On Secure Index Coding with Side Information

Son Hoang Dau
Division of Mathematical Sciences
School of Phys. and Math. Sciences
Nanyang Technological University
21 Nanyang Link, Singapore 637371
Email: *daus0002@ntu.edu.sg*

Vitaly Skachek[1]
Coordinated Science Laboratory
University of Illinois, Urbana-Champaign
1308 W. Main Street
Urbana, IL 61801, USA
Email: *vitalys@illinois.edu*

Yeow Meng Chee
Division of Mathematical Sciences
School of Phys. and Math. Sciences
Nanyang Technological University
21 Nanyang Link, Singapore 637371
Email: *YMChee@ntu.edu.sg*

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 $\boldsymbol{L}$, whose column space code $\mathcal{C}(\boldsymbol{L})$ 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 $t \leq d-2$ 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: (i) possesses t messages in advance; (ii) eavesdrops at most μ transmissions; (iii) 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.

I. Introduction

A. Background

The problem of Index Coding with Side Information (ICSI) was introduced by Birk and Kol [1]. It considers a communications scenario with one server and many clients. Each client misses a certain part of the data, due to intermittent reception, limited storage capacity or any other reasons. Before the transmission starts, the clients let the server know which packets they already have in their possession, and which packets they are interested to receive. The server needs to deliver all the messages each client requests, yet spending a minimum number of transmissions. As it was shown in [1], the server can significantly reduce the number of transmissions by coding the messages.

Possible applications of index coding include communications scenarios, in which a satellite or a server broadcasts a set of messages to a set clients, such as daily newspaper delivery or video-on-demand. ICSI can also be used in opportunistic wireless networks [2].

The ICSI problem has been a subject of several recent studies [3]–[8]. This problem can be viewed as a special case of the Network Coding (NC) problem [9], [10]. In particular, it was shown in [7] that every instance of the NC problem can be reduced to an instance of the ICSI problem.

[1]This work was done while Vitaly Skachek was with the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371.

B. Our contribution

In this paper, we initiate a study of the security aspects of linear index coding schemes. For each linear index code, we have a matrix $\boldsymbol{L}$, which represents a linear encoding function (it will be defined formally in the sequel). We introduce a notion of block security and establish two bounds on the security level of a deterministic linear index code based on $\boldsymbol{L}$. The 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.

We also introduce a natural generalization of the ICSI problem, called the *Index Coding with Side and Restricted Information (ICSRI)* problem. The results on the security of linear index codes are employed to analyze the existence of solutions to the ICSRI problem.

Finally, we consider linear index codes which use random messages. We establish new bounds on the length of such linear ICs, which are resistant to errors, eavesdropping, and information leaking. We also show that the coset coding technique (which has been successfully employed in network coding literature, see [11], [12]) yields an optimal strongly secure linear index code.

Whereas most of the known results on the security aspects of network coding were derived for the multicast scenario, the ICSI problem can be modeled as a special case of the non-multicast NC problem ([7], [8]). Being modeled in that way, the symbols transmitted on a set of special edges, which carry the side information, are not allowed to be corrupted. By contrast, for network coding any edge can be corrupted. These two differences suggest that the existing results on the security in network coding can not be directly generalized to index coding.

For detailed proofs, we refer the reader to the full version of this paper [13].

II. Preliminaries

Let $\mathbb{F}_q$ be the finite field of q elements, where q is a power of prime, and $\mathbb{F}_q^* = \mathbb{F}_q \backslash \{0\}$. Let $[n] = \{1, 2, \ldots, n\}$. The

978-1-4577-0594-6/11/$26.00 ©2011 IEEE

support of a vector $\boldsymbol{u} \in \mathbb{F}_q^n$ is defined by $\mathsf{supp}(\boldsymbol{u}) \triangleq \{i \in [n] : u_i \neq 0\}$. Suppose $E \subseteq [n]$. We write $\boldsymbol{u} \lhd E$ whenever $\mathsf{supp}(\boldsymbol{u}) \subseteq E$. Let $\boldsymbol{e}_i$ denote the unit vector, which has a one at the ith position, and zeros elsewhere. In the sequel, we use many standard notions from coding theory such as (Hamming) weight, minimum distance, dual distance, linear $[n, k, d]_q$ codes, dual codes, MDS codes (for instance, see [14]). We recall the following well-known result in coding theory.

Theorem 2.1 ([15], 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 $r \leq d^\perp - 1$ 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 vector $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ and a subset $B = \{i_1, i_2, \ldots, i_b\}$ of $[n]$, where $i_1 < i_2 < \cdots < i_b$, let $\boldsymbol{Y}_B$ denote the vector $(Y_{i_1}, Y_{i_2}, \ldots, Y_{i_b})$. For an $n \times k$ matrix $\boldsymbol{M}$, let $\boldsymbol{M}_i$ denote the ith row of $\boldsymbol{M}$, and $\boldsymbol{M}[j]$ its jth column. For a set $E \subseteq [n]$, let $\boldsymbol{M}_E$ denote the $|E| \times k$ sub-matrix of $\boldsymbol{M}$ formed by rows of $\boldsymbol{M}$ which are indexed by the elements of E. For a set $F \subseteq [k]$, let $\boldsymbol{M}[F]$ denote the $n \times |F|$ sub-matrix of $\boldsymbol{M}$ formed by columns of $\boldsymbol{M}$ which are indexed by the elements of F.

III. Index Coding and Some Basic Results

The Index Coding with Side Information problem considers the following scenario. There is a unique sender (or source) S, who has a vector of messages $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{F}_q^n$ in his possession, which is a realized value of a random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$. $X_1, X_2, \ldots, X_n$ hereafter are assumed to be independent uniformly distributed random variables over $\mathbb{F}_q$. There are also m receivers $R_1, R_2, \ldots, R_m$. For each $i \in [m]$, R_i has some side information, i.e. R_i owns a subset of messages $\{x_j\}_{j \in \mathcal{X}_i}$, $\mathcal{X}_i \subsetneq [n]$. 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 $\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_m)$. An instance of the ICSI problem is given by a quadruple $(m, n, \mathcal{X}, f)$.

Definition 3.1: A (deterministic) *index code* (IC) 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 IC. When the IC $\mathfrak{E}$ is used, S broadcasts a vector $\mathfrak{E}(\boldsymbol{x})$ of length N over $\mathbb{F}_q$.

Definition 3.2: An IC of the shortest possible length is called *optimal*. An IC is said to be *linear* if its encoding function $\mathfrak{E}$ is a linear transformation over $\mathbb{F}_q$. In other words, $\mathfrak{E}(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{L}$, for all $\boldsymbol{x} \in \mathbb{F}_q^n$, where $\boldsymbol{L}$ is an $n \times N$ matrix over $\mathbb{F}_q$. The matrix $\boldsymbol{L}$ is called the *matrix corresponding to the IC* $\mathfrak{E}$. We also refer to $\mathfrak{E}$ as the *IC based on* $\boldsymbol{L}$. Notice that the length of $\mathfrak{E}$ is the number of columns of $\boldsymbol{L}$.

Hereafter, we assume that the sets $\mathcal{X}_i$, for all $i \in [m]$, are known to S. Moreover, we also assume that $\mathfrak{E}$ 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.

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 [3] for the case where $m = n$, $f(i) = i$ for all $i \in [m]$, and $q = 2$. However, it can be formulated in a more general form as follows.

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 $\boldsymbol{u} \in \mathbb{F}_q^n$ satisfying $\boldsymbol{u} \lhd \mathcal{X}_i$ and $\boldsymbol{u} + \boldsymbol{e}_{f(i)} \in \mathcal{C}(\boldsymbol{L})$.

Proof: Assume that $\boldsymbol{u} \lhd \mathcal{X}_i$ and $\boldsymbol{u} + \boldsymbol{e}_{f(i)} \in \mathcal{C}(\boldsymbol{L})$. Since $\boldsymbol{u} + \boldsymbol{e}_{f(i)} \in \mathcal{C}(\boldsymbol{L})$, there exist $\boldsymbol{\beta} \in \mathbb{F}_q^N$ such that $\boldsymbol{u} + \boldsymbol{e}_{f(i)} = \boldsymbol{\beta}\boldsymbol{L}^T$. By taking the transpose and pre-multiplying by $\boldsymbol{x}$, we obtain that $\boldsymbol{x}(\boldsymbol{u} + \boldsymbol{e}_{f(i)})^T = (\boldsymbol{x}\boldsymbol{L})\boldsymbol{\beta}^T$. Therefore, $x_{f(i)} = \boldsymbol{x}\boldsymbol{e}_{f(i)}^T = (\boldsymbol{x}\boldsymbol{L})\boldsymbol{\beta}^T - \boldsymbol{x}\boldsymbol{u}^T$. Observe that R_i is able to find $\boldsymbol{u}$ and $\boldsymbol{\beta}$ from the knowledge of $\boldsymbol{L}$. Moreover, R_i is also able to compute $\boldsymbol{x}\boldsymbol{u}^T$ since $\boldsymbol{u} \lhd \mathcal{X}_i$. 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 4.3 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 [3], [16]), 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 presence of an adversary A which can listen to all the transmissions. In other words, A knows $\boldsymbol{x}\boldsymbol{L}$. The adversary is assumed to possess side information $\{x_j\}_{j \in \mathcal{X}_A}$, where $\mathcal{X}_A \subsetneq [n]$ (A knows $\boldsymbol{x}_{\mathcal{X}_A}$). The strength of an adversary is defined to be $|\mathcal{X}_A|$. Denote $\widehat{\mathcal{X}}_A \triangleq ([n] \backslash \mathcal{X}_A)$.

Definition 4.1: Suppose that S possesses a vector of messages $\boldsymbol{x} \in \mathbb{F}_q^n$, which is a realized value of $\boldsymbol{X}$. Suppose also that 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 $B \subseteq \widehat{\mathcal{X}}_A$, A is said to *have no information about* $\boldsymbol{x}_B$ if $\mathsf{H}(\boldsymbol{X}_B | \boldsymbol{X}\boldsymbol{L}, \boldsymbol{X}_{\mathcal{X}_A}) = \mathsf{H}(\boldsymbol{X}_B)$, where $\mathsf{H}(\cdot)$ is a binary entropy function.
2) The IC is said to be *b-block secure against $\mathcal{X}_A$* if for every b-subset $B \subseteq \widehat{\mathcal{X}}_A$, A has no information about $\boldsymbol{x}_B$. It is said to be *b-block secure against all adversaries of strength t* ($t \leq n - 1$) if it is b-block secure against $\mathcal{X}_A$ for every $\mathcal{X}_A \subset [n]$, $|\mathcal{X}_A| = t$.
3) The IC is said to be *weakly secure against $\mathcal{X}_A$* if it is 1-block secure against $\mathcal{X}_A$. It is said to be *weakly secure*

against all adversaries of strength t ($t \leq n-1$) if it is weakly secure against $\mathcal{X}_A$ for every t-subset $\mathcal{X}_A$ of $[n]$.

4) The IC is said to be *completely insecure against* $\mathcal{X}_A$ if A is able to determine x_i for all $i \in \widehat{\mathcal{X}}_A$. It is said to be *completely insecure against any adversary of strength* t ($t \leq n-1$) if it is completely insecure against $\mathcal{X}_A$ for every t-subset $\mathcal{X}_A$ of $[n]$.

B. Necessary and Sufficient Conditions for Block Security

We introduce the following new lemma, which is a generalization of Lemma 3.1. It provides both necessary and sufficient conditions for successful reconstruction of the information by A. Note that A in Lemma 4.1 (and similarly in Theorem 4.7) can be viewed as a legitimate receiver. Thus, Lemma 4.1 also provides necessary and sufficient conditions for a receiver to be able (or not) to recover certain messages.

Lemma 4.1: Let $\boldsymbol{L}$ be an $n \times N$ matrix over $\mathbb{F}_q$ and let S broadcast $\boldsymbol{x}\boldsymbol{L}$. For a subset $B \subseteq \widehat{\mathcal{X}}_A = [n] \backslash \mathcal{X}_A$, the adversary A (or any participant who owns $\boldsymbol{x}_{\mathcal{X}_A}$), after listening to all transmissions, has no information about $\boldsymbol{x}_B$ if and only if

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

In particular, for each $i \in \widehat{\mathcal{X}}_A$, A has no information about x_i if and only if $\boldsymbol{u} + \boldsymbol{e}_i \notin \mathcal{C}(\boldsymbol{L})$ for all $\boldsymbol{u} \lhd \mathcal{X}_A$.

Corollary 4.2: Let $\boldsymbol{L}$ be an $n \times N$ matrix over $\mathbb{F}_q$ and assume that S broadcasts $\boldsymbol{x}\boldsymbol{L}$. Then for each $i \in [m]$, the receiver R_i can reconstruct $x_{f(i)}$ if and only if there exists $\boldsymbol{u}^{(i)} \in \mathbb{F}_q^n$ such that $\boldsymbol{u}^{(i)} \lhd \mathcal{X}_i$ and $\boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)} \in \mathcal{C}(\boldsymbol{L})$.

Corollary 4.3: 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 $\boldsymbol{u}^{(i)} \in \mathbb{F}_q^n$ satisfying $\boldsymbol{u}^{(i)} \lhd \mathcal{X}_i$ and $\boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)} \in \mathcal{C}(\boldsymbol{L})$.

Remark 4.4: It follows from Corollary 4.3 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 $\boldsymbol{u}^{(i)} \lhd \mathcal{X}_i$, $i \in [m]$. Define

$$\kappa_q = \kappa_q(m, n, \mathcal{X}, f) \triangleq \min\{\mathsf{rank}_q(\{\boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)}\}_{i \in [m]}) : \boldsymbol{u}^{(i)} \in \mathbb{F}_q^n, \boldsymbol{u}^{(i)} \lhd \mathcal{X}_i\},$$

then κ_q is the shortest possible length of a linear $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$. This 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 $i \in [n]$, which was introduced in [3], [17].

Corollary 4.5: The length of an optimal linear $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$ is $\kappa_q = \kappa_q(m, n, \mathcal{X}, f)$.

Theorem 4.6: 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 IC 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 IC is not weakly secure against at least one adversary of strength $t = d-1$. Generally, if there exists a codeword of $\mathcal{C}(\boldsymbol{L})$ of weight w, then this IC is not weakly secure against at least one adversary of strength $t = w-1$.
3) Every adversary of strength $t \leq d-1$ can determine a list of q^{n-t-N} vectors in $\mathbb{F}_q^n$ which includes $\boldsymbol{x}$.

Proof: We only prove part 1) here. Assume that $t \leq d-2$. By Lemma 4.1, 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 $\mathsf{supp}(\boldsymbol{u}) \cap B = \varnothing$ and α_i's, $i \in B$, are not all zero, we deduce that $\boldsymbol{u} + \sum_{i \in B} \alpha_i \boldsymbol{e}_i \neq \boldsymbol{0}$. Hence $\boldsymbol{u} + \sum_{i \in B} \alpha_i \boldsymbol{e}_i \notin \mathcal{C}(\boldsymbol{L})$. ■

C. Block Security and Complete Insecurity

In general, the IC based on $\boldsymbol{L}$ might still be block secure against some adversaries of strength t for $t \geq d$. 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 4.7: The linear IC based on $\boldsymbol{L}$ is completely insecure against any adversary of strength $t \geq n - d^{\perp} + 1$, where $d^{\perp}$ denotes the dual distance of $\mathcal{C}(\boldsymbol{L})$.

Proof: Suppose that $|\mathcal{X}_A| = t \geq n - d^{\perp} + 1$. By Corollary 4.2, it suffices to show that for each $j \in \widehat{\mathcal{X}}_A$, there exists $\boldsymbol{u} \in \mathbb{F}_q^n$ satisfying $\boldsymbol{u} \lhd \mathcal{X}_A$ and $\boldsymbol{u} + \boldsymbol{e}_j \in \mathcal{C}(\boldsymbol{L})$.

Indeed, take any $j \in \widehat{\mathcal{X}}_A$, and let $\rho = n - t \leq d^{\perp} - 1$. Consider the ρ indices which are not in $\mathcal{X}_A$. By Theorem 2.1, there exists a codeword $\boldsymbol{c} \in \mathcal{C}(\boldsymbol{L})$ with $c_j = 1$ and $c_\ell = 0$ if $\ell \notin \mathcal{X}_A \cup \{j\}$. 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 $u_\ell = 0$. Then $\boldsymbol{c} = \boldsymbol{u} + \boldsymbol{e}_j$. Hence by Corollary 4.2, the adversary can reconstruct x_j. ■

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 4.6 and 4.7 are actually tight. The following example further illustrates the results stated in these theorems.

Example 4.1: Let $n = m = 7$, $q = 2$, and $f(i) = i$ for all $i \in [m]$.

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 $\mathrm{supp}(\boldsymbol{u}^{(i)}) = \mathcal{X}_i$. Consider an IC based on $\boldsymbol{L}$ with $\mathcal{C}(\boldsymbol{L}) = \mathsf{span}_q(\{\boldsymbol{u}^{(i)} + \boldsymbol{e}_i\}_{i \in [7]})$. We can take $\boldsymbol{L}$ to be the matrix whose set of columns is $\{\boldsymbol{L}[i] \triangleq \boldsymbol{u}^{(i)} + \boldsymbol{e}_i\}_{i \in [4]}$. Then $\mathcal{C}(\boldsymbol{L})$ 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: $s_i = \boldsymbol{x}(\boldsymbol{u}^{(i)} + \boldsymbol{e}_i)^T$, $i \in [4]$. 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 $x_i = \boldsymbol{x}\boldsymbol{e}_i^T$. Since $d = 3$, if one message is leaked, the adversary has no information about any other particular message. If none of the messages are leaked, then the adversary has no information about any group of 2 messages. On the other hand, if $t \geq 4$ messages are leaked, the adversary is able to determine the remaining $7 - t$ messages.

D. Application: Index Coding with Side and Restricted Information

In this section, we consider an extension of the ICSI problem, which we call the *Index Coding with Side and Restricted Information (ICSRI)* problem. This problem arises in applications such as audio and video-on-demand. Consider a client who has subscribed to certain media content, and 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.

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

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

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

Proposition 4.8: 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 $\mathcal{C}(\boldsymbol{L}) \cap \mathcal{F}(m, n, \mathcal{X}, \mathcal{Z}, f) = \varnothing$.

Example 4.2: 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 4.1. Moreover, $\mathcal{Z} = (\mathcal{Z}_1, \mathcal{Z}_2, \ldots, \mathcal{Z}_7)$, where $\mathcal{Z}_1 = \{2, 3, 4, 5\}$, $\mathcal{Z}_2 = \{1, 3, 4, 6\}$, $\mathcal{Z}_3 = \{1, 2, 4, 7\}$, and $\mathcal{Z}_4 = \mathcal{Z}_5 = \mathcal{Z}_6 = \mathcal{Z}_7 = \varnothing$. Consider the IC based on $\boldsymbol{L}$ constructed in Example 4.1. It can be verify that $\mathcal{C}(\boldsymbol{L}) \cap \mathcal{F}(m, n, \mathcal{X}, \mathcal{Z}, f) = \varnothing$. Hence by Proposition 4.8, this IC provides a solution to this instance of the ICSRI problem.

Let

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

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

$$\mathsf{span}_q\left(\{\boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)}\}_{i \in [m]}\right) \cap \mathcal{F}(m, n, \mathcal{X}, \mathcal{Z}, f) = \varnothing. \quad (1)$$

Let $\kappa_q^* = +\infty$ if there are no choices of $\boldsymbol{u}^{(i)}$'s, $i \in [m]$, which satisfy (1). The following proposition follows immediately.

Proposition 4.9: 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 an adversary with an additional ability to corrupt some transmissions of S.

A. A Lower Bound on the Length

We first generalize the definition of ICs to *randomized ICs*. Let $\boldsymbol{G} = (G_1, G_2, \ldots, G_\eta)$ be a vector of η random variables which are distributed independently and uniformly over $\mathbb{F}_q$. Let $\boldsymbol{g} = (g_1, g_2, \ldots, g_\eta)$ be a realization of $\boldsymbol{G}$.

Definition 5.1: 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

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

An η-randomized IC is linear over $\mathbb{F}_q$ if its encoding function is linear, i.e. $\mathfrak{E}(\boldsymbol{x}, \boldsymbol{g}) = (\boldsymbol{x}|\boldsymbol{g})\boldsymbol{L}$, where $\boldsymbol{L}$ is an $(n + \eta) \times N$ matrix over $\mathbb{F}_q$. Observe that by simply treating x_i's and g_i's as messages, the results from previous sections still apply to linear randomized ICs.

Definition 5.2: 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*, i.e., 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*, i.e., an adversary A who possesses $\boldsymbol{x}_{\mathcal{X}_A}$ ($|\mathcal{X}_A| = t$), and listens to at most μ transmissions ($\mu \leq N$) gains no information about other messages. Equivalently, for any $W \subseteq [N]$, $|W| \leq \mu$,

$$\mathsf{H}(\boldsymbol{X}_{\widehat{\mathcal{X}}_A} \mid (\boldsymbol{X}|\boldsymbol{G})\boldsymbol{L}[W], \boldsymbol{X}_{\mathcal{X}_A}) = \mathsf{H}(\boldsymbol{X}_{\widehat{\mathcal{X}}_A}). \quad (2)$$

Lemma 5.1: If $\boldsymbol{L}$ corresponds to a (μ, t)-strongly secure linear η-randomized $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$, then $\eta \geq \mu$.

Sketch of the proof: 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 $|W| \leq \mu$, the equality (2) holds. From Lemma 4.1 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 $\boldsymbol{c}_E = \boldsymbol{0}$ (we denote this 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. It is possible to show that $\mathsf{rank}_q(\boldsymbol{L}') \leq \mu - 1$. Now

let $\mathsf{r} \triangleq \mathsf{rank}_q(\boldsymbol{L}')$, and let $\{\boldsymbol{L}'_{j_1}, \boldsymbol{L}'_{j_2}, \ldots, \boldsymbol{L}'_{j_\mathsf{r}}\}$ be a basis of the space spanned by the rows of $\boldsymbol{L}'$.

- On one hand, by Corollary 4.3, $\mathcal{C}(\boldsymbol{L})$ contains a vector $\boldsymbol{c} = \boldsymbol{u}^{(i)} + \boldsymbol{e}_{f(i)}$ where $\boldsymbol{u}^{(i)} \lhd \mathcal{X}_i$ and $f(i) \in \widehat{\mathcal{X}}_A$. Therefore, $\boldsymbol{c}_E = \boldsymbol{0}$ and $\boldsymbol{c}_{\widehat{\mathcal{X}}_A} \neq \boldsymbol{0}$.
- On the other hand, there exist $\beta_1, \beta_2, \ldots, \beta_\mathsf{r}$ 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 5.2: From Lemma 5.1, a (μ, t, δ)-strongly secure linear randomized IC requires at least μ random symbols. We show in Section V-B that there exists a (μ, t, δ)-strongly secure IC that uses precisely μ random symbols.

Lemma 5.3: Suppose that $\boldsymbol{L}$ corresponds to a linear μ-randomized $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$. If this randomized IC 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 $\boldsymbol{v}^{(i)} \lhd [n]$ and $\boldsymbol{v}^{(i)} + \boldsymbol{e}_{n+i} \in \mathcal{C}(\boldsymbol{L})$.

Sketch of the 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 who owns $\{x_j\}_{j \in [n]}$ and requests the symbol g_i. By Corollary 4.2, this virtual receiver has no information about g_i after listening to all transmissions. It can be shown that discarding G_i from the scheme does not affect its strong security. However, this contradicts Lemma 5.1, since the resulting code has less than μ random symbols. ∎

Theorem 5.4: The length of a (μ, t)-strongly secure linear η-randomized $(m, n, \mathcal{X}, f)$-IC over $\mathbb{F}_q$ is at least $\kappa_q + \mu$.

Sketch of the proof: If $\eta = \mu$ then by Corollary 4.3 and Lemma 5.3, the length of the code is at least

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

Similar argument applies to the case when $\eta > \mu$ and for all $i \in [\eta]$ there exists some $\boldsymbol{v}^{(i)} \lhd [n]$ such that $\boldsymbol{v}^{(i)} + \boldsymbol{e}_{n+i} \in \mathcal{C}(\boldsymbol{L})$. For the remaining case, we keep discarding some random variable G_i from the code, until we reach a code which falls into the first two cases. ∎

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

Theorem 5.5: 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 $\kappa_q + \mu + 2\delta$.

B. 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 $\mathbb{F}_q$ whose length attaining the lower bound established in Theorem 5.5. The proposed construction is based on the coset coding technique, originally introduced by Ozarow and Wyner [18].

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, $\boldsymbol{M}$ can be chosen to be a generator matrix of a Reed-Solomon code). Let $\boldsymbol{P}$ be the sub-matrix of $\boldsymbol{M}$ formed by the first κ_q rows, and $\boldsymbol{Q}$ the sub-matrix formed by the last μ rows of $\boldsymbol{M}$. Take

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{L}^{(0)} \boldsymbol{P} \\ \hline \boldsymbol{Q} \end{pmatrix}.$$

Theorem 5.6: 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. Acknowledgements

The authors would like to thank Frédérique Oggier for helpful discussions. This work is supported by the National Research Foundation of Singapore (Research Grant NRF-CRP2-2007-03).

References

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

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

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

[4] ——, "Index coding with side information," *IEEE Trans. Inform. Theory*, to appear.

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

[6] 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 Inform. Theory Workshop (ITW)*, 2007, pp. 120–125.

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

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

[9] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1204–1216, 2000.

[10] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, pp. 782–795, 2003.

[11] D. Silva and F. R. Kschischang, "Universal secure error-correcting schemes for network coding," in *Proc. IEEE Symp. on Inform. Theory (ISIT)*, Austin, Texas, USA, 2010, pp. 2428–2432.

[12] A. E. Rouayheb and E. Soljanin, "On wiretap networks II," in *Proc. IEEE Symp. on Inform. Theory (ISIT)*, Nice, France, 2007, pp. 551–555.

[13] S. H. Dau, V. Skachek, and Y. M. Chee, "On the security of index coding with side information," available at `http://arxiv.org/abs/1101.2728`, 2011.

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

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

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

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

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