On the Security of Index Coding With Side Information

Son Hoang Dau, Vitaly Skachek, and Yeow Meng Chee, *Senior Member, IEEE*

Abstract*—Security aspects of the index coding with side information (ICSI) problem are investigated. Building on the results of Bar-Yossef *et al.* (2006), the properties of linear index codes are further explored. The notion of weak security, considered by Bhattad and Narayanan (2005) in the context of network coding, is generalized to *block security*. It is shown that the linear index code based on a matrix L, whose column space code $\mathcal{C}(L)$ has length n, minimum distance d, and dual distance $d^{\perp}$, is $(d-1-t)$-block secure (and hence also weakly secure) if the adversary knows in advance $t \le d-2$ messages, and is completely insecure if the adversary knows in advance more than $n-d^{\perp}$ messages. Strong security is examined under the conditions that the adversary: 1) possesses t messages in advance; 2) eavesdrops at most μ transmissions; 3) corrupts at most δ transmissions. We prove that for sufficiently large q, an optimal linear index code which is strongly secure against such an adversary has length $\kappa_q+\mu+2\delta$. Here, κ_q is a generalization of the min-rank over $\mathbb{F}_q$ of the side information graph for the ICSI problem in its original formulation in the work of Bar-Yossef *et al.

***Index Terms*—Index coding, network coding, side information, strong security, weak security.**

I. Introduction

The problem of index coding with side information (ICSI) was introduced by Birk and Kol [1], [2]. It was motivated by applications such as audio and video-on-demand, and daily newspaper delivery. In these applications, a server (sender) has to deliver some sets of data, audio or video files to a set of clients (receivers), different sets are requested by different receivers. Assume that before the transmission starts, the receivers have already (from previous transmissions) some files or movies in their possession. Via a slow backward channel, the receivers can let the sender know which messages they already have in their possession, and which messages they request. By exploiting this information, the amount of the overall transmissions can be reduced. As was observed in [1], this can be achieved by coding the messages at the server before broadcasting them out.

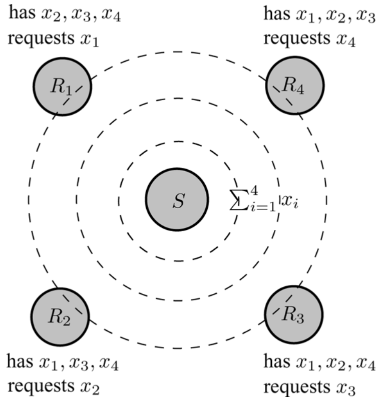

Fig. 1. Example of the ICSI problem.

Another possible application of the ICSI problem is in opportunistic wireless networks. These are networks in which a wireless node can opportunistically listen to the wireless channel. As a result, the node may obtain packets that were not designated to it [3]–[5]. This way, a node obtains some side information about the transmitted data. Exploiting this additional knowledge may help to increase the throughput of the system.

Consider the toy example in Fig. 1. It presents a scenario with one sender and four receivers. Each receiver requires a different information packet (or message). The naïve approach requires four separate transmissions, one transmission per an information packet. However, by exploiting the knowledge of the subsets of messages that clients already have, and by using coding of the transmitted data, the server can satisfy all the demands by broadcasting just one coded packet.

The ICSI problem has been a subject of several recent studies [3], [6]–[12]. This problem can be regarded as a special case of the well-known network coding (NC) problem [13], [14]. In particular, it was shown that every instance of the NC problem can be reduced to an instance of the ICSI problem [3], [10].

Several previous works focused on the design of an efficient index code for the ICSI problem. Given an instance of the ICSI problem, Bar-Yossef *et al.* [6] proved that finding the best scalar linear binary index code is equivalent to finding the so-called min-rank of a graph, which is known to be an NP-hard problem [6], [15]. Here, scalar linear index codes refer to linear index codes in which each message is a symbol in the field $\mathbb{F}_q$. By contrast, in vector linear index codes, each message is a vector over $\mathbb{F}_q$. Lubetzky and Stav [7] showed that there exist instances in which scalar linear index codes over nonbinary fields and

Manuscript received February 14, 2011; revised September 12, 2011; accepted December 21, 2011. Date of publication March 09, 2012; date of current version May 15, 2012. This work was supported by the National Research Foundation of Singapore under Research Grant NRF-CRP2-2007-03. The material in this paper was presented in part at the 2011 IEEE International Symposium on Information Theory.

S. H. Dau and Y. M. Chee are with the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 637371 Singapore (e-mail: daus0002@e.ntu.edu.sg; ymchee@ntu.edu.sg).

V. Skachek is with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC H3A 2A7, Canada (e-mail: vitaly.skachek@gmail.com). The work of this author was done in part while he was with the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 637371, Singapore.

Communicated by C. Fragouli, Associate Editor for Communication Networks.

Digital Object Identifier 10.1109/TIT.2012.2188777

linear index codes over mixed fields outperform the scalar linear binary index codes. El Rouayheb *et al.* [3], [10] showed that for certain instances of the ICSI problem, vector linear index codes achieve strictly higher transmission rate than scalar linear index codes do. They also pointed out that there exist instances in which vector nonlinear index codes outperform vector linear index codes. Vector nonlinear index codes were also shown to outperform scalar nonlinear index codes for certain instances by Alon *et al.* [12]. Several heuristic solutions for the ICSI problem were proposed in [9] and [11].

In this paper, we study the security aspects of linear index codes. We restrict ourselves to *scalar linear* index codes. It is known that vector linear index codes can achieve better transmission rate than their scalar counterparts, for certain instances of the ICSI problem [3], [10]. However, if the block length is fixed, one can model a vector index code as a scalar index code applied to another instance of the ICSI problem. If the block length is ℓ, the number of messages is n, and the number of receivers is m in the original (vector) instance, then the equivalent (scalar) instance can be viewed as having ℓn messages and ℓm receivers.

Let $\mathbb{F}_q$ be a finite field with q elements. A linear index code maps $\boldsymbol{x} \in \mathbb{F}_q^n$ onto $\boldsymbol{x}\boldsymbol{L}$, where $\boldsymbol{L}$ is an $n \times N$ matrix over $\mathbb{F}_q$, and $n,\ N \in \mathbb{N}$. In this paper, we show that each linear index code provides a certain level of information security. More specifically, let the code $\mathcal{C}(\boldsymbol{L})$ be spanned by the columns of $\boldsymbol{L}$, and let d and $d^{\perp}$ be its minimum distance and dual distance, respectively. We say that a particular adversary is of strength t if it has t messages in its possession. Then, we show that the index code based on $\boldsymbol{L}$ is $(d-1-t)$-block secure against all adversaries of strength $t \le d-2$ and is completely insecure against any adversary of strength at least $n-d^{\perp}+1$. If $\mathcal{C}(\boldsymbol{L})$ is an MDS code, then the two bounds coincide. The technique used in the proof for this result is reminiscent of that used in the constructions of (multiple) secret sharing schemes from linear error-correcting codes [16], [17]. The results on the security of linear index codes can be further employed to analyze the existence of solutions for a natural generalization of the ICSI problem, so-called the index coding with side and restricted information (ICSRI) problem. In that problem, it is required that some receivers have no information about some messages.

In the sequel, we also consider a linear randomized index code, which is based on the use of random symbols. We show that the coset coding technique (which has been successfully employed in secure NC literature; see, for instance, [18]–[22]) yields an optimal strongly secure linear randomized index code of length $\kappa_q + \mu + 2\delta$. This randomized index code is strongly secure against an adversary which

(i) has t arbitrary messages in advance;
(ii) eavesdrops at most μ transmissions;
(iii) corrupts at most δ transmissions.

Here, κ_q denotes the min-rank over $\mathbb{F}_q$ of the side information graph corresponding to the instance of the index coding (IC) problem.

Most of previous works on the security aspects (and on the error-correction aspect, as a special case) of network coding dealt with the multicast scenario. One of the main reasons for this limitation is that the optimal simultaneous transmission rates for nonmulticast networks have not been fully characterized yet. The ICSI problem can be modeled as a special case of the nonmulticast NC problem [3], [12]. Moreover, being modeled in that way, it requires that there are directed edges from particular sources to each sink, which provide the side information. The symbols transmitted on these special edges are not allowed to be corrupted, where usually for NC any edge can be corrupted. These two differences restrict the ability to derive the results on the security of the IC schemes from the existing results on security of NC schemes.

II. Preliminaries

Recall that we use the notation $\mathbb{F}_q$ for the finite field with q elements, where q is a power of prime. We also use $\mathbb{F}_q^*$ for the set of all nonzero elements of $\mathbb{F}_q$. Let $[n]$ denote the set of integers $\{1, 2, \ldots, n\}$. For the vectors $\boldsymbol{u} = (u_1, u_2, \ldots, u_n) \in \mathbb{F}_q^n$ and $\boldsymbol{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{F}_q^n$, the (Hamming) distance between $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined to be the number of coordinates where $\boldsymbol{u}$ and $\boldsymbol{v}$ differ, namely

$$\mathsf{d}(\boldsymbol{u}, \boldsymbol{v}) = |\{i \in [n] \ : \ u_i \neq v_i\}|.$$

The *support* of a vector $\boldsymbol{u} \in \mathbb{F}_q^n$ is defined to be the set $\operatorname{supp}(\boldsymbol{u}) = \{i \in [n] : u_i \neq 0\}$. The (Hamming) weight of a vector $\boldsymbol{u}$, denoted $\mathsf{wt}(\boldsymbol{u})$, is defined to be $|\operatorname{supp}(\boldsymbol{u})|$, the number of nonzero coordinates of $\boldsymbol{u}$.

A k-dimensional subspace $\mathcal{C}$ of $\mathbb{F}_q^n$ is called a linear $[n, k, d]_q$ (q-ary) code if the minimum distance of $\mathcal{C}$

$$\mathsf{d}(\mathcal{C}) \overset{\triangle}{=} \min_{\boldsymbol{u} \in \mathcal{C},\ \boldsymbol{v} \in \mathcal{C},\ \boldsymbol{u} \neq \boldsymbol{v}} \mathsf{d}(\boldsymbol{u}, \boldsymbol{v})$$

is equal to d. Sometimes, we may use the notation $[n, k]_q$ for the sake of simplicity. The vectors in $\mathcal{C}$ are called codewords. It is easy to see that the minimum weight of a nonzero codeword in a linear code $\mathcal{C}$ is equal to its minimum distance $\mathsf{d}(\mathcal{C})$. A generator matrix G of an $[n, k]_q$-code $\mathcal{C}$ is a $k \times n$ matrix whose rows are linearly independent codewords of $\mathcal{C}$. Then, $\mathcal{C} = \{\boldsymbol{y}\mathsf{G} : \boldsymbol{y} \in \mathbb{F}_q^k\}$.

The *dual code* or *dual space* of $\mathcal{C}$ is defined as $\mathcal{C}^{\perp} = \{\boldsymbol{u} \in \mathbb{F}_q^n : \ \boldsymbol{u}\boldsymbol{c}^T = 0 \text{ for all } \boldsymbol{c} \in \mathcal{C}\}$. The minimum distance of $\mathcal{C}^{\perp}$, $\mathsf{d}(\mathcal{C}^{\perp})$, is called the dual distance of $\mathcal{C}$.

The following upper bound on the minimum distance of a q-ary linear code is well known (see [23, Ch. 1]).

Theorem 2.1 (Singleton Bound): For an $[n, k, d]_q$-code, we have $d \le n - k + 1$.

Codes attaining this bound are called maximum distance separable (MDS) codes. For a subset of vectors

$$\{\boldsymbol{c}^{(1)}, \boldsymbol{c}^{(2)}, \ldots, \boldsymbol{c}^{(k)}\} \subseteq \mathbb{F}_q^n$$

define its linear span over $\mathbb{F}_q$

$$\mathsf{span}_q\left(\{\boldsymbol{c}^{(1)}, \boldsymbol{c}^{(2)}, \ldots, \boldsymbol{c}^{(k)}\}\right) \overset{\triangle}{=} \left\{ \sum_{i=1}^{k} \alpha_i \boldsymbol{c}^{(i)} \ : \ \alpha_i \in \mathbb{F}_q,\ i \in [k] \right\}.$$

We use $\boldsymbol{e}_i = (\underbrace{0,\dots,0}_{i-1},1,\underbrace{0,\dots,0}_{n-i}) \in \mathbb{F}_q^n$ to denote the unit vector, which has a one at the ith position, and zeros elsewhere. We also use $\mathbf{I}_n$, $n \in \mathbb{N}$, to denote the $n \times n$ identity matrix.

We recall the following well-known result in coding theory.

Theorem 2.2 ([24, p. 66]): Let $\mathcal{C}$ be an $[n,k,d]_q$-code with dual distance $d^{\perp}$ and $\boldsymbol{M}$ denote the $q^k \times n$ matrix whose q^k rows are codewords of $\mathcal{C}$. If $r \le d^{\perp}-1$, then each r-tuple from $\mathbb{F}_q$ appears in an arbitrary set of r columns of $\boldsymbol{M}$ exactly q^{k-r} times.

For a random vector $\boldsymbol{Y} = (Y_1, Y_2, \dots, Y_n)$ and a subset $B = \{i_1, i_2, \dots, i_b\}$ of $[n]$, where $i_1 < i_2 < \cdots < i_b$, let $\boldsymbol{Y}_B$ denote the vector $(Y_{i_1}, Y_{i_2}, \dots, Y_{i_b})$. For an $n \times k$ matrix $\boldsymbol{L}$, let $\boldsymbol{L}_i$ denote the ith row of $\boldsymbol{L}$, and $\boldsymbol{L}[j]$ its jth column. For a set $E \subseteq [n]$, let $\boldsymbol{L}_E$ denote the $|E| \times k$ submatrix of $\boldsymbol{L}$ formed by rows of $\boldsymbol{L}$ which are indexed by the elements of E. For a set $F \subseteq [k]$, let $\boldsymbol{L}[F]$ denote the $n \times |F|$ submatrix of $\boldsymbol{L}$ formed by columns of $\boldsymbol{L}$ which are indexed by the elements of F.

Let X and Y be discrete random variables taking values in the sets Σ_X and Σ_Y, respectively. Let $\Pr(X = x)$ denote the probability that X takes a particular value $x \in \Sigma_X$. Let $\mathsf{H}(X)$, $\mathsf{H}(X|Y)$, $\mathsf{I}(X;Y)$, and $\mathsf{I}(X;Y|Z)$ denote the (binary) entropy, conditional entropy, mutual information, and conditional mutual information (see [25] for the background).

III. Index Coding and Some Basic Results

The ICSI problem considers the following communications scenario. There is a unique sender (or source) S, who has a vector of messages $\boldsymbol{x} = (x_1, x_2, \dots, x_n) \in \mathbb{F}_q^n$ in his possession, which is a realized value of a random vector $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$. $X_1, X_2, \dots, X_n$, hereafter, are assumed to be independent uniformly distributed random variables over $\mathbb{F}_q$. There are also m receivers $R_1, R_2, \dots, R_m$. For each $i \in [m]$, R_i has some side information, i.e., R_i owns a subset of messages $\{x_j\}_{j \in \mathcal{X}_i}$, $\mathcal{X}_i \subseteq [n]$. In addition, each R_i, $i \in [m]$, is interested in receiving the message $x_{f(i)}$, for some demand function $f : [m] \to [n]$. Here, we assume that $f(i) \notin \mathcal{X}_i$ for all $i \in [m]$. Let $\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_m)$. An instance of the ICSI problem is given by a quadruple $(m, n, \mathcal{X}, f)$. Here, we assume that every receiver requests exactly one message. This assumption is not a limitation of the model, as we can consider an equivalent problem by splitting each receiver who requests multiple messages into multiple receivers, each of whom requests exactly one message and have the same set of side information [1], [6].

Definition 3.1: An *index code* over $\mathbb{F}_q$ for an instance $(m, n, \mathcal{X}, f)$ of the ICSI problem, referred to as an $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$, is an encoding function

$$\mathfrak{E} : \mathbb{F}_q^n \to \mathbb{F}_q^N$$

such that for each receiver R_i, $i \in [m]$, there exists a decoding function

$$\mathfrak{D}_i : \mathbb{F}_q^N \times \mathbb{F}_q^{|\mathcal{X}_i|} \to \mathbb{F}_q$$

satisfying

$$\forall \boldsymbol{x} \in \mathbb{F}_q^n \ : \ \mathfrak{D}_i(\mathfrak{E}(\boldsymbol{x}), \boldsymbol{x}_{\mathcal{X}_i}) = x_{f(i)}.$$

The parameter N is called the *length* of the index code. In the scheme corresponding to this code, S broadcasts a vector $\mathfrak{E}(\boldsymbol{x})$ of length N over $\mathbb{F}_q$.

Definition 3.2: An index code of the shortest possible length is called *optimal.*

Definition 3.3: A linear index code is an index code, for which the encoding function $\mathfrak{E}$ is a linear transformation over $\mathbb{F}_q$. Such a code can be described as

$$\forall \boldsymbol{x} \in \mathbb{F}_q^n \ : \ \mathfrak{E}(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{L}$$

where $\boldsymbol{L}$ is an $n \times N$ matrix over $\mathbb{F}_q$. The matrix $\boldsymbol{L}$ is called the matrix corresponding to the index code $\mathfrak{E}$. We also refer to $\mathfrak{E}$ as the index code based on $\boldsymbol{L}$. Notice that the length of $\mathfrak{E}$ is the number of columns of $\boldsymbol{L}$.

Let $E \subseteq [n]$ and $\boldsymbol{u} \in \mathbb{F}_q^n$. In the sequel, we write $\boldsymbol{u} \lhd E$ if $\mathsf{supp}(\boldsymbol{u}) \subseteq E$. Intuitively, this means that if some receiver knows x_j for all $j \in E$ (and also knows $\boldsymbol{u}$), then this receiver is also able to compute the value of $\boldsymbol{x}\boldsymbol{u}^T$.

Hereafter, we assume that the sets $\mathcal{X}_i$, for all $i \in [m]$, are known to S. Moreover, we also assume that the index code $\mathfrak{E}$ is known to each receiver R_i, $i \in [m]$. In practice, this can be achieved by a preliminary communication session, when the knowledge of the sets $\mathcal{X}_i$, for all $i \in [m]$, and of the code $\mathfrak{E}$ are disseminated between the participants of the scheme.

In [6], for the case $m = n$ and $f(i) = i$ for all $i \in [n]$, the *side information graph* $\mathcal{G}$ of an instance $(m, n, \mathcal{X}, f)$ of the ICSI problem is defined by $\mathcal{G} = (\mathcal{V}_{\mathcal{G}}, \mathcal{E}_{\mathcal{G}})$, where $\mathcal{V}_{\mathcal{G}} = [n]$ and

$$\mathcal{E}_{\mathcal{G}} = \{e = (i,j) : i, j \in [n], \ j \in \mathcal{X}_i\}.$$

A matrix $\boldsymbol{A}$ over $\mathbb{F}_q$ is said to *fit* $\mathcal{G}$ ([26]) if

$$\begin{cases} a_{i,j} \neq 0, & \text{if } i = j, \\ a_{i,j} = 0, & \text{if } i \neq j, \ (i,j) \notin \mathcal{E}_{\mathcal{G}}. \end{cases} \tag{1}$$

Then, the *min-rank* over $\mathbb{F}_q$ of the side information graph $\mathcal{G}$ is defined by

$$\min\{\mathsf{rank}_q(\boldsymbol{A}) : \boldsymbol{A} \text{ fits } \mathcal{G}\}. \tag{2}$$

Let $\mathcal{C}(\boldsymbol{L}) = \mathsf{span}_q(\{\boldsymbol{L}[j]^T\}_{j \in [N]})$, the subspace spanned by the (transposed) columns of $\boldsymbol{L}$. The following lemma was implicitly formulated in [6] for the case where $m = n$, $f(i) = i$ for all $i \in [m]$, and $q = 2$. This lemma specifies a sufficient condition on $\mathcal{C}(\boldsymbol{L})$ so that a receiver can reconstruct a particular message. We reproduce this lemma with its proof in its general form for the sake of completeness of the presentation.

Lemma 3.1: Let $\boldsymbol{L}$ be an $n \times N$ matrix over $\mathbb{F}_q$. Assume that S broadcasts $\boldsymbol{x}\boldsymbol{L}$. Then, for each $i \in [m]$, the receiver R_i can reconstruct $x_{f(i)}$ if there exists a vector $\boldsymbol{u}^{(i)} \in \mathbb{F}_q^n$ satisfying

$$\boldsymbol{u}^{(i)} \lhd \mathcal{X}_i \tag{3}$$

$$\boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)} \in \mathcal{C}(\boldsymbol{L}). \tag{4}$$

Proof: Assume that $\boldsymbol{u}^{(i)} \lhd \mathcal{X}_i$ and $\boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)} \in \mathcal{C}(\boldsymbol{L})$. Since $\boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)} \in \mathcal{C}(\boldsymbol{L})$, there exists $\beta \in \mathbb{F}_q^N$ such that

$$\boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)} = \beta \boldsymbol{L}^T.$$

By taking the transpose and premultiplying by $\boldsymbol{x}$, we obtain

$$\boldsymbol{x}(\boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)})^T = (\boldsymbol{x}\boldsymbol{L})\beta^T.$$

Therefore

$$x_{f(i)} = \boldsymbol{x}\boldsymbol{e}_{f(i)}^T = (\boldsymbol{x}\boldsymbol{L})\beta^T - \boldsymbol{x}\boldsymbol{u}^{(i)^T}.$$

Observe that R_i is able to find $\boldsymbol{u}^{(i)}$ and β from the knowledge of $\boldsymbol{L}$. Moreover, R_i is also able to compute $\boldsymbol{x}\boldsymbol{u}^{(i)^T}$ since $\boldsymbol{u}^{(i)} \lhd \mathcal{X}_i$. Additionally, R_i knows $\boldsymbol{x}\boldsymbol{L}$, which is transmitted by S. Therefore, R_i is able to compute $x_{f(i)}$. ■

Remark 3.2: It follows from Lemma 3.1 that $\boldsymbol{L}$ corresponds to a linear $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$ if $\mathcal{C}(\boldsymbol{L}) \supseteq \mathsf{span}_q(\{\boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)}\}_{i \in [m]})$, for some $\boldsymbol{u}^{(i)} \lhd \mathcal{X}_i$, $i \in [m]$. We show later in Corollary 4.5 that this condition is also necessary. Finding such an $\boldsymbol{L}$ with minimal number of columns by careful selection of $\boldsymbol{u}^{(i)}$'s is a difficult task (in fact it is NP-hard to do so, see [6], [15]), which, however, yields a linear coding scheme with the minimal number of transmissions.

IV. Block Secure Linear Index Codes

A. Block Security and Weak Security

In this section, we assume the presence of an adversary A who can listen to all transmissions. Assume that S employs a linear index code based on $\boldsymbol{L}$. The adversary is assumed to possess side information $\{x_j\}_{j \in \mathcal{X}_A}$, where $\mathcal{X}_A \subsetneq [n]$. For short, we say that A knows (or possesses, owns) $\boldsymbol{x}_{\mathcal{X}_A}$. The *strength* of A is defined to be $|\mathcal{X}_A|$. Denote $\widehat{\mathcal{X}}_A \overset{\triangle}{=} ([n] \backslash \mathcal{X}_A)$. Note that by listening to S, the adversary also knows $\boldsymbol{s} \overset{\triangle}{=} \mathfrak{E}(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{L}$. We define in the following several levels of security for linear index codes.

Definition 4.1: Suppose that the sender S possesses a vector of messages $\boldsymbol{x} \in \mathbb{F}_q^n$, which is a realized value of a random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, whose coordinates X_i, $i \in [n]$, are all independent and uniformly distributed over $\mathbb{F}_q$. An adversary A possesses $\boldsymbol{x}_{\mathcal{X}_A}$. Consider a linear $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$ based on $\boldsymbol{L}$.

1) For $B \subseteq \widehat{\mathcal{X}}_A$, the adversary is said to *have no information about* $\boldsymbol{x}_B$ if

$$\mathsf{H}(\boldsymbol{X}_B | \boldsymbol{X}\boldsymbol{L}, \boldsymbol{X}_{\mathcal{X}_A}) = \mathsf{H}(\boldsymbol{X}_B). \tag{5}$$

In other words, despite the partial knowledge on $\boldsymbol{x}$ that the adversary has (his side information and the transmissions he eavesdrops), the symbols $\boldsymbol{x}_B$ still looks completely random to him.

2) The index code is said to be *b-block secure against $\mathcal{X}_A$* if for every b-subset $B \subseteq \widehat{\mathcal{X}}_A$, the adversary has no information about $\boldsymbol{x}_B$.

3) The index code is said to be *b-block secure against all adversaries of strength* t $(0 \le t \le n-1)$ if it is b-block secure against $\mathcal{X}_A$ for every $\mathcal{X}_A \subset [n]$, $|\mathcal{X}_A| = t$.

4) The index code is said to be *weakly secure against $\mathcal{X}_A$* if it is 1-block secure against $\mathcal{X}_A$. In other words, after listening to all transmissions, the adversary has no information about each particular message that he does not possess in the first place.

5) The index code is said to be *weakly secure against all adversaries of strength* t $(0 \le t \le n-1)$ if it is weakly secure against $\mathcal{X}_A$ for every t-subset $\mathcal{X}_A$ of $[n]$.

6) The index code is said to be *completely insecure against* $\mathcal{X}_A$ if an adversary, who possesses $\{x_i\}_{i \in \mathcal{X}_A}$, by listening to all transmissions, is able to determine x_i for all $i \in \widehat{\mathcal{X}}_A$.

7) The index code is said to be *completely insecure against any adversary of strength* t $(0 \le t \le n-1)$ if an adversary, who possesses an arbitrary set of t messages, is always able to reconstruct all of the other $n-t$ messages after listening to all transmissions.

Remark 4.1: Even when the index code is b-block secure $(b \ge 1)$ as defined previously, the adversary is still able to obtain information about dependences between various x_i's in $\widehat{\mathcal{X}}_A$ (but he gains no information about any group of b particular messages). This definition of b-block security is a generalization of that of weak security (see [27], [28]). Obviously, if an index code is b-block secure against $\mathcal{X}_A (b \ge 1)$, then it is also weakly secure against $\mathcal{X}_A$, but the converse is not always true.

B. Necessary and Sufficient Conditions for Block Security

In the sequel, we consider the sets $B \subseteq [n]$, $B \ne \emptyset$, and $E \subseteq [n]$, $E \ne \emptyset$. Moreover, we assume that the sets $\mathcal{X}_A$, B, and E are disjoint, and that they form a partition of $[n]$, namely $\mathcal{X}_A \cup B \cup E = [n]$. In particular, $\widehat{\mathcal{X}}_A = B \cup E$.

Lemma 4.2: Assume that for all $\boldsymbol{u} \lhd \mathcal{X}_A$ and for all $\alpha_i \in \mathbb{F}_q$, $i \in B$ (not all α_i's are zeros)

$$\boldsymbol{u} + \sum_{i \in B} \alpha_i \boldsymbol{e}_i \notin \mathcal{C}(\boldsymbol{L}). \tag{6}$$

Then, we have the following.

1) For all $i \in B$

$$\boldsymbol{L}_i \in \mathsf{span}_q(\{\boldsymbol{L}_j\}_{j \in E}). \tag{7}$$

2) The system

$$\boldsymbol{y}\boldsymbol{L}_E = \boldsymbol{w}\boldsymbol{L}_B \tag{8}$$

has at least one solution $\boldsymbol{y} \in \mathbb{F}_q^{|E|}$ for every choice of $\boldsymbol{w} \in \mathbb{F}_q^{|B|}$.

Proof:

1) If $\mathsf{rank}_q(\boldsymbol{L}_E) = N$, then the first claim follows immediately. Otherwise, assume that $\mathsf{rank}_q(\boldsymbol{L}_E) < N$. As the N columns of $\boldsymbol{L}_E$ are linearly dependent, there exists $\boldsymbol{y} \in \mathbb{F}_q^N \backslash \{\boldsymbol{0}\}$ such that $\boldsymbol{y}\boldsymbol{L}_E^T = \boldsymbol{0}$.

   a) If for all such $\boldsymbol{y}$ and for all $i \in B$ we have $\boldsymbol{y}\boldsymbol{L}_i^T = 0$, then $\boldsymbol{L}_i \in ((\mathsf{span}_q(\{\boldsymbol{L}_j\}_{j \in E}))^\perp)^\perp = \mathsf{span}_q(\{\boldsymbol{L}_j\}_{j \in E})$ for all $i \in B$.

   b) Otherwise, there exist $\boldsymbol{y} \in \mathbb{F}_q^N$ and $i \in B$ such that $\boldsymbol{y}\boldsymbol{L}_E^T = \boldsymbol{0}$ and $\boldsymbol{y}\boldsymbol{L}_i^T \ne 0$. Without loss of generality, assume that

$$\boldsymbol{L} = \left[\begin{array}{c} \boldsymbol{L}_{\mathcal{X}_A} \\ \hline \boldsymbol{L}_B \\ \hline \boldsymbol{L}_E \end{array}\right].$$

Let $\boldsymbol{c} = \boldsymbol{y}\boldsymbol{L}^T \in \mathcal{C}(\boldsymbol{L})$. Then

$$\boldsymbol{c} = (\boldsymbol{c}_{\mathcal{X}_A}|\boldsymbol{c}_B|\boldsymbol{c}_E) = \left(\boldsymbol{y}\boldsymbol{L}_{\mathcal{X}_A}^T|\boldsymbol{y}\boldsymbol{L}_B^T|\boldsymbol{y}\boldsymbol{L}_E^T\right).$$

Hence, $\boldsymbol{c}_B = \boldsymbol{y}\boldsymbol{L}_B^T \neq \boldsymbol{0}$ and $\boldsymbol{c}_E = \boldsymbol{y}\boldsymbol{L}_E^T = \boldsymbol{0}$. Let $\boldsymbol{u} = (\boldsymbol{c}_{\mathcal{X}_A}|\boldsymbol{0}|\boldsymbol{0}) \lhd \mathcal{X}_A$ and $\alpha_i = c_i$ for all $i \in B$. Then, α_i's are not all zero and $\boldsymbol{u} + \sum_{i\in B}\alpha_i\boldsymbol{e}_i = \boldsymbol{c} \in \mathcal{C}(\boldsymbol{L})$, which contradicts (6).

2) By (7), each row of $\boldsymbol{L}_B$ is a linear combination of rows of $\boldsymbol{L}_E$. Hence, $\boldsymbol{w}\boldsymbol{L}_B$ is also a linear combination of rows of $\boldsymbol{L}_E$. Therefore, (8) has at least one solution. ■

The following lemma generalizes the weak security to the *block security*.

Lemma 4.3: Let $\boldsymbol{L}$ be an $n \times N$ matrix over $\mathbb{F}_q$. Assume that S broadcasts $\boldsymbol{x}\boldsymbol{L}$. For a subset $B \subseteq \widehat{\mathcal{X}}_A$, an adversary A who owns $\boldsymbol{x}_{\mathcal{X}_A}$, after listening to all transmissions, has no information about $\boldsymbol{x}_B$ if and only if

$$\begin{aligned}&\forall \boldsymbol{u} \lhd \mathcal{X}_A,\ \forall \alpha_i \in \mathbb{F}_q \text{ with } \alpha_i, i \in B, \text{ not all zero}\\ &\boldsymbol{u} + \sum_{i\in B}\alpha_i\boldsymbol{e}_i \notin \mathcal{C}(\boldsymbol{L}).\end{aligned} \tag{9}$$

In particular, for each $i \in \widehat{\mathcal{X}}_A$, A has no information about x_i if and only if

$$\forall \boldsymbol{u} \lhd \mathcal{X}_A\ : \boldsymbol{u} + \boldsymbol{e}_i \notin \mathcal{C}(\boldsymbol{L}).$$

Proof: Assume that (9) holds. We need to show that $\mathsf{H}(\boldsymbol{X}_B|\boldsymbol{X}\boldsymbol{L}, \boldsymbol{X}_{\mathcal{X}_A}) = \mathsf{H}(\boldsymbol{X}_B)$. It suffices to show that for all $\boldsymbol{g} \in \mathbb{F}_q^{|B|}$

$$\Pr(\boldsymbol{X}_B = \boldsymbol{g}|\boldsymbol{X}\boldsymbol{L} = \boldsymbol{s},\ \boldsymbol{X}_{\mathcal{X}_A} = \boldsymbol{x}_{\mathcal{X}_A}) = \frac{1}{q^{|B|}} \tag{10}$$

where $\boldsymbol{s} = \boldsymbol{x}\boldsymbol{L}$ for some $\boldsymbol{x} \in \mathbb{F}_q^n$.

Consider the following linear system with the unknown $\boldsymbol{z} \in \mathbb{F}_q^n$:

$$\begin{cases}\boldsymbol{z}_B = \boldsymbol{g}\\ \boldsymbol{z}_{\mathcal{X}_A} = \boldsymbol{x}_{\mathcal{X}_A}\\ \boldsymbol{z}\boldsymbol{L} = \boldsymbol{s}\end{cases}$$

which is equivalent to

$$\begin{cases}\boldsymbol{z}_B = \boldsymbol{g}\\ \boldsymbol{z}_{\mathcal{X}_A} = \boldsymbol{x}_{\mathcal{X}_A}\\ \boldsymbol{z}_E\boldsymbol{L}_E = \boldsymbol{s} - \boldsymbol{g}\boldsymbol{L}_B - \boldsymbol{x}_{\mathcal{X}_A}\boldsymbol{L}_{\mathcal{X}_A}.\end{cases} \tag{11}$$

In order to prove that (10) holds, it suffices to show that for all choices of $\boldsymbol{g} \in \mathbb{F}_q^{|B|}$, (11) always has the same number of solutions $\boldsymbol{z}$. Notice that the number of solutions $\boldsymbol{z}$ of (11) is equal to the number of solutions $\boldsymbol{z}_E$ of

$$\boldsymbol{z}_E\boldsymbol{L}_E = \boldsymbol{s} - \boldsymbol{g}\boldsymbol{L}_B - \boldsymbol{x}_{\mathcal{X}_A}\boldsymbol{L}_{\mathcal{X}_A} \tag{12}$$

where $\boldsymbol{s}$, $\boldsymbol{g}$, and $\boldsymbol{x}_{\mathcal{X}_A}$ are known. For any $\boldsymbol{g} \in \mathbb{F}_q^{|B|}$, if (12) has a solution, then it has exactly $q^{|E|-\mathsf{rank}_q(\boldsymbol{L}_E)}$ different solutions. Therefore, it suffices to prove that (12) has at least one solution for every $\boldsymbol{g} \in \mathbb{F}_q^{|B|}$.

Since $\boldsymbol{s} = \boldsymbol{x}\boldsymbol{L}$, we have

$$\boldsymbol{x}_E\boldsymbol{L}_E = \boldsymbol{s} - \boldsymbol{x}_B\boldsymbol{L}_B - \boldsymbol{x}_{\mathcal{X}_A}\boldsymbol{L}_{\mathcal{X}_A}. \tag{13}$$

Subtracting (13) from (12) we obtain

$$(\boldsymbol{z}_E - \boldsymbol{x}_E)\boldsymbol{L}_E = (\boldsymbol{x}_B - \boldsymbol{g})\boldsymbol{L}_B$$

which can be rewritten as

$$\boldsymbol{y}\boldsymbol{L}_E = \boldsymbol{w}\boldsymbol{L}_B \tag{14}$$

where $\boldsymbol{y}\overset{\triangle}{=}\boldsymbol{z}_E - \boldsymbol{x}_E$, $\boldsymbol{w}\overset{\triangle}{=}\boldsymbol{x}_B - \boldsymbol{g}$. Due to Lemma 4.2, (14) always has a solution $\boldsymbol{y}$, for every choice of $\boldsymbol{w}$. Therefore, (12) has at least one solution for every $\boldsymbol{g} \in \mathbb{F}_q^{|B|}$.

Now we prove the converse. Assume that (9) does not hold. Then, there exist $\boldsymbol{u} \lhd \mathcal{X}_A$ and $\alpha_i \in \mathbb{F}_q$, $i \in B$, where α_i's, $i \in B$ are not all zero, such that

$$\sum_{i\in B}\alpha_i\boldsymbol{e}_i = \boldsymbol{c} - \boldsymbol{u}$$

for some $\boldsymbol{c} \in \mathcal{C}(\boldsymbol{L})$. Hence, similar to the proof of Lemma 3.1, the adversary obtains

$$\begin{aligned}\sum_{i\in B}\alpha_i x_i &= \boldsymbol{x}\left(\sum_{i\in B}\alpha_i\boldsymbol{e}_i\right)^T\\ &= \boldsymbol{x}(\boldsymbol{c} - \boldsymbol{u})^T\\ &= \boldsymbol{x}\boldsymbol{c}^T - \boldsymbol{x}\boldsymbol{u}^T.\end{aligned}$$

Note that the adversary can calculate $\boldsymbol{x}\boldsymbol{c}^T$ from $\boldsymbol{s}$, and can also find $\boldsymbol{x}\boldsymbol{u}^T$ based on his own side information. Therefore, A is able to compute a nontrivial linear combination of x_i's, $i \in B$. Hence, the entropy $\mathsf{H}(\boldsymbol{X}_B|\boldsymbol{X}\boldsymbol{L}, \boldsymbol{X}_{\mathcal{X}_A}) < \mathsf{H}(\boldsymbol{X}_B)$. Thus, the adversary gains some information about the $\boldsymbol{x}_B$. ■

Corollary 4.4 generalizes Lemma 3.1 by providing both necessary and sufficient conditions for a receiver's ability to recover the desired message. Equivalently, this corollary provides necessary and sufficient conditions for a receiver R_i (or the adversary A) to have no information about a particular message.

Corollary 4.4: Let $\boldsymbol{L}$ be an $n \times N$ matrix over $\mathbb{F}_q$ and let S broadcast $\boldsymbol{x}\boldsymbol{L}$. Then, for each $i \in [m]$, the receiver R_i can reconstruct $x_{f(i)}$ if and only if there exists a vector $\boldsymbol{u}^{(i)} \in \mathbb{F}_q^n$ satisfying (3) and (4). The receiver R_i has no information about $x_{f(i)}$ if there exists no vector $\boldsymbol{u}^{(i)} \in \mathbb{F}_q^n$ as earlier.

Proof: The proof of this lemma is straightforward from Lemma 3.1 and Lemma 4.3. ■

Corollary 4.4 can be reformulated as follows.

Corollary 4.5: The matrix $\boldsymbol{L}$ corresponds to a linear $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$ if and only if for all $i \in [m]$, there exists a vector $\boldsymbol{u}^{(i)} \in \mathbb{F}_q^n$ satisfying (3) and (4).

Remark 4.6: It follows from Corollary 4.5 that $\boldsymbol{L}$ corresponds to a linear $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$ if and only if $\mathcal{C}(\boldsymbol{L}) \supseteq \mathsf{span}_q(\{\boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)}\}_{i\in[m]})$, for some $\boldsymbol{u}^{(i)} \lhd \mathcal{X}_i$, $i \in [m]$. If we define

$$\begin{aligned}\kappa_q &= \kappa_q(m, n, \mathcal{X}, f)\\ &\overset{\triangle}{=} \min\{\mathsf{rank}_q(\{\boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)}\}_{i\in[m]}) : \boldsymbol{u}^{(i)} \in \mathbb{F}_q^n, \boldsymbol{u}^{(i)} \lhd \mathcal{X}_i\}\end{aligned} \tag{15}$$

then κ_q is the shortest possible length of a linear $(m,n,\mathcal{X},f)$-IC over $\mathbb{F}_q$.

Corollary 4.7: The length of an optimal linear $(m,n,\mathcal{X},f)$-IC over $\mathbb{F}_q$ is $\kappa_q = \kappa_q(m,n,\mathcal{X},f)$.

Remark 4.8: The quantity κ_q defined in (15) is precisely the min-rank over $\mathbb{F}_q$ of the side information graph of an ICSI instance in the case $m = n$ and $f(i) = i$ for all $i \in [n]$ (see also [6]).

Proof: Suppose that $\boldsymbol{u}^{(i)} \lhd \mathcal{X}_i$ for all $i \in [n]$. Let $\boldsymbol{A} = (a_{i,j})$ be the $n \times n$ matrix whose ith row is precisely $\boldsymbol{u}^{(i)} + \boldsymbol{e}_i$, for each $i \in [n]$. Then, $\boldsymbol{A}$ fits $\mathcal{G}$. Conversely, if $\boldsymbol{A}'$ fits $\mathcal{G}$, then by multiplying each row of $\boldsymbol{A}'$ with a suitable nonzero constant (which does not change the rank of $\boldsymbol{A}'$), one obtains a matrix $\boldsymbol{A}$ of the form (1). In other words, for each $i \in [n]$, the ith row of the resulting matrix $\boldsymbol{A}$ equals $\boldsymbol{u}^{(i)} + \boldsymbol{e}_i$ for some $\boldsymbol{u}^{(i)} \lhd \mathcal{X}_i$. Therefore, κ_q defined in (15) is indeed the minimum rank over $\mathbb{F}_q$ of a matrix which fits the side information graph $\mathcal{G}$. Thus, κ_q is precisely the min-rank over $\mathbb{F}_q$ of $\mathcal{G}$. ■

Theorem 4.9: Consider a linear $(m,n,\mathcal{X},f)$-IC over $\mathbb{F}_q$ based on $\boldsymbol{L}$. Let d be the minimum distance of $\mathcal{C}(\boldsymbol{L})$.

1) This index code is $(d-1-t)$-block secure against all adversaries of strength $t \leq d-2$. In particular, it is weakly secure against all adversaries of strength $t = d-2$.
2) This index code is not $(d-t)$-block secure against at least one adversary of strength t, for any $t \leq d-1$. In particular, it is not weakly secure against at least one adversary of strength $t = d-1$.
3) Every adversary of strength $t \leq d-1$ is able to find a list of q^{n-t-N} vectors in $\mathbb{F}_q^n$ which includes the vector of messages $\boldsymbol{x}$.

Proof:

1) Assume that $t \leq d-2$. By Lemma 4.3, it suffices to show that for every t-subset $\mathcal{X}_A$ of $[n]$ and for every $(d-1-t)$-subset B of $\widehat{\mathcal{X}}_A$
$$\forall \boldsymbol{u} \lhd \mathcal{X}_A,\ \forall \alpha_i \in \mathbb{F}_q \text{ with } \alpha_i, i \in B, \text{ not all zero}$$
$$\boldsymbol{u} + \sum_{i\in B} \alpha_i \boldsymbol{e}_i \notin \mathcal{C}(\boldsymbol{L}).$$
For such $\boldsymbol{u}$ and α_i's, we have $\mathsf{wt}(\boldsymbol{u} + \sum_{i\in B}\alpha_i\boldsymbol{e}_i) = \mathsf{wt}(\boldsymbol{u}) + \mathsf{wt}(\sum_{i\in B}\alpha_i\boldsymbol{e}_i) \leq t + (d-1-t) = d-1 < d$. Moreover, as $\mathsf{supp}(\boldsymbol{u}) \cap B = \emptyset$ and α_i's, $i \in B$, are not all zero, we deduce that $\boldsymbol{u} + \sum_{i\in B}\alpha_i\boldsymbol{e}_i \neq \boldsymbol{0}$. We conclude that $\boldsymbol{u} + \sum_{i\in B}\alpha_i\boldsymbol{e}_i \notin \mathcal{C}(\boldsymbol{L})$.
2) We now show that the index code is not $(d-t)$-block secure against at least one adversary of strength t, for any $t \leq d-1$.
Pick a codeword $\boldsymbol{c} = (c_1, c_2, \ldots, c_n) \in \mathcal{C}(\boldsymbol{L})$ such that $\mathsf{wt}(\boldsymbol{c}) = d$ and let $\mathsf{supp}(\boldsymbol{c}) = \{i_1, i_2, \ldots, i_d\}$. Without loss of generality assume that $\mathcal{X}_A = \{i_1, i_2, \ldots, i_t\}$, $|\mathcal{X}_A| = t$, $t \leq d-1$. Then, $B \stackrel{\triangle}{=} \{i_{t+1}, i_{t+2}, \ldots, i_d\} \subseteq \widehat{\mathcal{X}}_A$.
Let
$$\boldsymbol{u} = \boldsymbol{c} - \sum_{j=t+1}^{d} c_{i_j}\boldsymbol{e}_{i_j}.$$
Then, $\boldsymbol{u} \lhd \mathcal{X}_A$ and $\boldsymbol{u} + \sum_{j=t+1}^{d} c_{i_j}\boldsymbol{e}_{i_j} = \boldsymbol{c} \in \mathcal{C}(\boldsymbol{L})$. By Lemma 4.3, after listening to all transmissions, A gains some information about x_B, namely $\mathsf{H}(\boldsymbol{X}_B|\boldsymbol{X}\boldsymbol{L}, \boldsymbol{X}_{\mathcal{X}_A}) < \mathsf{H}(\boldsymbol{X}_B)$. Hence, the index code is not $(d-t)$-block secure against at least one adversary of strength t.
3) Let $\boldsymbol{s} = \boldsymbol{x}\boldsymbol{L}$. Consider the following linear system of equations with unknown $\boldsymbol{z} \in \mathbb{F}_q^n$:
$$\begin{cases} \boldsymbol{z}_{\mathcal{X}_A} = \boldsymbol{x}_{\mathcal{X}_A} \\ \boldsymbol{z}\boldsymbol{L} = \boldsymbol{s} \end{cases}$$
which is equivalent to
$$\begin{cases} \boldsymbol{z}_{\mathcal{X}_A} = \boldsymbol{x}_{\mathcal{X}_A} \\ \boldsymbol{z}_{\widehat{\mathcal{X}}_A}\boldsymbol{L}_{\widehat{\mathcal{X}}_A} = \boldsymbol{s} - \boldsymbol{x}_{\mathcal{X}_A}\boldsymbol{L}_{\mathcal{X}_A}. \end{cases} \tag{16}$$
The adversary A attempts to solve this system. Given that $\boldsymbol{s}$ and $\boldsymbol{x}_{\mathcal{X}_A}$ are known, the system (16) has $n-t$ unknowns and N equations. Note that $t \leq d-1$, and thus by applying Theorem 2.1 to $\mathcal{C}(\boldsymbol{L})$ we have $n - t \geq n - d + 1 \geq N$. If $\mathsf{rank}_q(\boldsymbol{L}_{\widehat{\mathcal{X}}_A}) = N$, then (16) has exactly q^{n-t-N} solutions, as required.
Next, we show that $\mathsf{rank}_q(\boldsymbol{L}_{\widehat{\mathcal{X}}_A}) = N$. Assume, by contrary, that the N columns of $\boldsymbol{L}_{\widehat{\mathcal{X}}_A}$, denoted by $\boldsymbol{c}^{(1)}, \boldsymbol{c}^{(2)}, \ldots, \boldsymbol{c}^{(N)}$, are linearly dependent. Then, there exist $\beta_i \in \mathbb{F}_q$, $i \in [N]$, not all zero, such that $\sum_{i=1}^{N}\beta_i\boldsymbol{c}^{(i)} = \boldsymbol{0}$. Let
$$\boldsymbol{c} = \sum_{i=1}^{N}\beta_i\boldsymbol{L}[i] \in \mathcal{C}(\boldsymbol{L})\backslash\{\boldsymbol{0}\}.$$
(Recall that $\boldsymbol{L}[i]$ denotes the ith column of $\boldsymbol{L}$). Then, $\boldsymbol{c}_{\widehat{\mathcal{X}}_A} = \sum_{i=1}^{N}\beta_i\boldsymbol{c}^{(i)} = \boldsymbol{0}$ and hence $\mathsf{wt}(\boldsymbol{c}) = \mathsf{wt}(\boldsymbol{c}_{\mathcal{X}_A}) \leq t \leq d-1$. This is a contradiction, which follows from the assumption that the N rows of $\boldsymbol{L}_{\widehat{\mathcal{X}}_A}$ are linearly dependent.

■

Example 4.1: Let $q = 2$. Assume that $\mathcal{X}_A = \emptyset$ and that $\mathcal{X}_i \neq \emptyset$ for all $i \in [m]$. For each $i \in [m]$, choose some $j_i \in \mathcal{X}_i$. Let $\boldsymbol{L}$ be the binary matrix whose columns form a basis of the space $\mathcal{C}(\boldsymbol{L}) = \mathsf{span}_q(\{\boldsymbol{e}_{j_i} + \boldsymbol{e}_{f(i)}\}_{i\in[m]})$. Then, $\mathsf{d}(\mathcal{C}(\boldsymbol{L})) = 2$. Since $t = |\mathcal{X}_A| = 0$, we have $d - 1 - t = 1$. Therefore, by Theorem 4.9, the index code based on $\boldsymbol{L}$ is weakly secure against A. By the Singleton bound, $\boldsymbol{L}$ has $N \leq n-d+1 = n-1$ columns. In other words, the index code based on $\boldsymbol{L}$ requires at most $n-1$ transmissions.

C. Block Security and Complete Insecurity

Theorem 4.9 provides a threshold for the security level of a linear index code based on $\boldsymbol{L}$. If A has a prior knowledge of any $t \leq d-2$ messages, where $d = \mathsf{d}(\mathcal{C}(\boldsymbol{L}))$, then the scheme is still secure, i.e., the adversary has no information about any group of $d-1-t$ particular messages from $\{x_j\}_{j\in\widehat{\mathcal{X}}_A}$. On the other hand, the scheme may no longer be secure against an adversary of strength $t = d-1$. The last assertion of Theorem 4.9 shows us the difference between being block secure and being strongly secure (the notion of strong security is rigorously defined in Definition 5.3 in the sequel). More specifically, if the scheme is strongly secure, the messages $\boldsymbol{x}_{\widehat{\mathcal{X}}_A}$, which are not leaked to the adversary in advance, look completely random to the adversary, i.e., the probability to guess them correctly is $1/q^{n-t}$. However, if the scheme is $(d-1-t)$-block secure (for $t \le d-2$), then the adversary is able to guess these messages correctly with probability $1/q^{n-t-N}$.

For an adversary of strength $t \ge d$, the security of the scheme depends on the properties of the code employed, in particular, it depends on the weight distribution of $\mathcal{C}(\boldsymbol{L})$. It is possible to show in the way analogous to the proof of part (2) in Theorem 4.9 that if there exists $\boldsymbol{c} \in \mathcal{C}(\boldsymbol{L})$ with $\mathsf{wt}(\boldsymbol{c}) = w$, then the scheme is not weakly secure against corresponding adversary of strength $t = w-1$. In general, the index code might still be (b-block or weakly) secure against some adversaries of strength t for $t \ge d$. While we cannot make a general conclusion on the security of the scheme when the adversary's strength is larger than $d-1$, Lemma 4.3 is still a useful tool to evaluate the security in that situation. However, as the next theorem shows, if the size of $\mathcal{X}_A$ is sufficiently large, then A is able to determine all the messages in $\{x_j\}_{j\in\widehat{\mathcal{X}}_A}$.

Theorem 4.10: The linear index code based on $\boldsymbol{L}$ is completely insecure against any adversary of strength $t \ge n - d^\perp + 1$, where $d^\perp$ denotes the dual distance of $\mathcal{C}(\boldsymbol{L})$.

Proof: Suppose the adversary knows a subset $\{x_j\}_{j\in\mathcal{X}_A}$, $\mathcal{X}_A \subsetneq [n]$ and $|\mathcal{X}_A| = t \ge n - d^\perp + 1$. By Corollary 4.4, it suffices to show that for all $j \in \widehat{\mathcal{X}}_A$, there exists $\boldsymbol{u} \in \mathbb{F}_q^n$ satisfying simultaneously $\boldsymbol{u} \lhd \mathcal{X}_A$ and $\boldsymbol{u} + \boldsymbol{e}_j \in \mathcal{C}(\boldsymbol{L})$.

Indeed, take any $j \in \widehat{\mathcal{X}}_A$, and let $\rho = n - t \le d^\perp - 1$. Consider the ρ indices which are not in $\mathcal{X}_A$. By Theorem 2.2, there exists a codeword $\boldsymbol{c} \in \mathcal{C}(\boldsymbol{L})$ with

$$c_\ell = \begin{cases} 1 & \text{if } \ell = j \\ 0 & \text{if } \ell \notin \mathcal{X}_A \cup \{j\}. \end{cases}$$

Then, $\mathrm{supp}(\boldsymbol{c}) \subseteq \mathcal{X}_A \cup \{j\}$. We define $\boldsymbol{u} \in \mathbb{F}_q^n$ such that $\boldsymbol{u} \lhd \mathcal{X}_A$, as follows. For $\ell \in \mathcal{X}_A$, we set $u_\ell = c_\ell$, and for $\ell \notin \mathcal{X}_A$, we set $u_\ell = 0$. It is immediately clear that $\boldsymbol{c} = \boldsymbol{u} + \boldsymbol{e}_j$. Therefore, by Corollary 4.4, the adversary can reconstruct x_j. We have shown that the index code is completely insecure against an arbitrary set $\mathcal{X}_A$ satisfying $|\mathcal{X}_A| \ge n - d^\perp + 1$, hence completing the proof. ■

When $\mathcal{C}(\boldsymbol{L})$ is an MDS code, we have $n - d^\perp + 1 = d - 1$, and hence, the two bounds established in Theorems 4.9 and 4.10 are actually tight. In that case, the third statement in Theorem 4.9 implies Theorem 4.10 as follows. This statement asserts that an adversary of strength $t = d - 1$ can find a list of $q^{n-N-d+1}$ vectors that includes the vector of messages $\boldsymbol{x}$. Since $\mathcal{C}(\boldsymbol{L})$ is an MDS code, we have $n - N - d + 1 = 0$. Therefore, the list contains only one element, namely $\boldsymbol{x}$ itself. Thus, the index code is completely insecure against any adversary of strength $d - 1$.

The following example further illustrates the results stated in these theorems.

Example 4.2: Let $n = 7$, $m = 7$, $q = 2$, and $f(i) = i$ for all $i \in [m]$. Suppose that the receivers have in their possession sets of messages as appear in the third column of the table that follows. Suppose also that the demands of all receivers are as in the second column of the table.

Receiver	Demand	$\{x_j\}_{i\in\mathcal{X}_i}$
R_1	x_1	$\{x_6, x_7\}$
R_2	x_2	$\{x_5, x_7\}$
R_3	x_3	$\{x_5, x_6\}$
R_4	x_4	$\{x_5, x_6, x_7\}$
R_5	x_5	$\{x_1, x_2, x_6\}$
R_6	x_6	$\{x_1, x_3, x_4\}$
R_7	x_7	$\{x_2, x_3, x_6\}$

For $i \in [7]$, let $\boldsymbol{u}^{(i)} \in \mathbb{F}_2^7$ such that $\mathrm{supp}(\boldsymbol{u}^{(i)}) = \mathcal{X}_i$. Assume that an index code based on $\boldsymbol{L}$ with $\mathcal{C}(\boldsymbol{L}) = \mathsf{span}_q(\{\boldsymbol{u}^{(i)} + \boldsymbol{e}_i\}_{i\in[7]})$ is used. For instance, we can take $\boldsymbol{L}$ to be the matrix whose set of columns is $\{\boldsymbol{L}[i] \triangleq \boldsymbol{u}^{(i)} + \boldsymbol{e}_i\}_{i\in[4]}$. It is easy to see that $\mathcal{C}(\boldsymbol{L})$ is a $[7,4,3]_2$ Hamming code with $d = 3$ and $d^\perp = 4$.

Following the coding scheme, S broadcasts the following four bits:

$$\begin{aligned} s_1 &= \boldsymbol{x}(\boldsymbol{u}^{(1)} + \boldsymbol{e}_1)^T \\ s_2 &= \boldsymbol{x}(\boldsymbol{u}^{(2)} + \boldsymbol{e}_2)^T \\ s_3 &= \boldsymbol{x}(\boldsymbol{u}^{(3)} + \boldsymbol{e}_3)^T \\ s_4 &= \boldsymbol{x}(\boldsymbol{u}^{(4)} + \boldsymbol{e}_4)^T. \end{aligned}$$

Each R_i, $i \in [7]$, can compute $\boldsymbol{x}(\boldsymbol{u}^{(i)} + \boldsymbol{e}_i)^T$ by using a linear combination of s_1, s_2, s_3, s_4. Then, each R_i can subtract $\boldsymbol{x}\boldsymbol{u}^{(i)T}$ (his side information) from $\boldsymbol{x}(\boldsymbol{u}^{(i)} + \boldsymbol{e}_i)^T$ to retrieve $x_i = \boldsymbol{x}\boldsymbol{e}_i^T$.

For example, consider R_5. Since

$$\boldsymbol{x}\left(\boldsymbol{u}^{(5)} + \boldsymbol{e}_5\right)^T = \boldsymbol{x}\left((\boldsymbol{u}^{(1)} + \boldsymbol{e}_1) + (\boldsymbol{u}^{(2)} + \boldsymbol{e}_2)\right)^T = s_1 + s_2$$

R_5 subtracts $x_1 + x_2 + x_6$ from $s_1 + s_2$ to obtain

$$\begin{aligned} &(s_1 + s_2) - (x_1 + x_2 + x_6) \\ &= (x_1 + x_2 + x_5 + x_6) - (x_1 + x_2 + x_6) \\ &= x_5. \end{aligned}$$

If an adversary A has a knowledge of a single message x_i, then by Theorem 4.9, A is not able to determine any other message x_ℓ, for $\ell \ne i$. Indeed, $\mathsf{d}(\mathcal{C}(\boldsymbol{L})) = 3$, while $t = 1$, so the code is weakly secure against all adversaries of strength $t = 1$. If none of the messages are leaked, then the adversary has no information about any group of two messages. On the other hand, the code is completely insecure against any adversary of strength $t \ge 4$; in that case, A is able to determine the remaining $7 - t$ messages.

Remark 4.11: So far we only discuss the case when the adversary can listen to all N transmissions. If we consider an adversary, which can eavesdrop at most μ ($\mu \le N$) messages, then analogous results can also be obtained. Consider a linear index code based on $\boldsymbol{L}$. Let

$$d_\mu \triangleq \min\left\{\mathsf{d}(\mathcal{C}(\boldsymbol{L}[W])) : W \subseteq [N],\ |W| = \mu\right\}$$

and

$$d_{\mu}^{\perp} \triangleq \min\left\{\mathsf{d}((\mathcal{C}(\boldsymbol{L}[W]))^{\perp}) : W \subseteq [N],\ |W| = \mu\right\}.$$

Then, it is straightforward to see that the results in Theorems 4.9 and 4.10 still hold, with d and $d^{\perp}$ being replaced by d_{μ} and $d_{\mu}^{\perp}$, respectively.

D. Role of the Field Size

The following example demonstrates that the use of index codes over larger fields might have a positive impact on the security level. More specifically, in that example, index codes over large fields significantly enhance the security, compared with index codes over small fields.

Example 4.3: Suppose that the source S has n messages $x_1, x_2, \ldots, x_n$. Assume that there are $m < n$ receivers $R_1, R_2, \ldots, R_m$, and each receiver R_i has the same set of side information, $\mathcal{X}_i = \{m+1, m+2, \ldots, n\}$. Assume also that each R_i requires x_i, for $i \in [m]$.

Any index code for this instance must have length at least m, since all the vectors $\boldsymbol{u}^{(i)} + \boldsymbol{e}_i$, for some $\boldsymbol{u}^{(i)} \lhd \mathcal{X}_i$, $i \in [m]$, are linearly independent over any field.

If we employ an index code over $\mathbb{F}_2$, by the fact that there are no nontrivial binary MDS codes, we deduce that the minimum distance d of $\mathcal{C}(\boldsymbol{L})$ is at most $n - m$. Hence, index codes over $\mathbb{F}_2$ are not secure against some adversaries of strength $t = n - m - 1$. However, if we consider index codes over $\mathbb{F}_q$ for sufficiently large $q (q \geq n - 1)$, there exists a q-ary MDS code $\mathcal{C}$ with minimum distance exactly $n - m + 1$. By choosing $\boldsymbol{L}$ so that $\mathcal{C}(\boldsymbol{L}) = \mathcal{C}$, the index code based on $\boldsymbol{L}$ is secure against all adversaries of strength at most $t = n - m - 1$, which is strictly more secure than the those over $\mathbb{F}_2$. To find such an $\boldsymbol{L}$, let $\boldsymbol{M} = (\boldsymbol{I}_m | \boldsymbol{P})$ be a generator matrix in standard form of an $[n, m]_q$-MDS code, and then take $\boldsymbol{L} = \boldsymbol{M}^T$. Then, $\boldsymbol{L}[i] = \boldsymbol{u}^{(i)} + \boldsymbol{e}^{(i)}$, for some $\boldsymbol{u}^{(i)} \lhd \mathcal{X}_i = \{m+1, m+2, \ldots, n\}$, $i \in [m]$. Therefore, by Corollary 4.5, $\boldsymbol{L}$ corresponds to a linear index code for this instance.

Note that if we employ an index code over $\mathbb{F}_2$, then for large values of n the minimum distance d of $\mathcal{C}(\boldsymbol{L})$ is bounded from earlier by the sphere-packing bound

$$d \leq 2n \cdot (\mathsf{H}^{-1}(1 - m/n) - \varepsilon)$$

where $\varepsilon \to 0$ as $n \to \infty$, and $\mathsf{H}^{-1}(\cdot)$ denotes the inverse of the binary entropy function.

There is a variety of stronger upper bounds on the minimum distance of binary codes, such as the Johnson bound, the Elias bound, and the McEliece–Rodemich–Rumsey–Welch bound (see [29, Ch. 4.5] for more details). These bounds provide even stronger bounds on the security of the binary scheme for this instance of the ICSI problem. By contrast, as shown previously, by using a q-ary MDS code, the distance d of $\mathcal{C}(\boldsymbol{L})$ can achieve the Singleton bound. It is well known that there is a significant gap between the Singleton bound and the sphere-packing bound (see [29, p. 111] for details). Therefore, for this instance of the ICSI problem, index codes over large fields provide significantly higher levels of security than those over binary field.

E. Application: ICSRI

In this section, we consider an extension of the ICSI problem, which we call the ICSRI problem. This problem arises in applications such as audio and video-on-demand. Consider a client who has subscribed for certain media content (audio or video programs, movies, newspapers, etc.) At the same time, this client has not subscribed to some other content. The content provider wants to restrict this client from obtaining a content which he is not eligible for, even though he might be able to obtain it "for free" from the transmissions provided by the server. As we show in sequel, the solution for the ICSRI problem is a straight-forward application of the results in Corollary 4.4.

More formally, the arguments of an instance $(m, n, \mathcal{X}, \mathcal{Z}, f)$ of the ICSRI problem are similar to their counterparts for the ICSI problem. The new additional parameter, $\mathcal{Z} = (\mathcal{Z}_1, \mathcal{Z}_2, \ldots, \mathcal{Z}_m)$, represents the sets $\mathcal{Z}_i \subseteq [n]$ of message indices that the respective receivers R_i, $i \in [m]$, are not allowed to obtain. The goal is that at the end of the communication round, the receiver R_i has the message $x_{f(i)}$ in its possession, for all $i \in [m]$, and it has no information about x_j for all $j \in \mathcal{Z}_i$. The notion of a linear $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$ is naturally extended to that of a linear $(m, n, \mathcal{X}, \mathcal{Z}, f)$-IC over $\mathbb{F}_q$.

Let

$$\mathcal{F}(m, n, \mathcal{X}, \mathcal{Z}, f) \triangleq \bigcup_{i=1}^{m} \{\boldsymbol{u} + \boldsymbol{e}_j : \boldsymbol{u} \lhd \mathcal{X}_i,\ j \in \mathcal{Z}_i\}.$$

The following proposition provides a necessary and sufficient condition for a linear index code to be also a solution to an instance of the ICSRI problem.

Proposition 4.12: The linear $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$ based on $\boldsymbol{L}$ is also a linear $(m, n, \mathcal{X}, \mathcal{Z}, f)$-IC if and only if $\mathcal{C}(\boldsymbol{L}) \cap \mathcal{F}(m, n, \mathcal{X}, \mathcal{Z}, f) = \emptyset$.

Proof: Let S employ the $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$ based on $\boldsymbol{L}$. Then, clearly R_i can recover $x_{f(i)}$ for all $i \in [m]$. Due to Lemma 4.3, for each $i \in [m]$ and $j \in \mathcal{Z}_i$, R_i has no information about x_j if and only if

$$\forall \boldsymbol{u} \lhd \mathcal{X}_i : \boldsymbol{u} + \boldsymbol{e}_j \notin \mathcal{C}(\boldsymbol{L}).$$

Hence, we complete the proof. ■

Example 4.4: Consider an instance $(m, n, \mathcal{X}, \mathcal{Z}, f)$ of the ICSRI problem where m, n, $\mathcal{X}$, and f are defined as in Example 4.2. Moreover, let $\mathcal{Z} = (\mathcal{Z}_1, \mathcal{Z}_2, \ldots, \mathcal{Z}_7)$, where $\mathcal{Z}_1 = \{2, 3, 4, 5\}$, $\mathcal{Z}_2 = \{1, 3, 4, 6\}$, $\mathcal{Z}_3 = \{1, 2, 4, 7\}$, and $\mathcal{Z}_4 = \mathcal{Z}_5 = \mathcal{Z}_6 = \mathcal{Z}_7 = \emptyset$. Consider the index code based on $\boldsymbol{L}$ constructed in Example 4.2. It is straightforward to verify that $\mathcal{C}(\boldsymbol{L}) \cap \mathcal{F}(m, n, \mathcal{X}, \mathcal{Z}, f) = \emptyset$. Therefore, by Proposition 4.12, this index code also provides a solution to this instance of the ICSRI problem.

Let

$$\begin{aligned}\kappa_q^* &= \kappa_q^*(m, n, \mathcal{X}, \mathcal{Z}, f) \\ &\triangleq \min\{\mathsf{rank}_q(\{\boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)}\}_{i \in [m]})\}\end{aligned}$$

where the minimum is taken over all choices of $\boldsymbol{u}^{(i)} \lhd \mathcal{X}_i$, $i \in [m]$, which satisfy

$$\mathsf{span}_q\left(\{\boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)}\}_{i \in [m]}\right) \cap \mathcal{F}(m, n, \mathcal{X}, \mathcal{Z}, f) = \emptyset. \quad (17)$$

Let $\kappa_q^* = +\infty$ if there are no choices of $\boldsymbol{u}^{(i)}$'s, $i \in [m]$, which satisfy (17). The following proposition follows immediately.

Proposition 4.13: The length of an optimal linear $(m, n, \mathcal{X}, \mathcal{Z}, f)$-IC over $\mathbb{F}_q$ is κ_q^*. If $\kappa_q^* = +\infty$, then there exist no linear $(m, n, \mathcal{X}, \mathcal{Z}, f)$-ICs over $\mathbb{F}_q$.

V. Strongly Secure Index Codes With Side Information

In this section, we consider a different model of adversary. Similarly, to its counterpart in Section IV, the adversary A in this section owns some prior side information. Additionally, A can listen to $\mu \leq N$ transmissions of S. It can also corrupt some transmissions of S, received by any of R_i, $i \in [m]$.

We start the analysis with some basic definitions of error-correcting index codes. This type of index codes has been studied very recently by the authors of this paper in [30]. We repeat some basic results for the sake of completeness.

A. Error-Correcting Index Codes

Assume that some of the symbols received by R_i, $i \in [m]$, are in error. Consider an ICSI instance $(m, n, \mathcal{X}, f)$, and assume that S broadcasts a vector $\mathfrak{E}(\boldsymbol{x}) \in \mathbb{F}_q^N$. Let $\boldsymbol{\xi}^{(i)} \in \mathbb{F}_q^N$ be the error affecting the information received by R_i, $i \in [m]$. Then, R_i actually receives the vector

$$\boldsymbol{y}^{(i)} = \mathfrak{E}(\boldsymbol{x}) + \boldsymbol{\xi}^{(i)} \in \mathbb{F}_q^N$$

instead of $\mathfrak{E}(\boldsymbol{x})$. The following definition is a generalization of Definition 3.1.

Definition 5.1: A δ-error-correcting index code over $\mathbb{F}_q$ for an instance $(m, n, \mathcal{X}, f)$ of the ICSI problem, referred to as a δ-error-correcting $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$, is an encoding function

$$\mathfrak{E} : \mathbb{F}_q^n \to \mathbb{F}_q^N$$

such that for each receiver R_i, $i \in [m]$, there exists a decoding function

$$\mathfrak{D}_i : \mathbb{F}_q^N \times \mathbb{F}_q^{|\mathcal{X}_i|} \to \mathbb{F}_q$$

satisfying

$$\forall \boldsymbol{x}, \boldsymbol{\xi}^{(i)} \in \mathbb{F}_q^n,\ \mathsf{wt}(\boldsymbol{\xi}^{(i)}) \leq \delta\ :\ \mathfrak{D}_i(\mathfrak{E}(\boldsymbol{x}) + \boldsymbol{\xi}^{(i)}, \boldsymbol{x}_{\mathcal{X}_i}) = x_{f(i)}.$$

The definitions of the length, of a linear index code, and of the matrix corresponding to an index code are naturally extended to δ-error-correcting index codes.

We define the following sets:

$$\begin{aligned}&\mathcal{I}(q, m, n, \mathcal{X}, f)\\ &\overset{\triangle}{=} \{\boldsymbol{z} \in \mathbb{F}_q^n : \exists i \in [m] \text{ such that } \boldsymbol{z}_{\mathcal{X}_i} = \boldsymbol{0} \text{ and } z_{f(i)} \neq 0\}.\end{aligned}$$

For each $i \in [m]$, we also define

$$\mathcal{Y}_i \overset{\triangle}{=} [n] \backslash \Big(\{f(i)\} \cup \mathcal{X}_i \Big).$$

Then, the collection of supports of all vectors in $\mathcal{I}(q, m, n, \mathcal{X}, f)$ is precisely

$$\mathcal{J}(m, n, \mathcal{X}, f) \overset{\triangle}{=} \bigcup_{i \in [m]} \Big\{ \{f(i)\} \cup Y_i : Y_i \subseteq \mathcal{Y}_i \Big\}. \qquad (18)$$

The necessary and sufficient condition for a matrix $\boldsymbol{L}$ to correspond to a linear δ-error-correcting index code is given in the following lemma.

Lemma 5.1: The matrix $\boldsymbol{L}$ corresponds to a linear δ-error-correcting $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$ if and only if

$$\mathsf{wt}(\boldsymbol{z}\boldsymbol{L}) \geq 2\delta + 1 \text{ for all } \boldsymbol{z} \in \mathcal{I}(q, m, n, \mathcal{X}, f). \qquad (19)$$

Equivalently, $\boldsymbol{L}$ corresponds to a linear δ-error-correcting $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$ if and only if

$$\mathsf{wt}\left(\sum_{i \in K} z_i \boldsymbol{L}_i \right) \geq 2\delta + 1$$

for all $K \in \mathcal{J}(m, n, \mathcal{X}, f)$ and for all choices of nonzero $z_i \in \mathbb{F}_q$, $i \in K$.

Proof: For each $\boldsymbol{x} \in \mathbb{F}_q^n$, we define

$$B(\boldsymbol{x}, \delta) = \{\hat{\boldsymbol{c}} \in \mathbb{F}_q^n : \hat{\boldsymbol{c}} = \boldsymbol{x}\boldsymbol{L} + \boldsymbol{\xi},\ \mathsf{wt}(\boldsymbol{\xi}) \leq \delta,\ \boldsymbol{\xi} \in \mathbb{F}_q^n\}$$

the set of all vectors resulting from at most δ errors in the transmitted vector associated with the information vector $\boldsymbol{x}$. Then, the receiver R_i can recover $x_{f(i)}$ correctly if and only if

$$B(\boldsymbol{x}, \delta) \cap B(\boldsymbol{x}', \delta) = \emptyset$$

for every pair $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{F}_q^n$ satisfying

$$\boldsymbol{x}_{\mathcal{X}_i} = \boldsymbol{x}'_{\mathcal{X}_i} \text{and } x_{f(i)} \neq x'_{f(i)}.$$

Therefore, $\boldsymbol{L}$ correspond to a linear δ-error-correcting $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$ if and only if the following condition is satisfied: for all $i \in [m]$ and for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{F}_q^n$ such that $\boldsymbol{x}_{\mathcal{X}_i} = \boldsymbol{x}'_{\mathcal{X}_i}$ and $x_{f(i)} \neq x'_{f(i)}$, it holds

$$\begin{aligned}\forall \boldsymbol{\xi}, \boldsymbol{\xi}' \in \mathbb{F}_q^N,\ \mathsf{wt}(\boldsymbol{\xi}) \leq \delta,\ \mathsf{wt}(\boldsymbol{\xi}') \leq \delta\ :\\ \boldsymbol{x}\boldsymbol{L} + \boldsymbol{\xi} \neq \boldsymbol{x}'\boldsymbol{L} + \boldsymbol{\xi}'. \qquad (20)\end{aligned}$$

Denote $\boldsymbol{z} = \boldsymbol{x}' - \boldsymbol{x}$. Then, the condition in (20) can be reformulated as follows: for all $i \in [n]$ and for all $\boldsymbol{z} \in \mathbb{F}_q^n$ such that $\boldsymbol{z}_{\mathcal{X}_i} = \boldsymbol{0}$ and $\boldsymbol{z}_{f(i)} \neq 0$, it holds

$$\forall \boldsymbol{\xi}, \boldsymbol{\xi}' \in \mathbb{F}_q^N,\ \mathsf{wt}(\boldsymbol{\xi}) \leq \delta,\ \mathsf{wt}(\boldsymbol{\xi}') \leq \delta\ :\ \boldsymbol{z}\boldsymbol{L} \neq \boldsymbol{\xi} - \boldsymbol{\xi}'. \qquad (21)$$

The equivalent condition is that for all $\boldsymbol{z} \in \mathcal{I}(q, m, n, \mathcal{X}, f)$

$$\mathsf{wt}(\boldsymbol{z}\boldsymbol{L}) \geq 2\delta + 1.$$

Since for $\boldsymbol{z} \in \mathcal{I}(q, m, n, \mathcal{X}, f)$ we have

$$\boldsymbol{z}\boldsymbol{L} = \sum_{i \in \mathsf{supp}(\boldsymbol{z})} z_i \boldsymbol{L}_i$$

the condition (19) can be restated as

$$\mathsf{wt}\left(\sum_{i\in K} z_i \boldsymbol{L}_i\right) \geq 2\delta + 1$$

for all $K \in \mathcal{J}(m,n,\mathcal{X},f)$ and for all choices of nonzero $z_i \in \mathbb{F}_q$, $i \in K$. ■

The next corollary follows directly from Lemma 5.1 by considering an error-free setup, i.e., $\delta = 0$.

Corollary 5.2: The matrix $\boldsymbol{L}$ corresponds to an $(m,n,\mathcal{X},f)$-IC over $\mathbb{F}_q$ if and only if

$$\mathsf{wt}\left(\sum_{i\in K} z_i \boldsymbol{L}_i\right) \geq 1$$

for all $K \in \mathcal{J}(m,n,\mathcal{X},f)$ and for all choices of nonzero $z_i \in \mathbb{F}_q$, $i \in K$.

The conditions stated in Corollary 5.2 and Corollary 4.5 are, as expected, equivalent. Indeed, the condition in Corollary 5.2 is equivalent to the condition that for each $i \in [m]$

$$\boldsymbol{L}_{f(i)} \notin \mathsf{span}_q(\{\boldsymbol{L}_j\}_{j\in\mathcal{Y}_i})$$

which can be rewritten as

$$\exists \boldsymbol{v}^{(i)} \in \mathbb{F}_q^N \;:\; \boldsymbol{v}^{(i)}\boldsymbol{L}_{f(i)}^T = 1 \text{ and } \boldsymbol{v}^{(i)}\boldsymbol{L}_{\mathcal{Y}_i}^T = \boldsymbol{0}.$$

For each $i \in [m]$ denote

$$\boldsymbol{c}^{(i)} \triangleq \boldsymbol{v}^{(i)}\boldsymbol{L}^T \in \mathcal{C}(\boldsymbol{L}).$$

We have for all $i \in [m]$

$$\boldsymbol{c}_{f(i)}^{(i)} = 1 \text{ and } \boldsymbol{c}_{\mathcal{Y}_i}^{(i)} = \boldsymbol{0}$$

or, equivalently

$$\exists \boldsymbol{u}^{(i)} \in \mathbb{F}_q^n \;:\; \boldsymbol{u}^{(i)} \lhd \mathcal{X}_i \text{ and } \boldsymbol{c}^{(i)} = \boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)} \in \mathcal{C}(\boldsymbol{L}).$$

Observe that all the transitions earlier are "if and only if," and therefore, Corollary 5.2 and Corollary 4.5 are equivalent, as claimed.

B. Lower Bound on the Length

We start this section with a generalization of the definition of index codes to randomized index codes. Consider $\eta \in \mathbb{N}$ random variables $G_1, G_2, \ldots, G_\eta$, which are distributed independently and uniformly over $\mathbb{F}_q$. Let $\boldsymbol{G} = (G_1, G_2, \ldots, G_\eta)$ and let $\boldsymbol{g} = (g_1, g_2, \ldots, g_\eta)$ be a realization of $\boldsymbol{G}$.

Definition 5.2: An η-randomized $(m,n,\mathcal{X},f)$-IC over $\mathbb{F}_q$ for an instance $(m,n,\mathcal{X},f)$ is an encoding function

$$\mathfrak{E} : \mathbb{F}_q^n \times \mathbb{F}_q^\eta \to \mathbb{F}_q^N$$

such that for each receiver R_i, $i \in [m]$, there exists a decoding function

$$\mathfrak{D}_i : \mathbb{F}_q^N \times \mathbb{F}_q^{|\mathcal{X}_i|} \to \mathbb{F}_q$$

satisfying

$$\forall \boldsymbol{x} \in \mathbb{F}_q^n \;:\; \mathfrak{D}_i(\mathfrak{E}(\boldsymbol{x},\boldsymbol{g}), \boldsymbol{x}_{\mathcal{X}_i}) = x_{f(i)}$$

for any $\boldsymbol{g} \in \mathbb{F}_q^\eta$, which is a realization of the random vector $\boldsymbol{G}$.

The definition of a δ-error-correcting index code can be naturally extended to that of a δ-error-correcting randomized index code. We simply replace $\mathfrak{E} : \mathbb{F}_q^n \to \mathbb{F}_q^N$ by $\mathfrak{E} : \mathbb{F}_q^n \times \mathbb{F}_q^\eta \to \mathbb{F}_q^N$, and $\mathfrak{E}(\boldsymbol{x})$ by $\mathfrak{E}(\boldsymbol{x},\boldsymbol{g})$ in Definition 5.1.

An η-randomized index code is linear over $\mathbb{F}_q$ if it has a linear encoding function $\mathfrak{E}$

$$\mathfrak{E}(\boldsymbol{x},\boldsymbol{g}) = (\boldsymbol{x} \mid \boldsymbol{g})\boldsymbol{L}$$

where $\boldsymbol{L}$ is an $(n+\eta) \times N$ matrix over $\mathbb{F}_q$. In the sequel, we assume that any message x_i, $i \in [n]$, is requested by at least one receiver. Observe that by simply treating $x_1, x_2, \ldots, x_n, g_1, g_2, \ldots, g_\eta$ as messages, the results from previous sections still apply to linear randomized index codes.

Definition 5.3: The linear η-randomized $(m,n,\mathcal{X},f)$-IC over $\mathbb{F}_q$ based on $\boldsymbol{L}$ is said to be (μ,t,δ)-*strongly secure* if it has the following two properties.

1) This code is δ-error-correcting. In other words, upon receiving $(\boldsymbol{x}|\boldsymbol{g})\boldsymbol{L}$ with at most δ coordinates in error, the receiver R_i can still recover $x_{f(i)}$, for all $i \in [m]$.
2) This code is (μ,t)-strongly secure. In other words, an adversary A who possesses $\boldsymbol{x}_{\mathcal{X}_A}$, for $\mathcal{X}_A \subseteq [n]$, $|\mathcal{X}_A| = t$, and listens to at most μ transmissions, $\mu \leq N$, gains no information about other messages. Equivalently

$$\mathsf{H}(\boldsymbol{X}_{\widehat{\mathcal{X}}_A} \mid (\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W], \boldsymbol{X}_{\mathcal{X}_A}) = \mathsf{H}(\boldsymbol{X}_{\widehat{\mathcal{X}}_A})$$

for any $W \subseteq [N]$, $|W| \leq \mu$.

Remark 5.3:

1) If $\mu = t = \eta = 0$, then a (μ,t,δ)-strongly secure η-randomized $(m,n,\mathcal{X},f)$-IC over $\mathbb{F}_q$ is simply a δ-error-correcting $(m,n,\mathcal{X},f)$-IC over $\mathbb{F}_q$.
2) If $\delta = 0$, the index code is strongly secure, but has no error-correcting capability. In that case, we simply say that the code is "(μ,t)-strongly secure" instead of "$(\mu,t,0)$-strongly secure."
3) A simple concatenation of an error-correcting IC scheme and a secure IC scheme may not necessarily yield a (μ,t,δ)-strongly secure η-randomized $(m,n,\mathcal{X},f)$-IC over $\mathbb{F}_q$. By contrast, in Section V-C we present a somewhat more sophisticated scheme that yields a (μ,t,δ)-strongly secure μ-randomized $(m,n,\mathcal{X},f)$-IC over $\mathbb{F}_q$.

In the lemma that follows, we assume that each message is requested by at least one receiver. Otherwise, that "useless" message can be discarded without affecting the model.

Lemma 5.4: If $\boldsymbol{L}$ corresponds to a (μ,t)-strongly secure linear η-randomized $(m,n,\mathcal{X},f)$-IC over $\mathbb{F}_q$, then $\eta \geq \mu$.

Proof: We prove this lemma by contradiction. Suppose that $\boldsymbol{L}$ corresponds to a (μ,t)-strongly secure η-randomized $(m,n,\mathcal{X},f)$-IC over $\mathbb{F}_q$, and that $\eta < \mu$. Let $E = \{n+1, n+2, \ldots, n+\eta\}$.

For $W \subseteq [N]$, let $\mathcal{C}(\boldsymbol{L}[W])$ be the space spanned by columns of $\boldsymbol{L}$ indexed by elements of W. Then, for all $W \subseteq [N]$ with $|W| \le \mu$, it holds that

$$\mathsf{H}(\boldsymbol{X}_{\widehat{\mathcal{X}}_A}|(\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W], \boldsymbol{X}_{\mathcal{X}_A}) = \mathsf{H}(\boldsymbol{X}_{\widehat{\mathcal{X}}_A})$$

i.e., an adversary who owns $\boldsymbol{x}_{\mathcal{X}_A}$ gains no information about $\boldsymbol{x}_{\widehat{\mathcal{X}}_A}$ after eavesdropping the transmissions corresponding to the set of indices W. From Lemma 4.3 with $\mathcal{C}(\boldsymbol{L})$ being replaced by $\mathcal{C}(\boldsymbol{L}[W])$, we conclude that $\mathcal{C}(\boldsymbol{L}[W])$ does not contain a vector $\boldsymbol{c}$ which satisfies $\boldsymbol{c}_{\widehat{\mathcal{X}}_A} \neq \boldsymbol{0}$ and $\boldsymbol{c}_E = \boldsymbol{0}$. In the sequel, we refer to this property of $\mathcal{C}(\boldsymbol{L}[W])$ as Property A.

Let $\boldsymbol{L}' = (\boldsymbol{L}_{\widehat{\mathcal{X}}_A \cup E})^T$ be the matrix obtained from $\boldsymbol{L}$ by first deleting rows of $\boldsymbol{L}$ indexed by $\mathcal{X}_A$, and then taking its transpose. We show that $\mathsf{rank}_q(\boldsymbol{L}') \le \mu - 1$. Indeed, take any μ rows of $\boldsymbol{L}'$, denote them $\boldsymbol{L}'_{j_1}, \dots, \boldsymbol{L}'_{j_\mu}$. Let $\boldsymbol{L}''$ be the submatrix of $\boldsymbol{L}'$ formed by the last η columns. Since $\eta < \mu$, the μ rows $\boldsymbol{L}''_{j_1}, \dots, \boldsymbol{L}''_{j_\mu}$ are linearly dependent. Hence, there exist $\alpha_1, \alpha_2, \dots, \alpha_\mu$, not all zeros, such that

$$\sum_{\ell=1}^{\mu} \alpha_\ell \boldsymbol{L}''_{j_\ell} = \boldsymbol{0}.$$

This implies

$$\sum_{\ell=1}^{\mu} \alpha_\ell \boldsymbol{L}'_{j_\ell} = \boldsymbol{0}$$

due to Property A. Thus, $\mathsf{rank}_q(\boldsymbol{L}') \le \mu - 1$.

Now let $\mathsf{r} \stackrel{\triangle}{=} \mathsf{rank}_q(\boldsymbol{L}') < \mu$, and let

$$\{\boldsymbol{L}'_{j_1}, \boldsymbol{L}'_{j_2}, \dots, \boldsymbol{L}'_{j_\mathsf{r}}\}$$

be a basis of the space spanned by the rows of $\boldsymbol{L}'$. Suppose that the receiver R_i requests $x_{f(i)}$ where $f(i) \in \widehat{\mathcal{X}}_A$.

1) On the one hand, by Corollary 4.5, $\mathcal{C}(\boldsymbol{L})$ contains a vector $\boldsymbol{c} = \boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)}$ where $\boldsymbol{u}^{(i)} \lhd \mathcal{X}_i$. Therefore, $\boldsymbol{c}_E = \boldsymbol{0}$ and $\boldsymbol{c}_{\widehat{\mathcal{X}}_A} \neq \boldsymbol{0}$.
2) On the other hand, there exist $\beta_1, \beta_2, \dots, \beta_\mathsf{r}$ such that

$$(\boldsymbol{c}_{\widehat{\mathcal{X}}_A}|\boldsymbol{c}_E) = \sum_{\ell=1}^{\mathsf{r}} \beta_\ell \boldsymbol{L}'_{j_\ell}.$$

Since $\mathsf{r} < \mu$ and $\boldsymbol{c}_E = \boldsymbol{0}$, by Property A, we have $\boldsymbol{c}_{\widehat{\mathcal{X}}_A} = \boldsymbol{0}$. We obtain a contradiction. ∎

Remark 5.5: From Lemma 5.4, a (μ, t)-strongly secure linear randomized index code requires at least μ random symbols. We show in Section V-C that there exists such a code that uses precisely μ random symbols.

Lemma 5.6: Suppose that $\boldsymbol{L}$ corresponds to a linear μ-randomized $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$. If this randomized index code is (μ, t)-strongly secure, then for all $i \in [\mu]$, there exists a vector $\boldsymbol{v}^{(i)} \in \mathbb{F}_q^{n+\mu}$ satisfying

1) $\boldsymbol{v}^{(i)} \lhd [n]$;
2) $\boldsymbol{v}^{(i)} + \boldsymbol{e}_{n+i} \in \mathcal{C}(\boldsymbol{L})$.

Proof: Assume, by contradiction, that for some $i \in [\mu]$, we have $\boldsymbol{v}^{(i)} + \boldsymbol{e}_{n+i} \notin \mathcal{C}(\boldsymbol{L})$ for all $\boldsymbol{v}^{(i)} \lhd [n]$. Consider a virtual receiver, which has a side information set $\{x_j\}_{j \in [n]}$, and requests the symbol g_i. By Corollary 4.4, this virtual receiver has no information about g_i after listening to all transmissions. In other words, we have

$$\mathsf{H}(G_i|(\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}, \boldsymbol{X}) = \mathsf{H}(G_i) \tag{22}$$

where all symbols in $\boldsymbol{X}$ and $\boldsymbol{G}$ are independent and uniformly distributed. In particular, for a smaller set of side information

$$\mathsf{H}(G_i|(\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}, \boldsymbol{X}_{\mathcal{X}_A}) = \mathsf{H}(G_i). \tag{23}$$

We recall Definition 5.3: for every μ-subset $W \subseteq [N]$ and every t-subset $\mathcal{X}_A \subseteq [n]$, we have

$$\mathsf{H}(\boldsymbol{X}_{\widehat{\mathcal{X}}_A}|(\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W], \boldsymbol{X}_{\mathcal{X}_A}) = \mathsf{H}(\boldsymbol{X}_{\widehat{\mathcal{X}}_A}). \tag{24}$$

In the sequel, we show that if the value of G_i is known to the adversary, this randomized index code is still (μ, t)-strongly secure. In other words, we aim to show that

$$\mathsf{H}(\boldsymbol{X}_{\widehat{\mathcal{X}}_A}|(\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W], \boldsymbol{X}_{\mathcal{X}_A}, G_i) = \mathsf{H}(\boldsymbol{X}_{\widehat{\mathcal{X}}_A}) \tag{25}$$

for every μ-subset $W \subseteq [N]$ and every t-subset $\mathcal{X}_A \subseteq [n]$. Indeed, the left-hand side of (25) is equal to

$$\mathsf{H}(\boldsymbol{X}_{\widehat{\mathcal{X}}_A}|(\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W], \boldsymbol{X}_{\mathcal{X}_A}) - \mathsf{I}(\boldsymbol{X}_{\widehat{\mathcal{X}}_A}; G_i|(\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W], \boldsymbol{X}_{\mathcal{X}_A})$$

which is

$$\mathsf{H}(\boldsymbol{X}_{\widehat{\mathcal{X}}_A}) - \mathsf{I}(\boldsymbol{X}_{\widehat{\mathcal{X}}_A}; G_i|(\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W], \boldsymbol{X}_{\mathcal{X}_A})$$

due to (24). Hence, it suffices to show that

$$\mathsf{I}(\boldsymbol{X}_{\widehat{\mathcal{X}}_A}; G_i \mid (\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W], \boldsymbol{X}_{\mathcal{X}_A}) = 0.$$

We have

$$\begin{aligned}
\mathsf{I}(\boldsymbol{X}_{\widehat{\mathcal{X}}_A}; G_i|(\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W], \boldsymbol{X}_{\mathcal{X}_A}) &\\
= \mathsf{H}(G_i|(\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W], \boldsymbol{X}_{\mathcal{X}_A}) &\\
- \mathsf{H}(G_i|(\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W], \boldsymbol{X}_{\mathcal{X}_A}, \boldsymbol{X}_{\widehat{\mathcal{X}}_A}) &\\
= \mathsf{H}(G_i|(\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W], \boldsymbol{X}_{\mathcal{X}_A}) &\\
- \mathsf{H}(G_i|(\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W], \boldsymbol{X}) &\\
= \mathsf{H}(G_i) - \mathsf{H}(G_i) &\\
= 0 &
\end{aligned}$$

where the third transition is due to (22) and (23).

To this end, we have shown that the randomized index code is still (μ, t)-strongly secure if the adversary knows the realized value of G_i. Equivalently, discarding the random variable G_i from the scheme does not affect its strong security. However, this contradicts Lemma 5.4, since the resulting code has less than μ random symbols. ∎

The following theorem proves a lower bound on the length of a (μ, t)-strongly secure linear randomized index code.

Theorem 5.7: The length of a (μ, t)-strongly secure linear η-randomized $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$ is at least $\kappa_q + \mu$.

Proof: Suppose the linear randomized index code is based on $\boldsymbol{L}$. We divide the proof into several cases.

Case 1: $\eta = \mu$. Then, by Corollary 4.5 and Lemma 5.6, the subspace $\mathcal{C}(\boldsymbol{L})$ must contain the following.

1) The vectors $\boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)}$ for some $\boldsymbol{u}^{(i)} \lhd \mathcal{X}_i$, for all $i \in [m]$.
2) The vectors $\boldsymbol{v}^{(i)} + \boldsymbol{e}_{n+i}$, for some $\boldsymbol{v}^{(i)} \lhd [n]$, for all $i \in [\mu]$.

Due to linear independence of these vectors and to the definition of κ_q, the length of the code is at least

$$\begin{aligned}\dim(\mathcal{C}(\boldsymbol{L})) &\geq \mathsf{rank}_q(\{\boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)}\}_{i\in[m]}) \\ &\quad + \mathsf{rank}_q(\{\boldsymbol{v}^{(i)} + \boldsymbol{e}_{n+i}\}_{i\in[\mu]}) \\ &\geq \kappa_q + \mu.\end{aligned}$$

Case 2: $\eta > \mu$, and for all $i \in [\eta]$ there exists some vector $\boldsymbol{v}^{(i)} \lhd [n]$ such that $\boldsymbol{v}^{(i)} + \boldsymbol{e}_{n+i} \in \mathcal{C}(\boldsymbol{L})$.

In this case, similarly to Case 1, we have

$$\dim(\mathcal{C}(\boldsymbol{L})) \geq \kappa_q + \eta > \kappa_q + \mu.$$

Therefore, $\boldsymbol{L}$ has at least $\kappa_q + \mu$ columns.

Case 3: $\eta > \mu$, and for some $i \in [\eta]$, $\boldsymbol{v}^{(i)} + \boldsymbol{e}_{n+i} \notin \mathcal{C}(\boldsymbol{L})$ for all $\boldsymbol{v}^{(i)} \lhd [n]$. By following exactly the same argument as in the proof of Lemma 5.6, we deduce that discarding G_i does not affect the strong security of the randomized index code. By doing so, we obtain a new randomized (μ, t)-strongly secure index code, which has $\eta - 1$ random variables. This code is based on $\boldsymbol{L}'$, which is obtained from $\boldsymbol{L}$ by deleting its $(n+i)$th row.

The aforementioned argument can be applied until either the number of random variables decreases to μ, or the code in consideration satisfies the condition of Case 2. In both cases, the resulting randomized index code has length at least $\kappa_q + \mu$. As the length of the code does not change during the process, we conclude that the length of the original code is at least $\kappa_q + \mu$. ■

The next theorem establishes a lower bound on the length of a (μ, t, δ)-strongly secure linear randomized index code.

Theorem 5.8: The length of a (μ, t, δ)-strongly secure linear η-randomized $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$ is at least $\kappa_q + \mu + 2\delta$.

Proof: Let $\boldsymbol{L}$ correspond to a (μ, t, δ)-strongly secure η-randomized $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$. Let $\boldsymbol{L}'$ be the matrix obtained by deleting any 2δ columns of $\boldsymbol{L}$. Since $\boldsymbol{L}$ corresponds to a δ-error-correcting index code, by Lemma 5.1, it satisfies

$$\mathsf{wt}\left(\sum_{i\in K} z_i \boldsymbol{L}_i\right) \geq 2\delta + 1$$

for all $K \in \mathcal{J}(m, n, \mathcal{X}, f)$ and all choices of nonzero $z_i \in \mathbb{F}_q$, $i \in K$. We obtain that the rows of $\boldsymbol{L}'$ satisfy

$$\mathsf{wt}\left(\sum_{i\in K} z_i \boldsymbol{L}'_i\right) \geq 1.$$

By Corollary 5.2, $\boldsymbol{L}'$ corresponds to an η-randomized $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$. Since all entries of $\boldsymbol{L}'$ are contained in $\boldsymbol{L}$, we deduce that $\boldsymbol{L}'$ corresponds to a (μ, t)-strongly secure η-randomized $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$. Therefore, by Theorem 5.7, $\boldsymbol{L}'$ has at least $\kappa_q + \mu$ columns. Therefore, $\boldsymbol{L}$ has at least $\kappa_q + \mu + 2\delta$ columns. ■

C. A Construction of Optimal Strongly Secure Index Codes

In this section, we present a construction of an optimal (μ, t, δ)-strongly secure μ-randomized linear $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$, which has length attaining the lower bound established in Theorem 5.8. It requires q to be at least $\kappa_q + \mu + 2\delta + 1$. The proposed construction is based on the coset coding technique, originally introduced by Ozarow and Wyner [31]. This technique has been adopted in a variety of NC applications, such as in [18]–[22].

Construction A: Let $\boldsymbol{L}^{(0)}$ correspond to a linear $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$ of optimal length κ_q. Let $\boldsymbol{M}$ be a generator matrix of an $[N = \kappa_q + \mu + 2\delta, \kappa_q + \mu, 2\delta + 1]_q$ MDS code, so that the last μ rows of $\boldsymbol{M}$ form a generator matrix of another MDS code. For instance, take

$$\boldsymbol{M} = \left(\begin{array}{cccc} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_N \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{\kappa_q - 1} & \alpha_2^{\kappa_q - 1} & \cdots & \alpha_N^{\kappa_q - 1} \\ \hline \alpha_1^{\kappa_q} & \alpha_2^{\kappa_q} & \cdots & \alpha_N^{\kappa_q} \\ \alpha_1^{\kappa_q + 1} & \alpha_2^{\kappa_q + 1} & \cdots & \alpha_N^{\kappa_q + 1} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{\kappa_q + \mu - 1} & \alpha_2^{\kappa_q + \mu - 1} & \cdots & \alpha_N^{\kappa_q + \mu - 1} \end{array}\right)$$

where $\alpha_1, \alpha_2, \ldots, \alpha_N$ are pairwise distinct nonzero elements in $\mathbb{F}_q$. Let $\boldsymbol{P}$ be the submatrix of $\boldsymbol{M}$ formed by the first κ_q rows, and $\boldsymbol{Q}$ the submatrix formed by the last μ rows of $\boldsymbol{M}$. Take

$$\boldsymbol{L} = \left(\begin{array}{c} \boldsymbol{L}^{(0)}\boldsymbol{P} \\ \hline \boldsymbol{Q} \end{array}\right).$$

Lemma 5.9: The matrix $\boldsymbol{L}$ in Construction A corresponds to a δ-error-correcting μ-randomized $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$.

Proof: Recall that $\boldsymbol{g} \in \mathbb{F}_q^{\mu}$ is a random vector. The encoding function $\mathfrak{E}$ has a form

$$\mathfrak{E}(\boldsymbol{x}, \boldsymbol{g}) = (\boldsymbol{x}|\boldsymbol{g})\boldsymbol{L} = \boldsymbol{x}\boldsymbol{L}^{(0)}\boldsymbol{P} + \boldsymbol{g}\boldsymbol{Q} = (\boldsymbol{x}\boldsymbol{L}^{(0)}|\boldsymbol{g})\boldsymbol{M}.$$

Since $\boldsymbol{M}$ is a generator matrix of a δ-error-correcting code, each receiver R_i, $i \in [m]$, is able to recover $(\boldsymbol{x}\boldsymbol{L}^{(0)}|\boldsymbol{g})$ if the number of errors in $\mathfrak{E}(\boldsymbol{x}, \boldsymbol{g})$ is less than or equal to δ. Therefore, each receiver R_i can recover $\boldsymbol{x}\boldsymbol{L}^{(0)}$, and hence, it can also recover $x_{f(i)}$, $i \in [m]$, as $\boldsymbol{L}^{(0)}$ corresponds to a linear $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$. ■

Lemma 5.10: The matrix $\boldsymbol{L}$ in Construction A corresponds to a (μ, t)-strongly secure μ-randomized $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$.

Proof: Suppose that the adversary A possess a message vector $\boldsymbol{x}_{\mathcal{X}_A}$, $|\boldsymbol{x}_{\mathcal{X}_A}| = t$. Additionally, A can eavesdrop μ transmissions, i.e., it has a knowledge of $\boldsymbol{b} \overset{\triangle}{=} (\boldsymbol{x}|\boldsymbol{g})\boldsymbol{L}[W]$, for some $W \subseteq [N]$, $|W| = \mu$. In the following, we show that the entropy of $\boldsymbol{X}_{\widehat{\mathcal{X}}_A}$ is not changed given the knowledge of $(\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W]$ and of $\boldsymbol{x}_{\mathcal{X}_A}$. It suffices to show that for all $\boldsymbol{a} \in \mathbb{F}_q^{n-t}$

$$\Pr(\boldsymbol{X}_{\widehat{\mathcal{X}}_A} = \boldsymbol{a} \mid (\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W] = \boldsymbol{b},\ \boldsymbol{X}_{\mathcal{X}_A} = \boldsymbol{x}_{\mathcal{X}_A}) = \frac{1}{q^{n-t}} \tag{26}$$

where all symbols in $\boldsymbol{X}$ and $\boldsymbol{G}$ are independent and uniformly distributed. The left-hand side of (26) can be re-written as

$$\frac{\Pr(\boldsymbol{X}_{\widehat{\mathcal{X}}_A} = \boldsymbol{a}, (\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W] = \boldsymbol{b} \mid \boldsymbol{X}_{\mathcal{X}_A} = \boldsymbol{x}_{\mathcal{X}_A})}{\Pr((\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W] = \boldsymbol{b} \mid \boldsymbol{X}_{\mathcal{X}_A} = \boldsymbol{x}_{\mathcal{X}_A})}. \tag{27}$$

The numerator in (27) is given by

$$\begin{aligned}
&\Pr(\boldsymbol{X}_{\widehat{\mathcal{X}}_A} = \boldsymbol{a}, (\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W] = \boldsymbol{b} \mid \boldsymbol{X}_{\mathcal{X}_A} = \boldsymbol{x}_{\mathcal{X}_A})\\
&\quad = \Pr(\boldsymbol{X}_{\widehat{\mathcal{X}}_A} = \boldsymbol{a} \mid \boldsymbol{X}_{\mathcal{X}_A} = \boldsymbol{x}_{\mathcal{X}_A})\\
&\qquad \times \Pr((\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W] = \boldsymbol{b} \mid \boldsymbol{x}_{\widehat{\mathcal{X}}_A} = \boldsymbol{a}, \boldsymbol{X}_{\mathcal{X}_A} = \boldsymbol{x}_{\mathcal{X}_A})\\
&\quad = \frac{1}{q^{n-t}} \Pr((\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W] = \boldsymbol{b} \mid \boldsymbol{X}_{\widehat{\mathcal{X}}_A} = \boldsymbol{a}, \boldsymbol{X}_{\mathcal{X}_A} = \boldsymbol{x}_{\mathcal{X}_A})\\
&\quad = \frac{1}{q^{n-t}} \frac{1}{q^{\mu}} \ = \ \frac{1}{q^{n-t+\mu}}.
\end{aligned} \tag{28}$$

The penultimate transition can be explained as follows. We have

$$\boldsymbol{b} = (\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W] = \boldsymbol{X}\boldsymbol{L}^{(0)}\boldsymbol{P}[W] + \boldsymbol{G}\boldsymbol{Q}[W]. \tag{29}$$

The matrix $\boldsymbol{Q}[W]$ is invertible due to the fact that $\boldsymbol{Q}$ is a generator matrix of an $[N, \mu]$-MDS code. Since $\boldsymbol{X}$ is known, the system (29) has a unique solution given by

$$\boldsymbol{G} = (\boldsymbol{b} - \boldsymbol{X}\boldsymbol{L}^{(0)}\boldsymbol{P}[W])(\boldsymbol{Q}[W])^{-1}.$$

Since $\boldsymbol{G}$ is uniformly distributed over $\mathbb{F}_q^{\mu}$

$$\begin{aligned}
&\Pr((\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W] = \boldsymbol{b} \mid \boldsymbol{X}_{\widehat{\mathcal{X}}_A} = \boldsymbol{a}, \boldsymbol{X}_{\mathcal{X}_A} = \boldsymbol{x}_{\mathcal{X}_A})\\
&\quad = \Pr(\boldsymbol{G} = (\boldsymbol{b} - \boldsymbol{X}\boldsymbol{L}^{(0)}\boldsymbol{P}[W])(\boldsymbol{Q}[W])^{-1})\\
&\quad = \frac{1}{q^{\mu}}\ .
\end{aligned}$$

Similarly to (28), the denominator in (27) is

$$\begin{aligned}
&\Pr((\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W] = \boldsymbol{b} \mid \boldsymbol{X}_{\mathcal{X}_A} = \boldsymbol{x}_{\mathcal{X}_A})\\
&\quad = \sum_{\boldsymbol{c} \in \mathbb{F}_q^{n-t}} \Pr(\boldsymbol{X}_{\widehat{\mathcal{X}}_A} = \boldsymbol{c} \mid \boldsymbol{X}_{\mathcal{X}_A} = \boldsymbol{x}_{\mathcal{X}_A})\\
&\qquad \times \Pr((\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W] = \boldsymbol{b} \mid \boldsymbol{X}_{\widehat{\mathcal{X}}_A} = \boldsymbol{c}, \boldsymbol{X}_{\mathcal{X}_A} = \boldsymbol{x}_{\mathcal{X}_A})\\
&\quad = q^{n-t} \frac{1}{q^{n-t}} \frac{1}{q^{\mu}} \ = \ \frac{1}{q^{\mu}}\ .
\end{aligned} \tag{30}$$

From (27), (28), and (30), we obtain (26), as claimed. ■

From Theorem 5.8, Lemma 5.9, and Lemma 5.10, we obtain the main result of this section.

Theorem 5.11: The length of an optimal (μ, t, δ)-strongly secure linear η-randomized $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q (q \geq \kappa_q + \mu + 2\delta + 1)$ is $\kappa_q + \mu + 2\delta$. Moreover, the code in Construction A achieves this optimal length.

VI. Conclusions and Open Questions

In this paper, we initiate a study of the security aspects of linear index coding schemes. We introduce a notion of block security and establish two bounds on the security level of a linear index code based on the matrix $\boldsymbol{L}$. These analysis makes use of the minimum distance and the dual distance of $\mathcal{C}(\boldsymbol{L})$, the code spanned by the columns of $\boldsymbol{L}$. While the dimension of this code corresponds to the number of transmissions in the scheme, the minimum distance characterizes its security strength.

Our second contribution is the analysis of the strong security of linear index codes. New bounds on the length of linear index codes, which are resistant to errors, eavesdropping, and information leaking, are established. Index codes that achieve these bounds are constructed. These new bounds cannot be deduced directly from the existing results in network coding literature.

One important problem, which remains open, deals with a design of an optimal secure index coding scheme. This problem can be formulated as follows: given an instance of the ICSI problem, how to design $\boldsymbol{L}$, such that $\mathcal{C}(\boldsymbol{L})$ has the largest possible minimum distance? More specifically, let us define the binary side information matrix $\boldsymbol{A} = (\mathsf{a}_{i,j})_{i \in [n],\, j \in [n]}$ as in [6], namely

$$\mathsf{a}_{i,j} = \begin{cases} 1 & \text{if } j = i \text{ or } j \in \mathcal{X}_i \\ 0 & \text{otherwise.} \end{cases}$$

The problem is equivalent to finding a way to turn certain off-diagonal 1's in $\boldsymbol{A}$ into 0's, such that the rows of the resulting matrix generate an error-correcting code of the largest possible minimum distance. It is very likely that this task is a hard problem. For comparison, even finding the minimum distance of an error-correcting code given by its generating matrix is known to be NP-hard [32].

Acknowledgment

The authors would like to thank Frédérique Oggier for helpful discussions.

References

[1] Y. Birk and T. Kol, "Informed-source coding-on-demand (ISCOD) over broadcast channels," in *Proc. IEEE Conf. Comput. Commun.*, San Francisco, CA, 1998, pp. 1257–1264.

[2] Y. Birk and T. Kol, "Coding-on-demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2825–2830, Jun. 2006.

[3] A. E. Rouayheb, A. Sprintson, and C. Georghiades, "On the index coding problem and its relation to network coding and matroid theory," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3187–3195, Jul. 2010.

[4] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "Xors in the air: Practical wireless network coding," in *Proc. ACM SIGCOMM*, 2006, pp. 243–254.

[5] S. Katti, D. Katabi, H. Balakrishnan, and M. Médard, "Symbol-level network coding for wireless mesh networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 4, pp. 401–412, 2008.

[6] Z. Bar-Yossef, Z. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," in *Proc. 47th Annu. IEEE Symp. Found. Comput. Sci.*, 2006, pp. 197–206.

[7] E. Lubetzky and U. Stav, "Non-linear index coding outperforming the linear optimum," in *Proc. 48th Annu. IEEE Symp. Found. Comput. Sci.*, 2007, pp. 161–168.

[8] J. P. Y. Wu, R. Chandra, V. Padmanabhan, and P. A. Chou, "The local mixing problem," presented at the presented at the Inf. Theory Appl. Workshop, San Diego, CA, 2006.

[9] S. E. Rouayheb, M. A. R. Chaudhry, and A. Sprintson, "On the minimum number of transmissions in single-hop wireless coding networks," in *Proc. IEEE Inf. Theory Workshop*, 2007, pp. 120–125.

[10] A. E. Rouayheb, A. Sprintson, and C. Georghiades, "On the relation between the index coding and the network coding problems," in *Proc. IEEE Symp. Inf. Theory*, Toronto, Canada, 2008, pp. 1823–1827.

[11] M. A. R. Chaudhry and A. Sprintson, "Efficient algorithms for index coding," in *Proc. IEEE Conf. Comput. Commun.*, 2008, pp. 1–4.

[12] N. Alon, A. Hassidim, E. Lubetzky, U. Stav, and A. Weinstein, "Broadcasting with side information," in *Proc. 49th Annu. IEEE Symp. Found. Comput. Sci.*, 2008, pp. 823–832.

[13] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[14] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[15] R. Peeters, "Orthogonal representations over finite fields and the chromatic number of graphs," *Combinatorica*, vol. 16, no. 3, pp. 417–431, 1996.

[16] J. L. Massey, "Minimal codewords and secret sharing," in *Proc. Joint Swedish-Russian Int. Workshop Inf. Theory*, 1993, pp. 276–279.

[17] C. Ding, R. Laihonen, and A. Renvall, "Linear multisecret-sharing schemes and error-correcting codes," *J. Universal Comput. Sci.*, vol. 3, no. 9, pp. 1023–1036, 1997.

[18] N. Cai and R. W. Yeung, "Secure network coding," in *Proc. IEEE Symp. Inf. Theory*, Lausanne, Switzerland, 2002, pp. 551–555.

[19] J. Feldman, T. Malkin, R. A. Servedio, and C. Stein, "On the capacity of secure network coding," in *Proc. Annu. Allerton Conf. Commun., Control, Comput.*, 2004.

[20] A. E. Rouayheb and E. Soljanin, "On wiretap networks II," in *Proc. IEEE Symp. Inf. Theory*, Nice, France, 2007, pp. 551–555.

[21] Z. Zhuang, Y. Luo, and A. J. H. Vinck, "Secure error-correcting network codes with side information from source," in *Proc. Int. Conf. Commun. Intell. Inf. Security*, 2010, pp. 55–59.

[22] D. Silva and F. R. Kschischang, "Universal secure error-correcting schemes for network coding," in *Proc. IEEE Symp. Inf. Theory*, Austin, TX, 2010, pp. 2428–2432.

[23] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1977.

[24] A. S. Hedayat, N. J. A. Sloane, and J. Stufken, *Orthogonal Arrays: Theory and Applications*. New York: Springer-Verlag, 1999.

[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.

[26] W. Haemers, "An upper bound for the Shannon capacity of a graph," *Algebr. Methods Graph Theory*, vol. 25, pp. 267–272, 1978.

[27] K. Bhattad and K. R. Narayanan, "Weakly secure network coding," in *Proc. 1st Workshop Netw. Coding, Theory, Appl.*, 2005.

[28] D. Silva and F. R. Kschischang, "Universal weakly secure network coding," in *Proc. Inf. Theory Workshop Netw. Inf. Theory*, 2009, pp. 281–285.

[29] R. Roth, *Introduction to Coding Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[30] S. H. Dau, V. Skachek, and Y. M. Chee, *Index coding and error correction*, 2011 [Online]. Available: http://arxiv.org/abs/1101.2728

[31] L. H. Ozarow and A. D. Wyner, "The wire-tap channel II," *Bell Syst. Tech. J.*, vol. 63, pp. 2135–2157, 1984.

[32] A. Vardy, "The intractability of computing the minimum distance of a code," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1757–1766, Nov. 1997.

Son Hoang Dau received the B.S. degree in applied mathematics and informatics from the College of Science, Vietnam National University, Hanoi, Vietnam, in 2006 and the M.S. degree in mathematical sciences from the Division of Mathematical Sciences, Nanyang Technological University, Singapore, where he is currently working towards the Ph.D. degree.

His research interests are coding theory, network coding, and combinatorics.

Vitaly Skachek received the B.A. (Cum Laude), M.Sc. and Ph.D. degrees in computer science from the Technion—Israel Institute of Technology, in 1994, 1998 and 2007, respectively.

In the summer of 2004, he visited the Mathematics of Communications Department at Bell Laboratories under the DIMACS Special Focus Program in Computational Information Theory and Coding. During 2007–2012, he held visiting positions with the Claude Shannon Institute and the School of Mathematical Sciences, University College Dublin, with the School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, and with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana. He is now visiting the Department of Electrical and Computer Engineering, McGill University, Montreal.

Dr. Skachek is a recipient of the Permanent Excellent Faculty Instructor award, given by Technion.

Yeow Meng Chee (SM'08) received the B.Math. degree in computer science and combinatorics and optimization and the M.Math. and Ph.D. degrees in computer science from the University of Waterloo, Waterloo, ON, Canada, in 1988, 1989, and 1996, respectively. Currently, he is an Associate Professor at the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore. Prior to this, he was Program Director of Interactive Digital Media R & D in the Media Development Authority of Singapore, Postdoctoral Fellow at the University of Waterloo and IBM's Zürich Research Laboratory, General Manager of the Singapore Computer Emergency Response Team, and Deputy Director of Strategic Programs at the Infocomm Development Authority, Singapore. His research interest lies in the interplay between combinatorics and computer science/engineering, particularly combinatorial design theory, coding theory, extremal set systems, and electronic design automation.