathcal{R}(\mathbf{H}_{A_4}) \cap \mathcal{R}(\mathbf{H}_{B_1}) \cap \mathcal{R}(\mathbf{H}_{B_2}) \cap \mathcal{R}(\mathbf{H}_{B_3}) \cap \mathcal{R}(\mathbf{H}_{A_1}) \cap \mathcal{R}(\mathbf{H}_{A_3})$, which is the overlap of all pairwise overlaps of the edges in M, i.e., $(\mathcal{R}(\mathbf{H}_{A_4}) \cap \mathcal{R}(\mathbf{H}_{B_1})) \cap \cdots \cap (\mathcal{R}(\mathbf{H}_{A_3}) \cap \mathcal{R}(\mathbf{H}_{B_3}))$. For example, consider the first two edges in M.

$$\begin{aligned} & d\big(\mathcal{R}(\mathbf{H}_{A_4}) \cap \mathcal{R}(\mathbf{H}_{B_1}) \cap \mathcal{R}(\mathbf{H}_{B_2})\big) \\ & = d\big((\mathcal{R}(\mathbf{H}_{A_4}) \cap \mathcal{R}(\mathbf{H}_{B_1})) \cap (\mathcal{R}(\mathbf{H}_{A_4}) \cap \mathcal{R}(\mathbf{H}_{B_2})) \big) \\ & \geq d\big(\mathcal{R}(\mathbf{H}_{A_4}) \cap \mathcal{R}(\mathbf{H}_{B_1})\big) + d\big(\mathcal{R}(\mathbf{H}_{A_4}) \cap \mathcal{R}(\mathbf{H}_{B_2})\big) \\ & \quad - \ d\big(\mathcal{R}(\mathbf{H}_{A_4})\big) \qquad (15) \\ & \overset{(12)}{\geq} \ L + L - N \end{aligned}$$

where (15) follows from the sub-modularity of linear rank functions and the direct sum (the space spanned by the union of the two sets of vectors) of the two pairwise overlaps is contained in $\mathcal{R}(\mathbf{H}_{A_4})$ as each pairwise overlap involves $\mathcal{R}(\mathbf{H}_{A_4})$. The last step follows from the pairwise overlap constraint (12) and the fact that the rank of $\mathbf{H}_{A_4}$ is N. We have now transformed the pairwise overlap of dimension L to 3-wise overlap of $2L - N$, where a term of $L - N$ is added. Next as the edges are connected, we may apply sub-modularity repeatedly and find the overlap of all noise spaces in M by including one connected edge at one time, whose dimension turns out to be no less than $L + (\rho - 1)(L - N) = 5L - 4N$, i.e., the $L - N$ term is added $\rho - 1$ times (from pairwise overlap to ρ-wise overlap). Then by (13) in Lemma 1, we know that such noise overlap leads to signal overlap for all nodes $V_P = \{A_1, B_2, A_3, B_1\}$

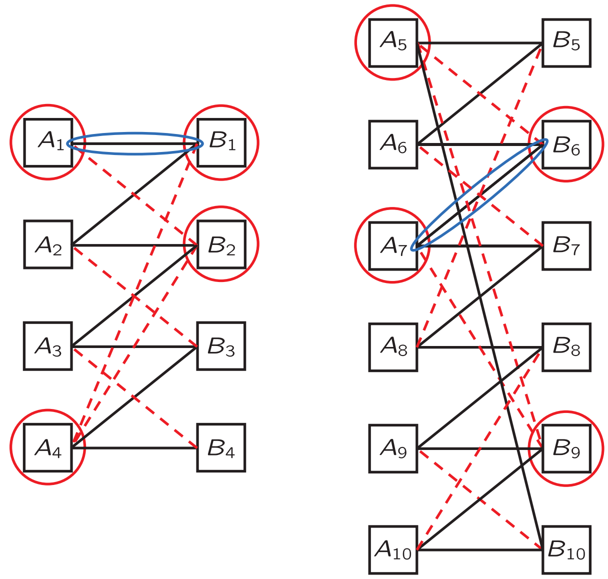


Fig. 3. A CDS instance where each qualified component is either a path or a cycle. For each qualified component, an internal qualified edge e is put in a blue circle and the nodes V_P in the unqualified path P are put in red circles. Then the connected edge cover M is the qualified path that connects to all nodes in V_P. For the left qualified component, $\rho(e, P) = 6$ (the path from A_1 to A_4); for the right qualified component, $\rho(e, P) = 7$ (the path from B_9 to A_7). The unqualified edges connecting two nodes from different qualified components are not drawn and can be arbitrary.

in the unqualified path P, in particular including the two nodes A_1, B_1 in the internal qualified edge e. As overlapping signal contributes no information for decoding, such overlap shall not exist (when the noise overlap of A_1, B_1 is exactly L in (12) and this will be relaxed in the general proof), i.e., $\rho L - (\rho - 1)N \leq 0$, and $R_{\text{linear}} = L/(2N) \leq (\rho - 1)/(2\rho)$.

B. Linear Capacity of a Class of CDS Instances

Next, we proceed to our second result, which shows that the linear converse in Theorem 1 is tight for a class of CDS instances and is stated in Theorem 2.

Theorem 2: For any CDS instance where the qualified edges in each qualified component form either a path or a cycle,[7] *the linear capacity is* $C_{\text{linear}} = (\rho - 1)/(2\rho)$.

Note that Theorem 2 only places constraints on the structure of qualified edges and works for any possible configuration of unqualified edges.

The proof of Theorem 2 is presented in Section V-B.

We give an example here to illustrate the idea. Consider the CDS instance in Fig. 3 and we show that the linear capacity is $C_{\text{linear}} = 5/12$. Theorem 2 can be applied as the instance contains two qualified components, where the qualified edges form a path in one qualified component and form a cycle in the other qualified component. $\rho = 6$, because for the left qualified component, there is an internal qualified edge $e = \{A_1, B_1\}$ (see the blue circle) in unqualified path $P = \{\{A_1, B_2\}, \{B_2, A_4\}, \{A_4, B_1\}\}$ (see the red circles), which is then covered by a qualified path with 6 edges $M = \{\{A_1, B_1\}, \{B_1, A_2\}, \{A_2, B_2\}, \{B_2, A_3\}, \{A_3, B_3\},$

[7]A cycle is a path where the first node is the same as the last node. If the qualified edges form either a path or a cycle, equivalently, we have that each node is connected to at most two qualified edges.

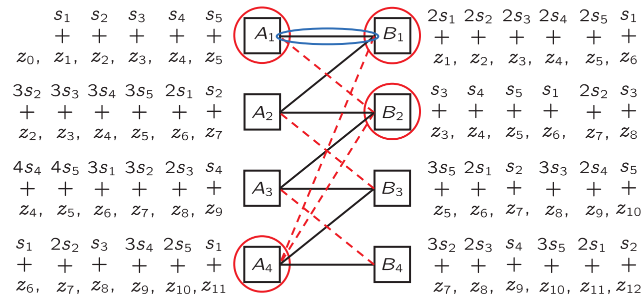


Fig. 4. A linear capacity achieving scheme for the left qualified component.

$\{B_3, A_4\}\}$. It can be verified that this M has the minimum cardinality, so $\rho = 6$. Then the converse bound follows from Theorem 1.

We now consider the achievable scheme. We will consider each qualified component one by one and use independent noise variables. Let us start from the left qualified component (a qualified path from A_1 to B_4), where the assignment of each signal is given in Fig. 4.

The uniform and i.i.d. noise variables are assigned sequentially to the nodes in the path following a sliding window manner, where the first node A_1 uses $z_0, z_1, \cdots, z_5 \triangleq z_{[0:5]}$, the second node uses $z_{[1:6]}$, and so on (every two consecutive nodes share $L = 5$ common noise variables). Note that this noise assignment does not depend on the structure of the unqualified edges (which is a key property that simplifies the scheme design). The secret symbols $s_{[1:5]}$ are assigned cyclicly to the noise variables, i.e., $s_{[1:5]}$ are assigned to $z_{[1:5]}$, $z_{[6:10]}$ etc. (i.e., s_j is assigned to z_{5B+j} for any integer B). The coefficients of s_j are the only left and most important part. To this end, focus on each z_i in an arbitrary order and consider only the nodes that contain z_i. For example, consider z_6, which appears in 6 nodes $B_1, A_2, B_2, A_3, B_3, A_4$ and consider the subgraph induced by these 6 nodes. For the induced subgraph, consider each unqualified component sequentially (any order will work and one node that connects to no unqualified edge is a trivial unqualified component) and assign the same signal to each node in the unqualified component. So here first consider the unqualified path $\{\{B_1, A_4\}, \{A_4, B_2\}\}$ and assign $s_1 + z_6$ to B_1, A_4, B_2; second consider the unqualified path $\{A_2, B_3\}$ and assign $2s_1 + z_6$ to A_2, B_3; lastly consider A_3 and assign $3s_1 + z_6$ to A_3. All other z_i can be treated in the same manner (essentially for each z_i, we apply the scheme from Theorem 1 of [13]). This completes the description of the scheme.

The security and correctness of the scheme follow from the assignment in a straightforward manner. For security, note that for each unqualified edge, the signal for overlapping noise is set to be the same so that the security constraint (9) is satisfied. For correctness, we note that in the qualified path, every two consecutive nodes are connected by a qualified edge and share L noise symbols. For each shared noise symbol, the secret symbols have different coefficients by noting that the connected edge cover number $\rho > \rho - 1$ and each z_i appears in consecutive $\rho - 1$ edges.

After completing the left qualified component, we proceed to the right component (a qualified cycle from A_5 to B_{10} and back to A_5), whose assignment is given in Fig. 5.

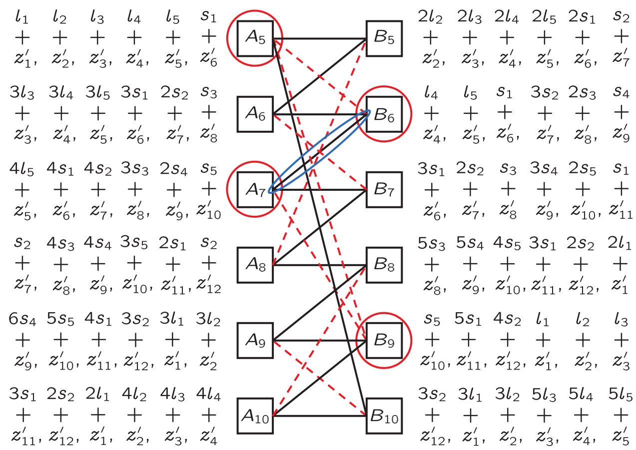


Fig. 5. A linear capacity achieving scheme for the right qualified component.

The assignment for a cycle is similar to that of a path and we only highlight the differences here. To cope with the fact that the first and last node is the same for a cycle, the sliding window based noise assignment needs to wrap back as well (see the $z'_{[1:5]}$ symbols in Fig. 5). Also, the secrets associated with nodes near the front and end need to be coded and generic (one linear combination of secret symbols $s_{[1:5]}$ is denoted by l_i in Fig. 5) so that when combined with any blocks of s_i, all secret symbols can be decoded as long as the collective number is sufficient. For example, consider the qualified edge $\{A_{10}, B_9\}$, where we need to recover $S = s_{[1:5]}$ from $(s_1, s_2, l_1, l_2, l_3)$. We will use Cauchy matrix to realize l_i over a sufficiently large field. The other elements for the cycle case are the same as those for a path, i.e., consider each z_i sequentially and set each unqualified component to have the same signal that contains z_i in the induced subgraph. Details are deferred to the general proof.

Lastly, we consider the unqualified edges connecting two nodes from different qualified components, which are not drawn in Fig. 3. The presence of any number of such unqualified edges will not change the result - for the converse, ρ is not influenced; for the achievability, security is preserved as independent noises are used (i.e., z_i, z'_i in Fig. 4 and Fig. 5, respectively) for different qualified components. The rate achieved is $R = L/(2N) = 5/12$ as the secret has $L = 5$ symbols and each signal has $N = 6$ symbols.

C. Converse $(\rho - 1)/(2\rho)$ Is Not Tight

The techniques from Theorem 1 and Theorem 2 are not sufficient in general. Our third result is the linear capacity characterization of a CDS instance in Fig. 6 that goes beyond previous theorems.

Theorem 3: The linear capacity of the CDS instance in Fig. 6 is $C_{\text{linear}} = 7/18$.

The proof of Theorem 3 is presented in Section V-C. The converse from Theorem 1 is not tight as $\rho = 5$ and the converse bound is $R_{\text{linear}} \leq 2/5$, which is strictly larger than $7/18$, the linear capacity. The converse proof of $R_{\text{linear}} \leq 7/18$ requires a highly non-trivial analysis of the noise and space spaces involved such that it goes well beyond the techniques from Theorem 1 and does not appear to admit a simple explanation (so we are not yet able to generalize it further). Once the converse bound is found, the achievable scheme follows by its guidance and falls in the general linear feasibility framework presented in Section II-B.

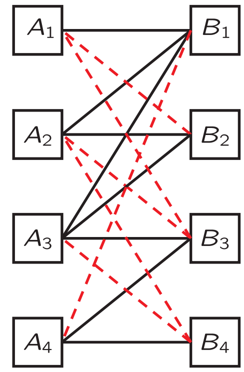


Fig. 6. A CDS instance whose linear capacity is $7/18$. The converse bound from Theorem 1 is not tight.

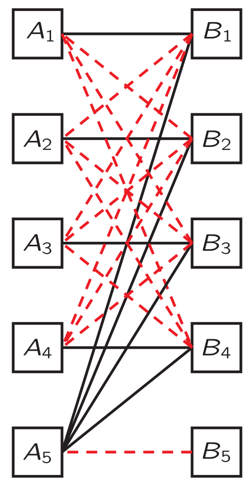


Fig. 7. A CDS instance whose linear capacity is open. The best known converse bound is from Theorem 1 and is equal to $2/5$.

IV. Conclusion

In this work, we take a Shannon theoretic perspective at the canonical conditional disclosure of secrets problem to seek capacity characterizations where the secret size is allowed to approach infinity while most cryptography work focuses on the scaling of communication cost with the input size[8] [2], [5], [6], [7], [8], [9]. This Shannon theoretic perspective follows the footsteps of recent attempts in the information theory community on other cryptographic primitives [19], [20], [21], [22], [23], [24], [25], [26], [27], [28]. Towards this end, we further develop the noise and signal alignment approach, which is a variation of interference alignment originally studied in wireless communication networks [15], [29], [30] and is introduced in [13], to characterize the linear capacity of a class of CDS instances, which go beyond the highest capacity scenarios found in [13]. Along the line, we identify a general

[8]One exception is recent work [18], where the amortization formulation essentially considers the same rate metric as our work. The difference is that we focus on each CDS instance and pursue exact linear capacity characterizations (so impossibility claims included) while [18] aims at worst case rate approximation for all CDS instances.

linear converse bound (see Theorem 1) and a linear feasibility framework that facilitates the design of linear schemes once the target rate value is fixed (see Section II-B). However, these results are not sufficient to fully understand the linear capacity of CDS in general. We conclude by giving an intriguing CDS instance whose linear capacity is open (see Fig. 7). Note that this instance is only slightly changed from the solved instance in Fig. 2. A general achievable scheme that works for all functions and has competitive rate performance is also missing, which comprises an interesting open problem.

V. Appendix

A. Proof of Theorem 1

The proof of Theorem 1 follows similarly from that of the CDS instance in Fig. 2 considered in the previous section. We first simplify a notation that will be frequently used. For nodes $v_1, \cdots, v_i$, denote the dimension of the overlap of their noise spaces as $\alpha_{v_1\cdots v_i}$, i.e.,

$$\alpha_{v_1\cdots v_i} \triangleq d(\mathcal{R}(\mathbf{H}_{v_1}) \cap \cdots \cap \mathcal{R}(\mathbf{H}_{v_i})). \quad (16)$$

Consider any CDS instance $G_f(V, E)$, where $\rho \neq +\infty$ and focus on an internal qualified edge e in an unqualified path P such that $\rho(e, P) = \rho$. Then the connected edge cover M for nodes V_P in P contains ρ edges and $\rho + 1$ nodes, denoted as $V_M = \{v_1, v_2, \cdots, v_{\rho+1}\} \subset V$. Note that such e, P, M are guaranteed to exist as $\rho \neq +\infty$ and according to the definition of ρ, the connected edge cover M attains the minimal cardinality so that M is a spanning tree of the nodes V_M.

Start with the internal qualified edge e in M, say $e = \{v_{i_1}, v_{i_2}\} \subset M, i_1, i_2 \in \{1, 2, \cdots, \rho+1\}$. As M is connected, there must exist a node $v_{i_3} \in V_M, i_3 \notin \{i_1, i_2\}$ and a node $u_1 \in \{v_{i_1}, v_{i_2}\}$ such that $\{u_1, v_{i_3}\}$ is a qualified edge. Then from sub-modularity, we have

$$\alpha_{v_{i_1}v_{i_2}v_{i_3}} \geq \alpha_{v_{i_1}v_{i_2}} + \alpha_{u_1 v_{i_3}} - N. \quad (17)$$

Then we proceed similarly to find $v_{i_4} \in V_M, i_4 \notin \{i_1, i_2, i_3\}$ such that $\{u_2, v_{i_4}\}$ is a qualified edge, where $u_2 \in \{v_{i_1}, v_{i_2}, v_{i_3}\}$. Again from sub-modularity, we have

$$\begin{aligned}\alpha_{v_{i_1}v_{i_2}v_{i_3}v_{i_4}} &\geq \alpha_{v_{i_1}v_{i_2}v_{i_3}} + \alpha_{u_2 v_{i_4}} - N \\ &\overset{(17)}{\geq} \alpha_{v_{i_1}v_{i_2}} + \alpha_{u_1 v_{i_3}} + \alpha_{u_2 v_{i_4}} - 2N. \quad (18)\end{aligned}$$

Continue this procedure, i.e., we include one node $v_{i_j} \in V_M, i_j \notin \{i_1, \cdots, i_{j-1}\}, j \in \{5, \cdots, \rho+1\}$ at one time such that $\{u_{j-2}, v_{i_j}\} \in M$ and $u_{j-2} \in \{v_{i_1}, \cdots, v_{i_{j-1}}\}$. Then we have

$$\begin{aligned}\alpha_{v_{i_1}v_{i_2}\cdots v_{i_{\rho+1}}} &\geq \alpha_{v_{i_1}\cdots v_{i_\rho}} + \alpha_{u_{\rho-1}v_{i_{\rho+1}}} - N \\ &\geq \alpha_{v_{i_1}v_{i_2}} + \alpha_{u_1 v_{i_3}} + \alpha_{u_2 v_{i_4}} + \cdots \\ &\quad + \alpha_{u_{\rho-1}v_{i_{\rho+1}}} - (\rho-1)N. \quad (19)\end{aligned}$$

Note that $i_1, \cdots, i_{\rho+1}$ are distinct so that $V_M = \{v_1, \cdots, v_{\rho+1}\} = \{v_{i_1}, \cdots, v_{i_{\rho+1}}\}$.

As the $\rho + 1$ noise spaces have an overlap of dimension $\alpha_{v_{i_1}v_{i_2}\cdots v_{i_{\rho+1}}}$, there exist $\rho + 1$ projection matrices $\mathbf{P}^{\cap}_{v_{i_1}}, \cdots, \mathbf{P}^{\cap}_{v_{i_{\rho+1}}}$ of rank $\alpha_{v_{i_1}v_{i_2}\cdots v_{i_{\rho+1}}}$ each such that

$$\mathbf{P}^{\cap}_{v_{i_1}}\mathbf{H}_{v_{i_1}} = \mathbf{P}^{\cap}_{v_{i_2}}\mathbf{H}_{v_{i_2}} = \cdots = \mathbf{P}^{\cap}_{v_{i_{\rho+1}}}\mathbf{H}_{v_{i_{\rho+1}}}. \quad (20)$$

Next, switch focus to the unqualified path P. Consider the nodes $V_P \subset V_M$ and denote $V_P = \{v_{i_1}, v_{j_1}, v_{j_2}, \cdots, v_{j_{|V_P|-2}}, v_{i_2}\} \subset \{v_{i_1}, v_{i_2}, \cdots, v_{i_{\rho+1}}\} = V_M$ such that $\{v_{i_1}, v_{j_1}\}$, $\{v_{j_1}, v_{j_2}\}$, $\cdots$, $\{v_{j_{|V_P|-2}}, v_{i_2}\}$ are unqualified edges. By (13), i.e., the signal alignment constraint from Lemma 1, and (20), we have

$$\begin{aligned}\mathbf{P}^{\cap}_{v_{i_1}}\mathbf{F}_{v_{i_1}} &= \mathbf{P}^{\cap}_{v_{j_1}}\mathbf{F}_{v_{j_1}} = \cdots = \mathbf{P}^{\cap}_{v_{j_{|V_P|-2}}}\mathbf{F}_{v_{j_{|V_P|-2}}} \\ &= \mathbf{P}^{\cap}_{v_{i_2}}\mathbf{F}_{v_{i_2}} \quad \Rightarrow \mathbf{P}^{\cap}_{v_{i_1}}\mathbf{F}_{v_{i_1}} = \mathbf{P}^{\cap}_{v_{i_2}}\mathbf{F}_{v_{i_2}}. \quad (21)\end{aligned}$$

Finally, consider the internal qualified edge $e = \{v_{i_1}, v_{i_2}\}$ and identify the noise overlap through matrices $\mathbf{P}_{v_{i_1}}, \mathbf{P}_{v_{i_2}}$ that have rank $\alpha_{v_{i_1}, v_{i_2}}$, i.e., $\mathbf{P}_{v_{i_1}}\mathbf{H}_{v_{i_1}} = \mathbf{P}_{v_{i_2}}\mathbf{H}_{v_{i_2}}$. Noting that $\mathcal{R}(\mathbf{P}^{\cap}_{v_{i_1}})$ is a subspace of $\mathcal{R}(\mathbf{P}_{v_{i_1}})$, we set

$$\begin{aligned}\mathbf{P}^{\cap}_{v_{i_1}} &= \mathbf{P}_{v_{i_1}}(1 : \alpha_{v_{i_1}v_{i_2}\cdots v_{i_{\rho+1}}}, :), \\ \mathbf{P}^{\cap}_{v_{i_2}} &= \mathbf{P}_{v_{i_2}}(1 : \alpha_{v_{i_1}v_{i_2}\cdots v_{i_{\rho+1}}}, :) \quad (22)\end{aligned}$$

without loss of generality, i.e., the first $\alpha_{v_{i_1}v_{i_2}\cdots v_{i_{\rho+1}}}$ rows of $\mathbf{P}_{v_{i_1}}$ are $\mathbf{P}^{\cap}_{v_{i_1}}$. Then from the correctness constraint (8) for qualified edge $e = \{v_{i_1}, v_{i_2}\}$, we have

$$\begin{aligned}L &\overset{(8)}{=} r\left(\mathbf{P}_{v_{i_1}}\mathbf{F}_{v_{i_1}} - \mathbf{P}_{v_{i_2}}\mathbf{F}_{v_{i_2}}\right) \quad (23a)\\
&\overset{(21)(22)}{=} r\Big(\mathbf{P}_{v_{i_1}}(\alpha_{v_{i_1}v_{i_2}\cdots v_{i_{\rho+1}}} + 1 : \alpha_{v_1 v_2}, :)\mathbf{F}_{v_{i_1}} \\
&\quad -\mathbf{P}_{v_{i_2}}(\alpha_{v_{i_1}v_{i_2}\cdots v_{i_{\rho+1}}} + 1 : \alpha_{v_1 v_2}, :)\mathbf{F}_{v_{i_2}}\Big) \quad (23b)\\
&\leq \alpha_{v_{i_1}v_{i_2}} - \alpha_{v_{i_1}v_{i_2}\cdots v_{i_{\rho+1}}} \quad (23c)\\
&\overset{(19)}{\leq} \alpha_{v_{i_1}v_{i_2}} - \Big(\alpha_{v_{i_1}v_{i_2}} + \alpha_{u_1 v_{i_3}} + \alpha_{u_2 v_{i_4}} + \cdots \\
&\quad + \alpha_{u_{\rho-1}v_{i_{\rho+1}}} - (\rho-1)N\Big) \quad (23d)\\
&=(\rho-1)N \\
&\quad - \left(\alpha_{u_1 v_{i_3}} + \alpha_{u_2 v_{i_4}} + \cdots + \alpha_{u_{\rho-1}v_{i_{\rho+1}}}\right) \quad (23e)\\
&\overset{(12)}{\leq} (\rho-1)N - (\rho-1)L \quad (23f)\\
&\Rightarrow \rho L \leq (\rho-1)N \\
&\Rightarrow R_{\text{linear}} = L/(2N) \leq (\rho-1)/(2\rho). \quad (23g)\end{aligned}$$

The proof of the linear converse bound in Theorem 1 is thus complete.

B. Proof of Theorem 2

In this section, we present a vector linear CDS scheme that achieves rate $(\rho-1)/(2\rho)$ as long as each qualified component of the CDS instance is either a path or a cycle. Recall that ρ is the minimum connected edge cover number defined in Definition 4. Specifically, we set $L = \rho - 1$, i.e., each secret has L symbols $S = (s_1, \cdots, s_{\rho-1}) \triangleq s_{[1:\rho-1]}$ from $\mathbb{F}_p$ and $N = \rho$, i.e., each signal (node) v has N symbols from $\mathbb{F}_p$. We assume that p is a prime number that is no smaller than $2\rho - 2$.

To prepare for the achievable scheme, we first define $L = \rho-1$ generic linear combinations $l_{[1:\rho-1]}$ of the secret symbols.

$$\begin{aligned}&(l_1; \cdots; l_{\rho-1})_{(\rho-1)\times 1} \\ &= \mathbf{C}_{(\rho-1)\times(\rho-1)} \times (s_1; \cdots, s_{\rho-1})_{(\rho-1)\times 1}\end{aligned}$$

$$\mathbf{C}_{(\rho-1)\times(\rho-1)}(i,j) = \frac{1}{x_i - y_j}, i,j \in \{1,\cdots,\rho-1\} \quad (24)$$

where x_i, y_j are distinct elements from $\mathbb{F}_p$ (the existence is guaranteed by the fact that the field size p is no smaller than $2\rho - 2$), so $\mathbf{C}_{(\rho-1)\times(\rho-1)}$ is a Cauchy matrix whose every square sub-matrix has full rank [31].

Consider any CDS instance $G_f(V,E)$ such that the minimum connected edge cover number for any internal qualified edge is ρ. Suppose the instance contains Q qualified components, where each qualified component is either a path or a cycle of qualified edges. Denote the node set of the q-th qualified component by $V^q, q \in \{1,\cdots,Q\}$ such that $V = V^1 \cup \cdots \cup V^Q$. For each qualified component, we will use independent uniform i.i.d. noise symbols from $\mathbb{F}_p$, denoted as $z^q = (z_0^q, z_1^q, z_2^q, \cdots)$. The exact number of noise symbols used in each z^q will be specified when we give the scheme and is indicated by the subscript. So $Z = (z^1,\cdots,z^Q)$. We are now ready to specify the signal design.

1) Consider each qualified component sequentially. If the q-th qualified component is a path, go to 2; otherwise the q-th qualified component is a cycle, go to 3.
2) The nodes V^q form a qualified path. Denote $V^q = \{v_1^q,\cdots,v_{|V^q|}^q\}$. Suppose $\{v_1^q,v_2^q\}$, $\{v_2^q,v_3^q\}$, $\cdots$, $\{v_{|V^q|-1}^q, v_{|V^q|}^q\}$ are qualified edges, i.e., we interpret v_1^q as the first node and $v_{|V^q|}^q$ as the end node of the path.
   a) Assign the noise variables in a sequential manner as follows.
$$v_1^q = (z_0^q, z_1^q, \cdots, z_{\rho-1}^q), v_2^q = (z_1^q, z_2^q,\cdots,z_\rho^q), \cdots, v_{|V^q|}^q = (z_{|V^q|-1}^q, \cdots, z_{|V^q|+\rho-2}^q). \quad (25)$$
   b) We now describe how to the include the secret symbols to each node. Consider the nodes that contain each noise symbol $z_1^q,\cdots,z_{|V^q|+\rho-2}^q$ sequentially (z_0^q will not be used) and the induced subgraph formed by these nodes. Note that each noise symbol $z_j^q, j \in \{1,\cdots,|V^q|+\rho-2\}$ appears at no more than ρ nodes and denote the induced subgraph by $G_j^q \subset G_f$. Suppose G_j^q contains K_j^q unqualified components,[9] each of which is considered sequentially as follows.

      For each node v_i^q in the k-th unqualified component of $G_j^q, k \in \{1,\cdots,K_j^q\}$, $j \in \{1,\cdots,|V^q|+\rho-2\}$, replace z_j^q by $k \times s_{j \bmod (\rho-1)} + z_j^q$. (26)

   Note that in $s_{j \bmod (\rho-1)}$, the subscript is defined over $\{1,\cdots,\rho-1\}$ as secret symbols are $s_1,\cdots,s_{\rho-1}$, i.e., $\{1\},\cdots,\{\rho-1\}$ are the representative of the equivalent classes of the modulo $\rho-1$ function. The signal assignment is complete for the path case.

[9]A node that connects to no unqualified edge is a trivial unqualified component. As there are at most ρ nodes in G_j^q, we have that $K_j^q \le \rho$.

3) The nodes V^q form a qualified cycle. Denote $V^q = \{v_1^q,\cdots,v_{|V^q|}^q\}$, and suppose $\{v_1^q,v_2^q\}$, $\{v_2^q,v_3^q\}$, $\cdots$, $\{v_{|V^q|-1}^q, v_{|V^q|}^q\}, \{v_{|V^q|}^q, v_1^q\}$ are qualified edges.
   a) Assign the noise variables in the following cyclic manner.
$$\begin{aligned} v_1^q =& (z_1^q, z_2^q, \cdots, z_\rho^q), v_2^q \\ &= (z_2^q, z_3^q, \cdots, z_{\rho+1}^q), \cdots, \\ v_{|V^q|-1}^q &= (z_{|V_q|-1}^q, z_{|V_q|}^q, z_1^q, \cdots, z_{\rho-2}^q), \\ v_{|V^q|}^q &= (z_{|V_q|}^q, z_1^q, \cdots, z_{\rho-1}^q). \end{aligned} \quad (27)$$
   b) We now describe how to the include the secret symbols to each node. Consider the nodes that contain each noise symbol $z_1^q,\cdots,z_{|V^q|}^q$ sequentially and the induced subgraph formed by these nodes. Note that each noise symbol $z_j^q, j \in \{1,\cdots,|V^q|\}$ appears at ρ nodes and denote the induced subgraph by $G_j^q \subset G_f$. Suppose G_j^q contains K_j^q unqualified components, each of which is considered sequentially as follows.

      For each node v_i^q in the k-th unqualified component of $G_j^q, k \in \{1,\cdots,K_j^q\}$, if $j \in \{1,\cdots,\rho-1\}$, replace z_j^q by $k \times l_j + z_j^q$; otherwise $j \in \{\rho,\cdots,|V_q|\}$, replace z_j^q by $k \times s_{j \bmod (\rho-1)} + z_j^q$. (28)

   Note that similar as above, $j \bmod (\rho-1)$ is defined over $\{1,\cdots,\rho-1\}$. The signal assignment is complete for the cycle case.

After describing the signal design for all nodes, we proceed to show that the scheme is correct and secure.

First, we prove that the correctness constraint (8) is satisfied. All qualified edges belong to some qualified component, so it suffices to consider each qualified component. We have two cases.

- The first case is when the qualified component is a path. From the noise assignment (25), we know that the two nodes u, v in any qualified edge share $L = \rho - 1$ noise symbols with consecutive subscripts. Further, according to the signal assignment (26), these L consecutive noise symbols are each mixed with one distinct secret symbol from the L symbols in S. In addition, each shared secret symbol $s_i, i \in \{1,\cdots,L\}$ in v and u is multiplied by different coefficients k (see (26)). We prove this claim by contradiction, i.e., suppose that the coefficients k are the same. Then due to the signal assignment (26), $e = \{u,v\}$ must be an internal qualified edge in an unqualified path P, and we can find a connected edge cover M for the nodes in P and all nodes in M share one same noise symbol. Recall from Definition 4 that M contains $\rho(e,P)$ edges and $\rho(e,P) \ge \rho$. As a result, M contains at least $\rho+1$ nodes and these nodes share one same noise symbol, which is not possible because from the noise assignment (25), each noise symbol only appears at ρ nodes at most. Thus the coefficients for the L secret symbols in v, u are

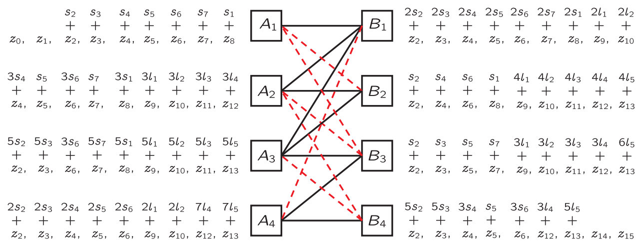


Fig. 8. A linear capacity achieving scheme with rate $7/18$. l_i is a generic linear combination of $s_1, \cdots, s_7$.

all distinct and from $\{v, u\}$ we can recover S with no error.

- The second case is when the qualified component is a cycle, whose proof is similar to the path case. Similarly from the noise assignment (27), any two nodes u, v in a qualified edge share $L = \rho - 1$ noise symbols with cyclicly consecutive subscripts. Further, according to the signal assignment (28), these L noise symbols are each mixed with either one distinct secret symbol s_i from the L symbols in S or one generic linear combination l_j. With a similar reasoning as above (due to the definition of ρ and each noise appears at ρ nodes), the multiplicative coefficients k for s_i, l_j are distinct. As l_j are from a Cauchy matrix (see (24)), whose every square sub-matrix has full rank, we conclude that from $\{v, u\}$ we can obtain L equations of form s_i, l_j thus recover S with no error.

Second, we prove that the security constraint (9) is satisfied. We have two cases for an unqualified edge.

- The first case is when the two nodes u, v of the unqualified edge are from the same qualified component. Security is guaranteed because in the signal assignment (26), (28), when the noise space overlaps, the same signal equation is assigned, i.e., signal alignment is ensured and (9) holds.
- The second case is when the two nodes u, v of the unqualified edge are from two different qualified components. As the noise symbols $z^q, z^{q'}$ are independent for distinct qualified components, the noise spaces of u, v have no overlap and (9) trivially holds.

The proof of Theorem 2 is now complete.

C. Proof of Theorem 3

We show that the linear capacity of the CDS instance in Fig. 6 is $7/18$. The achievable scheme is given in Fig. 8. The secret symbols $s_{[1:7]}$ are from $\mathbb{F}_{13}$ and $l_{[1:5]}$ are defined as follows.

$$(l_1; \cdots; l_5)_{5\times 1} = \mathbf{C}_{5\times 7} \times (s_1; \cdots, s_7)_{7\times 1}$$
$$\mathbf{C}_{5\times 7}(i,j) = \frac{1}{x_i - y_j}, \quad i \in \{1, \cdots, 5\}, j \in \{1, \cdots, 7\} \tag{29}$$

where x_i, y_j are distinct elements from $\mathbb{F}_{13}$ so that $\mathbf{C}_{5\times 7}$ is a Cauchy matrix whose every square sub-matrix has full rank. The correctness and security constraints (8) (9) are straightforward to verify.

Next we provide the converse proof. Recall that $\alpha_{v_1 \cdots v_J}$ denotes the dimension of the overlap of the row span of the noise precoding matrices of $v_1, \cdots, v_J$ and for each node, the rank of the noise precoding matrix is N (see Footnote 2).

We first give an upper bound for α_{A_2,A_3}, where the proof is similar to that of Theorem 1. Consider nodes A_2, A_3, B_3, A_4, B_4. To identify their noise overlap, we may find 5 matrices $\mathbf{P}^{\cap}_{A_2}$, $\mathbf{P}^{\cap}_{A_3}$, $\mathbf{P}^{\cap}_{B_3}$, $\mathbf{P}^{\cap}_{A_4}$, $\mathbf{P}^{\cap}_{B_4}$ of rank $\alpha_{A_2A_3B_3A_4B_4}$ each so that

$$\mathbf{P}^{\cap}_{A_2}\mathbf{H}_{A_2} = \mathbf{P}^{\cap}_{A_3}\mathbf{H}_{A_3} = \mathbf{P}^{\cap}_{B_3}\mathbf{H}_{B_3} = \mathbf{P}^{\cap}_{A_4}\mathbf{H}_{A_4} = \mathbf{P}^{\cap}_{B_4}\mathbf{H}_{B_4}. \tag{30}$$

Further, $\{\{A_3, B_4\}, \{B_4, A_2\}, \{A_2, B_3\}\}$ is an unqualified path. From (13) and (30), we have

$$\mathbf{P}^{\cap}_{A_3}\mathbf{F}_{A_3} = \mathbf{P}^{\cap}_{B_4}\mathbf{F}_{B_4} = \mathbf{P}^{\cap}_{A_2}\mathbf{F}_{A_2} = \mathbf{P}^{\cap}_{B_3}\mathbf{F}_{B_3} \Rightarrow \mathbf{P}^{\cap}_{A_3}\mathbf{F}_{A_3} = \mathbf{P}^{\cap}_{B_3}\mathbf{F}_{B_3}. \tag{31}$$

Consider now qualified edge $\{A_3, B_3\}$ and identify the noise overlap through $\mathbf{P}_{A_3}, \mathbf{P}_{B_3}$ of rank $\alpha_{A_3B_3}$ so that $\mathbf{P}_{A_3}\mathbf{H}_{A_3} = \mathbf{P}_{B_3}\mathbf{H}_{B_3}$. As $\mathcal{R}(\mathbf{P}^{\cap}_{A_3})$ is a subspace of $\mathcal{R}(\mathbf{P}_{A_3})$, without loss of generality we set

$$\mathbf{P}^{\cap}_{A_3} = \mathbf{P}_{A_3}(1 : \alpha_{A_2A_3B_3A_4B_4}, :), \quad \mathbf{P}^{\cap}_{B_3} = \mathbf{P}_{B_3}(1 : \alpha_{A_2A_3B_3A_4B_4}, :). \tag{32}$$

From the correctness constraint (8), we have

$$L \overset{(8)}{=} r\left(\mathbf{P}_{A_3}\mathbf{F}_{A_3} - \mathbf{P}_{B_3}\mathbf{F}_{B_3}\right) \tag{33a}$$
$$\overset{(31)(32)}{=} r\left(\mathbf{P}_{A_3}(\alpha_{A_2A_3B_3A_4B_4} + 1 : \alpha_{A_3B_3}, :)\mathbf{F}_{A_3} - \mathbf{P}_{B_3}(\alpha_{A_2A_3B_3A_4B_4} + 1 : \alpha_{A_3B_3}, :)\mathbf{F}_{B_3}\right) \tag{33b}$$
$$\le \alpha_{A_3B_3} - \alpha_{A_2A_3B_3A_4B_4} \tag{33c}$$
$$\le \alpha_{A_3B_3} - (\alpha_{A_2A_3} + \alpha_{A_3B_3} + \alpha_{B_3A_4} + \alpha_{A_4B_4} - 3N) \tag{33d}$$
$$= 3N - (\alpha_{A_2A_3} + \alpha_{B_3A_4} + \alpha_{A_4B_4}) \tag{33e}$$
$$\Rightarrow \alpha_{A_2A_3} \le 3N - (\alpha_{B_3A_4} + \alpha_{A_4B_4}) - L \tag{33f}$$

where (33d) follows from sub-modularity and we have obtained the desired upper bound for $\alpha_{A_2A_3}$.

We are now ready for the final step, which is a similar chain of arguments as above. Consider 6 nodes $A_1, B_1, A_2, B_2, A_3, B_3$ and identify their noise overlap through matrices $\mathbf{P}^{\cap\cap}_{A_1}$, $\mathbf{P}^{\cap\cap}_{B_1}$, $\mathbf{P}^{\cap\cap}_{A_2}$, $\mathbf{P}^{\cap\cap}_{B_2}$, $\mathbf{P}^{\cap\cap}_{A_3}$, $\mathbf{P}^{\cap\cap}_{B_3}$ of rank $\alpha_{A_1B_1A_2B_2A_3B_3}$ each.

$$\mathbf{P}^{\cap\cap}_{A_1}\mathbf{H}_{A_1} = \mathbf{P}^{\cap\cap}_{B_1}\mathbf{H}_{B_1} = \mathbf{P}^{\cap\cap}_{A_2}\mathbf{H}_{A_2} = \mathbf{P}^{\cap\cap}_{B_2}\mathbf{H}_{B_2} = \mathbf{P}^{\cap\cap}_{A_3}\mathbf{H}_{A_3} = \mathbf{P}^{\cap\cap}_{B_3}\mathbf{H}_{B_3} \Rightarrow \mathbf{P}^{\cap\cap}_{A_2}\mathbf{F}_{A_2} = \mathbf{P}^{\cap\cap}_{B_2}\mathbf{F}_{B_2} \tag{34}$$

where the last step follows from the unqualified path $\{\{A_2, B_3\}, \{B_3, A_1\}, \{A_1, B_2\}\}$ and (13).

Consider qualified edge $\{A_2, B_2\}$ and identify the noise overlap through $\mathbf{P}_{A_2}, \mathbf{P}_{B_2}$ of rank $\alpha_{A_2B_2}$ so that $\mathbf{P}_{A_2}\mathbf{H}_{A_2} = \mathbf{P}_{B_2}\mathbf{H}_{B_2}$. Then we have

$$\mathcal{R}(\mathbf{P}^{\cap\cap}_{A_2}) \subset \mathcal{R}(\mathbf{P}_{A_2}), \quad \mathcal{R}(\mathbf{P}^{\cap\cap}_{B_2}) \subset \mathcal{R}(\mathbf{P}_{B_2}) \tag{35a}$$
$$\Rightarrow L \overset{(8)}{=} r(\mathbf{P}_{A_2}\mathbf{F}_{A_2} - \mathbf{P}_{B_2}\mathbf{F}_{B_2}) \tag{35b}$$

$$\overset{(34)(35a)}{\leq} \alpha_{A_2B_2} - \alpha_{A_1B_1A_2B_2A_3B_3} \tag{35c}$$

$$\leq \alpha_{A_2B_2} - (\alpha_{A_1B_1A_2B_2A_3} + \alpha_{A_3B_3} - N) \tag{35d}$$

$$\leq \alpha_{A_2B_2} - (\alpha_{A_1B_1A_2A_3} + \alpha_{A_2B_2A_3} - \alpha_{A_2A_3} + \alpha_{A_3B_3} - N) \tag{35e}$$

$$\overset{(33f)}{\leq} \alpha_{A_2B_2} - \Big((\alpha_{A_1B_1} + \alpha_{B_1A_2} + \alpha_{A_3B_1} - 2N) + (\alpha_{A_2B_2} + \alpha_{B_2A_3} - N) - (3N - (\alpha_{B_3A_4} + \alpha_{A_4B_4}) - L) + \alpha_{A_3B_3} - N \Big) \tag{35f}$$

$$= 7N - L - (\alpha_{A_1B_1} + \alpha_{B_1A_2} + \alpha_{A_3B_1} + \alpha_{B_2A_3} + \alpha_{B_3A_4} + \alpha_{A_4B_4} + \alpha_{A_3B_3}) \tag{35g}$$

$$\overset{(12)}{\leq} 7N - 8L \tag{35h}$$

$$\Rightarrow R_{\text{linear}} = L/(2N) \leq 7/18 \tag{35i}$$

where sub-modularity is repeatedly applied in (35d), (35e), (35f); in (35g), $\{v, u\}$ is a qualified edge in every α_{vu} term, so (12) can be applied to obtain (35h). The converse proof and thus the linear capacity proof of Theorem 3 is complete.

REFERENCES

[1] Z. Li and H. Sun, "On the linear capacity of conditional disclosure of secrets," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 3202–3207.

[2] Y. Gertner, Y. Ishai, E. Kushilevitz, and T. Malkin, "Protecting data privacy in private information retrieval schemes," in *Proc. 13th Annu. ACM Symp. Theory Comput. STOC*, 1998, pp. 151–160.

[3] T. Liu, V. Vaikuntanathan, and H. Wee, "Towards breaking the exponential barrier for general secret sharing," in *Proc. Annu. Int. Conf. Theory Appl. Cryptograph. Techn.* Cham, Switzerland: Springer, 2018, pp. 567–596.

[4] B. Applebaum, A. Beimel, O. Nir, and N. Peter, "Better secret sharing via robust conditional disclosure of secrets," in *Proc. 52nd Annu. ACM SIGACT Symp. Theory Comput.*, Jun. 2020, pp. 280–293.

[5] R. Gay, I. Kerenidis, and H. Wee, "Communication complexity of conditional disclosure of secrets and attribute-based encryption," in *Proc. Annu. Cryptol. Conf.* Cham, Switzerland: Springer, 2015, pp. 485–502.

[6] B. Applebaum, B. Arkis, P. Raykov, and P. N. Vasudevan, "Conditional disclosure of secrets: Amplification, closure, amortization, lower-bounds, and separations," in *Proc. Annu. Int. Cryptol. Conf.* Cham, Switzerland: Springer, 2017, pp. 727–757.

[7] S. Laur and H. Lipmaa, "A new protocol for conditional disclosure of secrets and its applications," in *Proc. Int. Conf. Appl. Cryptography Netw. Secur.* Cham, Switzerland: Springer, 2007, pp. 207–225.

[8] B. Applebaum and P. N. Vasudevan, "Placing conditional disclosure of secrets in the communication complexity universe," in *Proc. 10th Innov. Theor. Comput. Sci. Conf. (ITCS)*. Wadern, Germany: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018, pp. 1–38.

[9] T. Liu, V. Vaikuntanathan, and H. Wee, "Conditional disclosure of secrets via non-linear reconstruction," in *Proc. Annu. Int. Cryptol. Conf.* Cham, Switzerland: Springer, 2017, pp. 758–790.

[10] V. Vaikuntanathan, "Some open problems in information-theoretic cryptography," in *Proc. 37th IARCS Annu. Conf. Found. Softw. Technol. Theor. Comput. Sci. (FSTTCS)*. Wadern, Germany: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018, pp. 1–7.

[11] Z. Li and H. Sun, "On extremal rates of secure storage over graphs," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 4721–4731, 2023.

[12] Z. Li and H. Sun, "On extremal rates of storage over graphs," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2023, pp. 2039–2044.

[13] Z. Li and H. Sun, "Conditional disclosure of secrets: A noise and signal alignment approach," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 4052–4062, Jun. 2022.

[14] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of linear coding in network information flow," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2745–2759, Aug. 2005.

[15] S. A. Jafar, "Topological interference management through index coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 529–568, Jan. 2014.

[16] S. H. Dau, V. Skachek, and Y. M. Chee, "On the security of index coding with side information," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3975–3988, Jun. 2012.

[17] H. Sun, "Secure groupcast: Extra-entropic structure and linear feasibility," 2020, *arXiv:2006.05944*.

[18] B. Applebaum and B. Arkis, "On the power of amortization in secret sharing: D-uniform secret sharing and CDS with constant information rate," in *Proc. Theory Cryptography Conf.* Cham, Switzerland: Springer, 2018, pp. 317–344.

[19] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4075–4088, Jul. 2017.

[20] E. J. Lee and E. Abbe, "Two Shannon-type problems on secure multiparty computations," in *Proc. 52nd Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, 2014, pp. 1287–1293.

[21] D. Data, V. M. Prabhakaran, and M. M. Prabhakaran, "Communication and randomness lower bounds for secure computation," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3901–3929, Jul. 2016.

[22] Y. Zhou, H. Sun, and S. Fu, "On the randomness cost of linear secure computation," in *Proc. 53rd Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2019, pp. 1–6.

[23] Y. Zhao and H. Sun, "Expand-and-randomize: An algebraic approach to secure computation," *Entropy*, vol. 23, no. 11, p. 1461, Nov. 2021.

[24] H. Sun, "The capacity of anonymous communications," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3871–3879, Jun. 2019.

[25] B. Tahmasebi and M. A. Maddah-Ali, "Private sequential function computation," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 1667–1671.

[26] Z. Wang, K. Banawan, and S. Ulukus, "Private set intersection: A multimessage symmetric private information retrieval perspective," *IEEE Trans. Inf. Theory*, vol. 68, no. 3, pp. 2001–2019, Mar. 2022.

[27] Z. Wang, K. Banawan, and S. Ulukus, "Multi-party private set intersection: An information-theoretic approach," *IEEE J. Sel. Areas Inf. Theory*, vol. 2, no. 1, pp. 366–379, Mar. 2021.

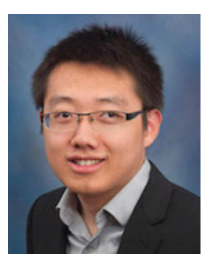
[28] H. Sun, "Compound secure groupcast: Key assignment for selected broadcasting," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 694–699.

[29] S. A. Jafar, "Interference alignment—A new look at signal dimensions in a communication network," *Found. Trends Commun. Inf. Theory*, vol. 7, no. 1, pp. 1–134, 2010, doi: 10.1561/0100000047.

[30] H. Sun and S. A. Jafar, "Index coding capacity: How far can one go with only Shannon inequalities?" *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3041–3055, Jun. 2015.

[31] S. Schechter, "On the inversion of certain matrices," *Math. Comput.*, vol. 13, no. 66, pp. 73–77, 1959.

Zhou Li (Member, IEEE) received the B.E. degree in communications engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2014. He is currently pursuing the Ph.D. degree with the University of North Texas. His research interests include information theory and privacy.



Hua Sun (Member, IEEE) received the B.E. degree in communications engineering from the Beijing University of Posts and Telecommunications, China, in 2011, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California at Irvine, USA, in 2013 and 2017, respectively.

He is currently an Associate Professor with the Department of Electrical Engineering, University of North Texas, USA. His research interests include information theory and its applications to communications, privacy, security, and storage. He is a recipient of the NSF CAREER Award in 2021, the UNT College of Engineering Junior Faculty Research Award in 2021, and the UNT College of Engineering Distinguished Faculty Fellowship in 2023. His coauthored papers received the IEEE Jack Keil Wolf ISIT Student Paper Award in 2016, an IEEE GLOBECOM Best Paper Award in 2016, and the 2020–2021 IEEE Data Storage Best Student Paper Award.