Secure Index Coding: Existence and Construction

Lawrence Ong[1], Badri N. Vellambi[2], Phee Lep Yeoh[3], Jörg Kliewer[2], and Jinhong Yuan[4]
[1]The University of Newcastle, Australia; [2]New Jersey Institute of Technology, USA;
[3]University of Melbourne, Australia; [4]University of New South Wales, Australia

***Abstract*—We investigate the construction of weakly-secure index codes for a sender to send messages to multiple receivers with side information in the presence of an eavesdropper. We derive a sufficient and necessary condition for the existence of index codes that are secure against an eavesdropper with access to any subset of messages of cardinality t, for any fixed t. In contrast to the benefits of using random keys in secure network coding, we prove that random keys do not promote security in three classes of index-coding instances.**

I. Introduction

In *classical*[1] index-coding problems, a sender sends multiple messages to multiple receivers through a common noiseless broadcast medium, where each receiver has a priori knowledge of a subset of messages [1]–[5]. The subsets that each receiver wants and knows can vary with the receiver. In this work, we consider *secure* index coding, where in addition to the classical setup, there is an eavesdropper who has access to a subset of messages from a collection of subsets of messages. The sender and the receivers know a priori the collection of message subsets, however, they do not know which subset of messages in this collection is actually accessed by the eavesdropper. A *weakly-secure* index code must satisfy all receivers' decoding requirements, while ensuring that the eavesdropper is not able to decode any message it has no access to.

A. Contributions of this paper and related work

The contributions of this work are three-fold:

1) Existence of secure index codes: Secure index coding was first studied by Dau, Skachek, and Chee [6]. The authors derived conditions that any given linear code (of a given message alphabet size) must satisfy to simultaneously meet the receivers' decoding requirements as well as be secure against an eavesdropper with access to a message subset.

In contrast to the code-centric results by Dau et al., we obtain problem-centric results. We derive a sufficient and necessary condition for the existence of both *linear and non-linear* weakly-secure index codes over *all finite-field alphabets* for any index-coding problem where the eavesdropper can access any message subset of cardinality t. We show how to construct such codes if they exist, and investigate their optimality.

2) Random keys: It has been shown [9] that there exist randomised secure network codes (using random keys) for instances where no deterministic secure network code exists. Owing to an equivalence between classical versions of network and index coding [7, 8], it is plausible that there exist index-coding instances where randomised encoding can enable security when deterministic encoding cannot. While we do not identify an instance where this is true, we have proven that random keys are *not* useful for weakly-secure index codes in the following three cases: *(i)* the eavesdropper has access to any t messages, *(ii)* the sender's encoding function is linear, or *(iii)* the eavesdropper has access to only one message subset.

3) Secure vs classical index coding: We highlight a significant difference between classical and secure index coding. In classical index coding, messages not required at any receiver are not useful and can be removed from the system. In weakly-secure index coding, these messages may be used as keys.

This work is supported by ARC grants FT140100219, DE140100420, and DP150100903, and US NSF grants CNS-1526547 and CCF-1439465.

[1]We use the term classical to indicate the absence of any security constraints.

II. Problem Definition and Notation

Let $m, n \in \mathbb{N}$. For each $i \in [n] \triangleq \{1, \ldots, n\}$, define two subsets $\mathcal{K}_i, \mathcal{W}_i \subseteq [m]$. A classical index-coding instance $(\mathcal{K}_i, \mathcal{W}_i)_{i=1}^n$ consists of a single sender and n receivers. The sender has m messages $\boldsymbol{X} = [X_1\, X_2\, \cdots X_m]$, where $\{X_i\}_{i=1}^m$ are independent and uniformly distributed over a finite field $\mathcal{F}_q$ with q elements. For a subset of integers $\mathcal{I} = \{i_1, i_2, \ldots, i_{|\mathcal{I}|}\}$ where $i_1 < i_2 < \cdots < i_{|\mathcal{I}|}$, let $\boldsymbol{X}_\mathcal{I} \triangleq [X_{i_1} X_{i_2} \cdots X_{i_{|\mathcal{I}|}}]$. Each receiver $i \in [n]$ has a priori knowledge of $\boldsymbol{X}_{\mathcal{K}_i}$, and needs to decode $\boldsymbol{X}_{\mathcal{W}_i}$. The sender encodes $\boldsymbol{X}$ and gives the codeword to all receivers. The codeword must be chosen so that each receiver $i \in [n]$ is able to decode the messages $\boldsymbol{X}_{\mathcal{W}_i}$ it wants using the codeword and the messages $\boldsymbol{X}_{\mathcal{K}_i}$ it already knows. Without loss of generality, we may assume that $\mathcal{W}_i \setminus \mathcal{K}_i \neq \emptyset$ for all $i \in [n]$, since receivers wanting only messages they already know can be expunged from the problem.

Let $\mathfrak{A} \subseteq 2^{[m]}$, where $2^{[m]}$ is the set of all subsets of $[m]$. A secure index-coding instance $((\mathcal{K}_i, \mathcal{W}_i)_{i=1}^n, \mathfrak{A})$ is a classical index-coding instance $(\mathcal{K}_i, \mathcal{W}_i)_{i=1}^n$ in the added presence of an eavesdropper who can access the sender's codeword and precisely one subset of messages $\boldsymbol{X}_\mathcal{A}$, where $\mathcal{A} \in \mathfrak{A}$. The eavesdropper cannot simultaneously access messages corresponding to the indices contained in more than one member of $\mathfrak{A}$. The set $\mathfrak{A}$ contains the possible subsets of indices of compromised messages. While the sender and the receivers are aware of $\mathfrak{A}$, they are oblivious to the exact subset of indices the eavesdropper knows. In addition to meeting the receivers' decoding requirements, a weakly-secure index codeword must ensure that the eavesdropper gains no additional information about each individual message $\boldsymbol{X}_j$, $j \in [m] \setminus \mathcal{A}$, given $\boldsymbol{X}_\mathcal{A}$ and the codeword. Formally, we have the following:

Definition 1 (Deterministic weakly-secure index code): Given a secure index-coding instance $((\mathcal{K}_i, \mathcal{W}_i)_{i=1}^n, \mathfrak{A})$, a determinis-

tic weakly-secure index code $(f_i, \{g_i\}_{i=1}^n)$ of codelength $\ell \in \mathbb{N}$ consists of

- an encoding function for the sender, $f : \mathcal{F}_q^m \to \mathcal{F}_q^\ell$, to encode $\boldsymbol{X}$ into $\boldsymbol{C} \triangleq f(\boldsymbol{X})$, and
- a decoding function for each receiver $i \in [n]$, $g_i : \mathcal{F}_q^\ell \times \mathcal{F}_q^{|\mathcal{K}_i|} \to \mathcal{F}_q^{|\mathcal{W}_i|}$, to decode $\boldsymbol{X}_{\mathcal{W}_i}$ from $\boldsymbol{C}$ and $\boldsymbol{X}_{\mathcal{K}_i}$

such that

- decodability: $g_i(f(\boldsymbol{X}), \boldsymbol{X}_{\mathcal{K}_i}) = \boldsymbol{X}_{\mathcal{W}_i}$ for each $i \in [n]$; and
- weak security: for all $\mathcal{A} \in \mathfrak{A}$, an eavesdropper accessing $\boldsymbol{X}_{\mathcal{A}}$ has no information about any single message in $\mathcal{A}^{\mathrm{c}} \triangleq [m] \backslash \mathcal{A}$, i.e., $H(X_i | f(\boldsymbol{X}), \boldsymbol{X}_{\mathcal{A}}) = H(X_i)$, for all $i \in \mathcal{A}^{\mathrm{c}}$. ∎

Remark 1: If $\mathfrak{A} = \{[m]\}$, we have a classical index-coding instance without any security constraint. ∎

The notion of weak security considered here, also known as 1-*block weakly secure* in the literature [6, 10], does not preclude the eavesdropper from gaining information about $\boldsymbol{X}_{\mathcal{A}^{\mathrm{c}}}$ despite gaining no knowledge about any single message thereof. Other notions of security have also been considered in the literature. For example, Mojahedian, Aref, and Gohari [11] considered *strongly*-secure index coding, where the eavesdropper has *no access* to any message, and must not gain any information about the messages $\boldsymbol{X}$. Their approach involves the sender encoding messages with keys that are pre-shared with the receivers, but are unknown to the eavesdropper.

It may be possible for the sender to use *random keys* along with the messages $\boldsymbol{X}$ during the encoding process to ensure security against the eavesdropper. We therefore introduce the following notion of random weakly-secure index codes that generalise deterministic weakly-secure index codes.

Definition 2 (Random weakly-secure index code): Let Y be a random variable taking values in a finite alphabet $\mathcal{Y}$ known only to the the sender, and unknown to the receivers and the eavesdropper. A random weakly-secure index code $(f_i, \{g_i\}_{i=1}^n)$ of codelength $\ell \in \mathbb{N}$ is identical to the deterministic index-code setup with the only exception that the sender encodes $\boldsymbol{X}$ into $\boldsymbol{C} \triangleq f(\boldsymbol{X}, Y)$ using the function $f : \mathcal{F}_q^m \times \mathcal{Y} \mapsto \mathcal{F}_q^\ell$. The decoding operations, decodability conditions, and security conditions are identical to those in Definition 1.

For the rest of this paper, unless otherwise stated, by secure index codes, we mean weakly-secure index codes.

Definition 3 (Linear index code): A random index code is linear if and only if the key $\boldsymbol{Y} = [Y_1\, Y_2\, \cdots Y_k]$ for some $k \in \mathbb{N}$, where $\{Y_i\}_{i=1}^k$ are independent and uniformly distributed over $\mathcal{F}_q$, and the encoding function

$$\boldsymbol{C} \triangleq f(\boldsymbol{X}, \boldsymbol{Y}) = \boldsymbol{X}\mathbb{G} + \boldsymbol{Y}\tilde{\mathbb{G}}, \tag{1}$$

for some matrices $\mathbb{G}$ and $\tilde{\mathbb{G}}$ over $\mathcal{F}_q$ of sizes $m \times \ell$ and $k \times \ell$, respectively. Similarly, a deterministic index code is linear if and only if the encoding function $f(\boldsymbol{X}) = \boldsymbol{X}\mathbb{G}$. ∎

We say that a secure index code exists for a secure index-coding instance $I = ((\mathcal{K}_i, \mathcal{W}_i)_{i=1}^n, \mathfrak{A})$ if and only if there exists a (deterministic or random) secure index code $(f, (g_i)_{i=1}^n)$ for some q that meets all the conditions in Definition 1. If one such code exists, we say that the code is *secure against* an eavesdropper having access to any message subset in $\mathfrak{A}$. As we will see later, a secure index code may or may not exist depending on $\mathfrak{A}$. The optimal secure index codelength $s(I)$ for a secure index-coding instance I, for which secure index codes exist, is defined as infimum of the codelengths of secure index codes over all alphabet sizes.

III. Fundamental Properties

We begin with the following counter-intuitive proposition:

Proposition 1: Let $\mathcal{A}' \subsetneq \mathcal{A} \subsetneq [m]$. An index code secure against an eavesdropper who knows $\boldsymbol{X}_{\mathcal{A}}$ may not be secure against an eavesdropper who knows $\boldsymbol{X}_{\mathcal{A}'}$, and vice versa.

Proof: The following example proves this claim. Consider four receivers, where $\mathcal{W}_i = \{i\}$, for all $i \in [4]$, $\mathcal{K}_1 = \{2\}$, $\mathcal{K}_2 = \{1\}$, $\mathcal{K}_3 = \{2, 4\}$, $\mathcal{K}_4 = \{2, 3\}$. Consider two eavesdroppers: the first eavesdropper has access to $\mathfrak{A}_1 = \{\{3, 4\}\}$; the second eavesdropper has access to $\mathfrak{A}_2 = \{\{3\}\}$. The index code $\boldsymbol{C}_1 \triangleq f_1(\boldsymbol{X}) = [X_1 + X_2 \;\; X_3 + X_4]$, where $+$ denotes addition of the finite field $\mathcal{F}_q$, is secure against the first eavesdropper (because $H(X_i | \boldsymbol{C}_1, X_3, X_4) = H(X_i)$ for each $i \in \{1, 2\}$) but not the second eavesdropper (because it can decode X_4). The index code $\boldsymbol{C}_2 \triangleq f_2(\boldsymbol{X}) = [X_1 + X_2 \;\; X_2 + X_3 + X_4]$ is secure against the second eavesdropper, but not the first. ∎

Proposition 1 is in contrast to secure network coding [9], where a network code that is strongly secure against an eavesdropper who can access a subset of links, say $\mathcal{L}$, is also secure against an eavesdropper who can access any $\mathcal{L}' \subsetneq \mathcal{L}$.

Proposition 2: No secure index code exists for any secure index-coding instance $((\mathcal{K}_i, \mathcal{W}_i)_{i=1}^n, \mathfrak{A})$ where there exists $\mathcal{A} \in \mathfrak{A}$ and $i \in [n]$ such that $\mathcal{K}_i \subseteq \mathcal{A}$ and $\mathcal{W}_i \cap \mathcal{A}^{\mathrm{c}} \neq \emptyset$.

Proof: Pick $j \in \mathcal{W}_i \cap \mathcal{A}^{\mathrm{c}}$. Let $\boldsymbol{C} = f(\boldsymbol{X}, Y)$ be the codeword of a random index code that uses a key Y. Since $H(X_j | \boldsymbol{C}, \boldsymbol{X}_{\mathcal{A}}) \leq H(X_j | \boldsymbol{C}, \boldsymbol{X}_{\mathcal{K}_i}) = 0$, the eavesdropper is able to decode $\boldsymbol{X}_j$. Thus, the code cannot be secure. ∎

IV. Existence of Secure Index Codes

Here, we present a necessary and sufficient condition for the existence of secure index codes, and their construction. Furthermore, we derive optimal secure index codes for certain classes of instances. We begin with a specific type of eavesdroppers.

Definition 4: For a given $(\mathcal{K}_i, \mathcal{W}_i)_{i=1}^n$, we say that an index code is *secure against an eavesdropper with t-level access*, for some $t \in \{0, 1, \ldots, m-1\}$, if and only if it is a secure index code for $((\mathcal{K}_i, \mathcal{W}_i)_{i=1}^n, \{\mathcal{A} \subsetneq [m] : |\mathcal{A}| = t\})$. ∎

Lemma 1: Any (deterministic or random) index code secure against an eavesdropper with t-level access is also secure against an eavesdropper with t'-level access, for any $t' < t$.

Proof: Consider an eavesdropper with t-level access who has access to any member of $\mathfrak{A} = \{\mathcal{A} \subsetneq [m] : |\mathcal{A}| = t\}$. An index code secure against this eavesdropper must satisfy

$$H(X_i | \boldsymbol{C}, \boldsymbol{X}_{\mathcal{A}}) = H(X_i), \quad \text{for all } \mathcal{A} \in \mathfrak{A}, i \in \mathcal{A}^{\mathrm{c}}. \tag{2}$$

Consider an eavesdropper with an access level $t' < t$ who has access to any member of $\mathfrak{A}' = \{\mathcal{B} \subsetneq [m] : |\mathcal{B}| = t'\}$. Pick any $\mathcal{A}' \in \mathfrak{A}'$ and any $i \in [m] \setminus \mathcal{A}'$. As $t' < t \leq m - 1$, we can always find a subset $\mathcal{A} \in \mathfrak{A}$ such that $\mathcal{A}' \subsetneq \mathcal{A}$ and $i \notin \mathcal{A}$. So,

$$H(X_i) \overset{(a)}{=} H(X_i | \boldsymbol{C}, \boldsymbol{X}_{\mathcal{A}}) \overset{(b)}{\leq} H(X_i | \boldsymbol{C}, \boldsymbol{X}_{\mathcal{A}'}) \overset{(c)}{\leq} H(X_i),$$

where (a) follows from (2), and (b) and (c) follow since conditioning cannot increase entropy. Since the choices of $\mathcal{A}'$ and i are arbitrary, we must have $H(X_i|\boldsymbol{C}, \boldsymbol{X}_{\mathcal{A}'}) = H(X_i)$, for all $\mathcal{A}' \in \mathfrak{A}'$ and all $i \in [m] \setminus \mathcal{A}'$. Thus, the index code is also secure against an eavesdropper with t'-level access. ■

Remark 2: Lemma 1 generalises the result by Dau et al. [6, Theorem 4.9] that pertains specifically to deterministic linear index codes to any (random or deterministic, linear or non-linear) index code. ■

Remark 3: Although Proposition 1 states that an index code secure against an eavesdropper with access to $\mathfrak{A} = \{\mathcal{A}\}$ may not be secure against an eavesdropper with access to $\mathfrak{A} = \{\mathcal{A}'\}$ where $\mathcal{A}' \subsetneq \mathcal{A}$, any index code secure against an eavesdropper with t-level access, i.e., $\mathfrak{A} = \{\mathcal{A} \subsetneq [m] : |\mathcal{A}| = t\}$, *is also* secure against an eavesdropper with any access level $t' < t$.■

A. Existence of secure index codes and their construction

We now present a necessary and sufficient condition for the existence of secure index codes.

Theorem 1: Consider a secure index-coding instance $((\mathcal{K}_i, \mathcal{W}_i)_{i=1}^n, \mathfrak{A})$ with $\mathfrak{A} = \{\mathcal{A} \subsetneq [m] : |\mathcal{A}| = t\}$ for some $t < m$, i.e., the eavesdropper has t-level access. Secure index codes exist if and only if

$$t < K_{\min} \triangleq \min_{i\in[n]} |\mathcal{K}_i|. \tag{3}$$

Deterministic linear secure index codes exist if (3) is satisfied.

Proof: We first prove the converse. Suppose that $t \geq K_{\min}$. By definition, there exists a receiver, say i, with $|\mathcal{K}_i| = K_{\min}$, and $\mathcal{W}_i \setminus \mathcal{K}_i \neq \emptyset$. Pick some $j \in \mathcal{W}_i \setminus \mathcal{K}_i$, i.e., $\mathcal{K}_i \subseteq [m] \setminus \{j\}$. Since $K_{\min} \leq t \leq m-1$, we can always find some $\mathcal{A} \in \mathfrak{A}$ such that $\mathcal{K}_i \subseteq \mathcal{A} \subseteq [m] \setminus \{j\}$. From Proposition 2, we conclude that no secure index code exists.

Next, we prove the forward part. Consider a deterministic linear index code of length $\ell = m - K_{\min}$, formed by

$$\boldsymbol{C} = \boldsymbol{X}\mathbb{G} = \sum_{i\in[m]} X_i \boldsymbol{g}_i, \tag{4}$$

where $\mathbb{G}$ is an $m \times \ell$ matrix over $\mathcal{F}_q$, and $\boldsymbol{g}_i \in \mathcal{F}_q^\ell$ is the i-th row of $\mathbb{G}$. Let $\mathbb{G}$ be the *transpose* of the generating matrix of a maximum-distance-separable (MDS) code, which always exists for a sufficiently large q. For any such code, it follows that any ℓ rows of $\mathbb{G}$ are linearly independent over $\mathcal{F}_q$.

(Decoding) Receiver $i \in [n]$ forms

$$\boldsymbol{C} - \sum_{k\in\mathcal{K}_i} X_k \boldsymbol{g}_k = \sum_{j\in[m]\setminus\mathcal{K}_i} X_j \boldsymbol{g}_j. \tag{5}$$

Since $|[m] \setminus \mathcal{K}_i| = m - |\mathcal{K}_i| \leq m - K_{\min} = \ell$, it follows that $\{\boldsymbol{g}_j : j \in [m] \setminus \mathcal{K}_i\}$ are linearly independent. So, by using $\boldsymbol{C}$ and $\{X_k : k \in \mathcal{K}_i\}$, receiver i can decode $\boldsymbol{X}_{[m]\setminus\mathcal{K}_i}$, and consequently the message(s) $\boldsymbol{X}_{\mathcal{W}_i}$ it wants, by solving (5).

(Security) Denote the Hamming distance between two vectors $\boldsymbol{a} = [a_1 a_2 \cdots a_m] \in \mathcal{F}_q^m$ and $\boldsymbol{b} = [b_1 b_2 \cdots b_m] \in \mathcal{F}_q^m$ by $d(\boldsymbol{a}, \boldsymbol{b}) \triangleq |\{i \in [m] : a_i \neq b_i\}|$, the minimum distance of a vector space $\mathcal{S}$ by $d(\mathcal{S}) = \min_{\{\boldsymbol{a},\boldsymbol{b}\in\mathcal{S}:\boldsymbol{a}\neq\boldsymbol{b}\}} d(\boldsymbol{a}, \boldsymbol{b})$, and the vector space spanned by the rows and columns of a matrix $\mathbb{M}$ by $\mathsf{rowsp}(\mathbb{M})$ and $\mathsf{colsp}(\mathbb{G})$, respectively. Dau et al. [6] showed that the linear index code of the form (4) is secure against an eavesdropper with an access level $d(\mathsf{colsp}(\mathbb{G})) - 2$. Note that $\mathbb{G}^{\mathrm{T}}$ is the generator matrix of an MDS code (m, ℓ, d) whose codewords are vectors in $\mathsf{rowsp}(\mathbb{G}^{\mathrm{T}})$. The minimum distance of this MDS code equals $d = d(\mathsf{rowsp}(\mathbb{G}^{\mathrm{T}})) = d(\mathsf{colsp}(\mathbb{G})) = m - \ell + 1$. Invoking Lemma 1, we see that the index code (4) is secure against an eavesdropper with an access level up to and including $(m - \ell + 1) - 2 = K_{\min} - 1$. ■

Some remarks are now in order.

Remark 4: MDS codes are also used in the partial-clique-cover coding scheme [1] and its time-shared version [4], and the local-chromatic-number coding scheme [12] for *unicast* index coding, where $\mathcal{W}_i = \{i\}$ for all receivers $i \in [n]$. ■

Remark 5: Receiver cooperation can increase the security level. Allowing two receivers, say i and j, to cooperate and share their messages is equivalent to solving a new secure index-coding instance where everything remains the same except that receivers i and j both know $\boldsymbol{X}_{\mathcal{K}_i\cup\mathcal{K}_j}$. Thus, cooperation can potentially increase $K_{\min}$ (see (3)), which then translates to security against eavesdroppers with higher access levels. ■

Remark 6: Theorem 1 also holds if we consider b-block security (see Dau et al. [6]) for $b \geq 1$ in Definition 1. In the setting of b-block security, an eavesdropper who knows $\boldsymbol{X}_{\mathcal{A}}$, $\mathcal{A} \in \mathfrak{A}$, gains no information about any b messages it does not know, i.e., $H(\boldsymbol{X}_{\mathcal{B}}|\boldsymbol{C}, \boldsymbol{X}_{\mathcal{A}}) = H(\boldsymbol{X}_{\mathcal{B}})$, for all $\mathcal{B} \subseteq \mathcal{A}^{\mathrm{c}}$ with $|\mathcal{B}| = b$. In this case, the necessary and sufficient condition for the existence of secure index codes in (3) is replaced by $t \leq K_{\min} - b$. ■

Corollary 1.1: If $A_{\max} \triangleq \max_{\mathcal{A}\in\mathfrak{A}} |\mathcal{A}| < K_{\min}$, then deterministic linear secure index codes exist.

Proof: Proof follows from Theorem 1 and Lemma 1. ■

Intuitively, Corollary 1.1 says that we can always find secure index codes if the eavesdropper can access fewer messages than each receiver can. However, unlike Theorem 1, we do not have a converse for Corollary 1.1. This is because even if an eavesdropper can access (numerically) more messages than some receivers can, we may still be able to construct secure index codes, depending on the sets of messages to which the eavesdropper has access. For example, see the secure index codes for the two instances with $A_{\max} \geq K_{\min}$ in the proof of Proposition 1.

B. Optimality of secure index codes

From the construction of secure index codes in Theorem 1, we have the following:

Corollary 1.2: If $A_{\max} < K_{\min}$, the optimal secure index codelength is upper-bounded as $s(I) \leq m - K_{\min}$. The upper bound is achievable by deterministic linear index codes.

Proof: See the proof of Theorem 1. ■

We say that a receiver i has *complementary message requests* if it wants all messages it does not know, i.e., $\mathcal{K}_i \cup \mathcal{W}_i = [m]$.

Proposition 3: If $A_{\max} < K_{\min}$, and if any receiver knowing exactly $K_{\min}$ messages has complementary message requests, then the optimal codelength $s(I) = m - K_{\min}$, and is achievable by deterministic secure linear index codes.

Proof: Without loss of generality, let $|\mathcal{K}_1| = K_{\min}$ and $\mathcal{K}_1 \cup \mathcal{W}_1 = [m]$. For any (deterministic or random) index code $\boldsymbol{C}$, we have $H(\boldsymbol{X}_{\mathcal{W}_1}|\boldsymbol{C}, \boldsymbol{X}_{\mathcal{K}_1}) = 0$. Thus,

$$\begin{aligned} m \log_2 q = H(\boldsymbol{X}) &\le H(\boldsymbol{C}, \boldsymbol{X}_{\mathcal{K}_1}, \boldsymbol{X}_{\mathcal{W}_1}) = H(\boldsymbol{C}, \boldsymbol{X}_{\mathcal{K}_1}) \\ &\le \log_2 q^{\ell} + \log_2 q^{K_{\min}}, \end{aligned}$$

where we have made use of the facts that $\boldsymbol{X} = \boldsymbol{X}_{\mathcal{K}_1 \cup \mathcal{W}_1}$, $H(\boldsymbol{X}_{\mathcal{W}_1}|\boldsymbol{C}, \boldsymbol{X}_{\mathcal{K}_1}) = 0$, $H(\boldsymbol{C}) \le \log_2 |\boldsymbol{C}| = \log_2 q^{\ell}$, and $H(\boldsymbol{X}_{\mathcal{K}_1}) = \log_2 q^{K_{\min}}$. Therefore, $\ell \ge m - K_{\min}$. Since $\boldsymbol{C}$ was arbitrary, it follows that $s(I) = \inf \ell \ge m - K_{\min}$. The proof is then complete by invoking Corollary 1.2. ∎

V. Secure vs Classical Index Coding

We can represent a secure index-coding instance $((\mathcal{G}_i, \mathcal{K}_i)_{i=1}^n, \mathfrak{A})$ by a directed bipartite graph $\mathcal{D} = (\mathcal{U}, \mathcal{M}, \mathcal{E})$, similar to that by Neely, Tehrani, and Zhang [13]. Here, $\mathcal{U}$ and $\mathcal{M}$ are independent vertex sets, where each arc (i.e., directed edge) in $\mathcal{E}$ connects a vertex in $\mathcal{U}$ to a vertex in $\mathcal{M}$. We further partition $\mathcal{U}$ into two disjoint sets: $\mathcal{R} = \{r_1, r_2, \ldots, r_n\}$ representing the n receivers, and $\mathcal{V} = \{v_1, v_2, \ldots, v_{|\mathfrak{A}|}\}$ representing the possible sets of messages to which the eavesdropper can access. The set $\mathcal{M} = [m]$ represents the message indices. The arc set $\mathcal{E}$ is defined as follows:

- There is an arc from $r_i \in \mathcal{R}$ to $j \in \mathcal{M}$ if and only if receiver i knows the message X_j, i.e., $j \in \mathcal{K}_i$.
- There is an arc from $j \in \mathcal{M}$ to $r_i \in \mathcal{R}$ if and only if receiver i wants the message X_j, i.e., $j \in \mathcal{W}_i$.
- For each $\mathcal{A} \in \mathfrak{A}$, we have a unique $v_i \in \mathcal{V}$ such that $\mathcal{N}_{\mathcal{D}}^+(v_i) = \mathcal{A}$, where $\mathcal{N}_{\mathcal{D}}^+(v_i) \triangleq \{j \in \mathcal{M} : (v_i \to j) \in \mathcal{E}\}$ is the out-neighbourhood of v_i.

For a given secure index-coding instance $\mathcal{D}$, if we ignore the security constraint, the subgraph $\mathcal{D}[\mathcal{R} \cup \mathcal{M}]$ induced by $(\mathcal{R}, \mathcal{M})$ is in fact the bipartite graph used by Neely et al. [13] to represent the classical index-coding instance.

Proposition 4: Consider a secure index-coding instance $((\mathcal{G}_i, \mathcal{K}_i)_{i=1}^n, \mathfrak{A})$, where $\mathfrak{A} \ne \{[m]\}$. Let $\mathcal{D} = ((\mathcal{R}, \mathcal{V}), \mathcal{M}, \mathcal{E})$ be its directed bipartite graph representation. If

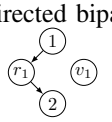
(C1) $\mathcal{D}[\mathcal{R} \cup \mathcal{M}]$ is acyclic, or equivalently, $\mathcal{D}$ is acyclic, and
(C2) every message is wanted by some receiver, i.e., for each $i \in [m]$, we have $i \in \mathcal{W}_j$ for some $j \in [n]$,

then no secure index code exists.

Proof: For the classical index-coding instance $\mathcal{D}[\mathcal{R} \cup \mathcal{M}]$, Neely et al. [13, Appendix A] have shown that if condition C1 is true (condition C2 is always assumed to be true for non-secure index coding), one can obtain all messages from any index code, even without using side information. Since any secure index code for $\mathcal{D}$, denoted by $\boldsymbol{C}$, is an index code for $\mathcal{D}[\mathcal{R} \cup \mathcal{M}]$, we have $H(\boldsymbol{X}_{[m]}|\boldsymbol{C}) = 0$. Therefore, for any $\mathcal{A} \subsetneq [m]$ and any $i \in [m] \setminus \mathcal{A}$, we have $H(X_i|\boldsymbol{C}, \boldsymbol{X}_{\mathcal{A}}) \le H(\boldsymbol{X}_{[m]}|\boldsymbol{C}) = 0 < H(X_i)$. Since $\mathfrak{A} \ne \{[m]\}$, there exists some $\mathcal{A} \subsetneq [m]$. It follows that no index code can be secure. ∎

Condition C2 in Proposition 4 that every message is wanted by some receiver is implicit in classical index coding as removing messages not wanted by any receiver will change neither the index code nor the optimal index codelength. However, removing unwanted messages may affect secure index coding, because these messages can be used as keys to protect the index code against the eavesdropper. The following example illustrates this idea.

Example 1: Consider the following secure index-coding instance depicted by its directed bipartite graph representation.



The message X_2 is not wanted by any receiver. If we remove it from the setup, by invoking Proposition 4, we conclude that there is no secure index code. However, keeping X_2 in the system, by invoking Corollary 1.1, we conclude that secure index codes exist. Indeed, the index code $C = X_1 + X_2$ is secure. Here, X_2 acts as a key between the sender and receiver 1 to protect message X_1 against the eavesdropper.

VI. Random Keys for Secure Index Coding

We saw in Example 1 that using unwanted messages as keys may be essential in ensuring security. One wonders if generating random keys unknown to the receivers and the eavesdropper can also help in secure index coding. While the answer to this question is not known in general, we show that in the following three scenarios, random keys are not useful in the sense that random secure index codes exist if and only if deterministic secure index codes also exist.

A. Eavesdroppers with t-level access

From Theorem 1, it follows that using random keys does not provide greater security against an eavesdropper with t-level access, i.e., when $\mathfrak{A} = \{\mathcal{A} \subsetneq [m] : |\mathcal{A}| = t\}$, for any $t < m$.

B. Linear index codes

We now restrict the secure index codes to be linear, while $\mathfrak{A}$ is arbitrary.

Theorem 2: Given any secure index-coding instance I. Random secure *linear* index codes of codelength ℓ exist for I if and only if deterministic secure *linear* index codes of codelength ℓ also exist for I.

Proof: We only need to prove the only if direction of the claim. Any random linear index code can be expressed as $\boldsymbol{C} = \boldsymbol{X}\mathbb{G} + \boldsymbol{Y}\tilde{\mathbb{G}}$. Since each receiver recovers its intended messages, for each receiver $i \in [n]$ and each $j \in \mathcal{W}_i$, there exist an $\ell \times 1$ vector $\mathbf{D}_{i,j}$ and a $|\mathcal{K}_i| \times 1$ vector $\mathbf{E}_{i,j}$ such that

$$X_j = \boldsymbol{C}\mathbf{D}_{i,j} + \boldsymbol{X}_{\mathcal{K}_i}\mathbf{E}_{i,j}. \tag{7}$$

Let $\mathbb{V}$ be defined as the nullspace of $\tilde{\mathbb{G}}$, i.e.,

$$\mathbb{V} = \mathsf{Null}(\tilde{\mathbb{G}}) \triangleq \{\mathbf{A} \in \mathcal{F}_q^{\ell} : \tilde{\mathbb{G}}\mathbf{A} = \mathbf{0}\}. \tag{8}$$

Note that $\mathbb{V}$ is a vector space. From (7), it follows that $\mathbf{D}_{i,j} \in \mathbb{V}$ for any $i \in [n]$ and $j \in \mathcal{W}_i$, since

$$X_j - \boldsymbol{X}_{\mathcal{K}_i}\mathbf{E}_{i,j} = \boldsymbol{C}\mathbf{D}_{i,j} = \boldsymbol{X}\mathbb{G}\mathbf{D}_{i,j} + \boldsymbol{Y}\tilde{\mathbb{G}}\mathbf{D}_{i,j}, \tag{9}$$

which can hold only if $\tilde{\mathbb{G}}\mathbf{D}_{i,j} = \mathbf{0}$ for any $i \in [n]$ and $j \in \mathcal{W}_i$.

Now, let $\mathbf{A}_1, \ldots, \mathbf{A}_{\hat{\ell}}$ be a basis for $\mathbb{V}$. Note that $\hat{l} \le l$, since $\mathbb{V} \subseteq \mathcal{F}_q^{\ell}$. If the sender broadcasts $\hat{\boldsymbol{C}} \triangleq [\hat{C}_1\ \hat{C}_2\ \cdots\ \hat{C}_{\hat{\ell}}]$, where $\hat{C}_i = \boldsymbol{C}\mathbf{A}_i = \boldsymbol{X}\mathbb{G}\mathbf{A}_i$, $i \in [\hat{\ell}]$, then each receiver will still be

able to recover its intended messages, since for any $i \in [m]$ and $j \in \mathcal{W}_i$, $X_j - \boldsymbol{X}_{\mathcal{K}_i}\mathbf{E}_{i,j} = \hat{\boldsymbol{C}}\mathbf{D}_{i,j}$ is a linear combination of $\hat{C}_1, \ldots, \hat{C}_{\hat{\ell}}$. Furthermore, for any $\mathcal{A} \in \mathfrak{A}$ and any $j \in \mathcal{A}^{\mathrm{c}}$,

$$H(X_j) \geq H(X_j|\boldsymbol{X}_{\mathcal{A}}, \hat{\boldsymbol{C}}) \geq H(X_j|\boldsymbol{X}_{\mathcal{A}}, \boldsymbol{C}) = H(X_j), \quad (10)$$

where the second inequality follows since $\hat{\boldsymbol{C}}$ is a function of $\boldsymbol{C}$. Hence, the new code $\hat{\boldsymbol{C}}$ is also secure. The proof is then complete by noting that $\hat{\boldsymbol{C}}$ is a deterministic index code. ■

Remark 7: Random keys have also been shown not to be useful for linear secure index codes in the strong-security setting considered by Mojahedian et al. [11].

C. Eavesdroppers having access to only one message subset

Lastly, we consider the class of secure index-coding instances where the eavesdropper can access only one message subset.

Proposition 5: Given any index-coding instance I with $|\mathfrak{A}| = 1$, random secure index codes exist for I if and only if deterministic secure index codes also exist for I.

Proof: Note that we only need to consider index-coding instances where for each receiver $i \in [n]$, either

- (Type 1) $\mathcal{K}_i \cup \mathcal{W}_i \subseteq \mathcal{A}$;
- (Type 2) $\mathcal{K}_i \setminus \mathcal{A} \neq \emptyset$ and $\mathcal{W}_i \setminus \mathcal{A} \neq \emptyset$; or
- (Type 3) $\mathcal{K}_i \setminus \mathcal{A} \neq \emptyset$ and $\mathcal{W}_i \subseteq \mathcal{A}$.

Otherwise, according to Proposition 2, no (deterministic or random) secure index code exists. Let receivers $1, \ldots, n'$ be of Type 2, for some $0 \leq n' \leq n$, and the rest, of Type 1 or 3.

Consider a related index-coding instance $I' = ((\mathcal{K}'_i, \mathcal{W}'_i)_{i=1}^{n'}, \mathfrak{A}')$ with only $|\mathcal{A}|^c$ messages $\boldsymbol{X}_{\mathcal{A}^{\mathrm{c}}}$, n' receivers that are of Type 2 in I, where $\mathfrak{A}' = \{\emptyset\}$, $\mathcal{K}'_i = (\mathcal{K}_i \cap \mathcal{A}^{\mathrm{c}}) \neq \emptyset$, and $\mathcal{W}'_i = (\mathcal{W}_i \cap \mathcal{A}^{\mathrm{c}}) \neq \emptyset$, for all $i \in [n']$. By the definition of Type-2 receiver, $|\mathcal{K}'_i| \geq 1$ for all $i \in [n']$. For I', as $A_{\max} = 0$ and $K_{\min} \geq 1$, by invoking Corollary 1.1, we see that there exists a deterministic secure index code, say, $\boldsymbol{C}' = f'(\boldsymbol{X}_{\mathcal{A}^{\mathrm{c}}})$. This means that there exists a function $g'_i(\boldsymbol{C}', \boldsymbol{X}_{\mathcal{K}'_i}) = \boldsymbol{X}_{\mathcal{W}'_i}$ for each $i \in [n']$, and $H(X_i|\boldsymbol{C}') = H(X_i)$ for each $i \in \mathcal{A}^{\mathrm{c}}$.

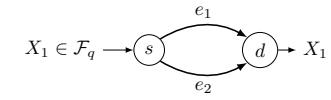
We now show that $\boldsymbol{C} = [\boldsymbol{X}_{\mathcal{A}}\ \boldsymbol{C}']$ is a secure index code for I. For any receiver of Type 1 or 3, its decoding requirement is fulfilled from observing $\boldsymbol{X}_{\mathcal{A}}$, because $\mathcal{W}_i \subseteq \mathcal{A}$. Any receiver i of Type 2 gets $\boldsymbol{X}_{\mathcal{W}_i \cap \mathcal{A}}$ from $\boldsymbol{X}_{\mathcal{A}}$, and $\boldsymbol{X}_{\mathcal{W}_i \cap \mathcal{A}^{\mathrm{c}}} = \boldsymbol{X}_{\mathcal{W}'_i}$ from $g'_i(\boldsymbol{C}', \boldsymbol{X}_{\mathcal{K}'_i})$, since it knows $\boldsymbol{X}_{\mathcal{K}_i} \supseteq \boldsymbol{X}_{\mathcal{K}'_i}$.

Finally, $H(X_i|\boldsymbol{C}, \boldsymbol{X}_{\mathcal{A}}) = H(X_i|\boldsymbol{C}', \boldsymbol{X}_{\mathcal{A}}) \overset{(a)}{=} H(X_i|\boldsymbol{C}') = H(X_i)$, for any $i \in \mathcal{A}^{\mathrm{c}}$, where (a) follows from the independence of $(X_i, \boldsymbol{C}')$ and $\boldsymbol{X}_{\mathcal{A}}$. Hence, $\boldsymbol{C}$ is a deterministic secure index code for I. So, for any I with $|\mathfrak{A}| = 1$, either no (deterministic or random) secure index codes exist, or we can always find a deterministic secure index code. ■

D. Secure index vs network coding

We now discuss some issues in extending the equivalence[2] [7, 8] between classical index and network coding to the secure setting. Consider the following network-coding instance N with a source s having two links e_1 and e_2 to a receiver d.

[2]The instances are equivalent in the sense that a code for one instance can be translated to a code for the other, and vice versa.



The codewords conveyed on links e_1 and e_2 in any network code can be written as $Y_1 = f_1(X_1)$ and $Y_2 = f_2(X_1)$, respectively, and the decoding operation $X_1 = g(Y_1, Y_2)$.

An equivalent [8] index-coding instance I has three independent messages $\hat{X}_1, \hat{Y}_1, \hat{Y}_2$, and four receivers, as follows:

Receiver	1	2	3	4
Has	$\hat{X}_1$	$\hat{X}_1$	$(\hat{Y}_1, \hat{Y}_2)$	$\hat{X}_1$
Wants	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{X}_1$	$(\hat{Y}_1, \hat{Y}_2)$

Since the codes for instances I and N can be translated to one another [8], we can translate the above code for N to an index code $\hat{\boldsymbol{C}} = [\hat{Y}_1 + f_1(\hat{X}_1)\ \ \hat{Y}_2 + f_2(\hat{X}_1)]$ for I.

Next, consider a secure version of N, with an eavesdropper who has access to any one link (e_1 or e_2) [9]. A secure network code must strongly secure X_1 against the eavesdropper. To this end, we need *random* network codes, e.g., $f_1(X_1) = K$, where K is a random key uniformly distributed on $\mathcal{F}_q$ and independent of X_1, and $f_2(X_1) = X_1 + K$.

Unfortunately, the code translation breaks down here in the presence of security constraints. In I, for receiver 1 to decode $\hat{Y}_1$ from $\hat{\boldsymbol{C}}$ and the message $\hat{X}_1$ it knows, it additionally needs to know the random key $K = f_1(\hat{X}_1)$ generated by the sender.

One difficulty in establishing an equivalence between secure network coding and secure index coding is that random keys used by the sender for encoding need not be available to the receivers for decoding. Furthermore, it is also not straightforward to translate (strong or weak) security constraints for the eavesdropper in N to equivalent and meaningful (strong or weak) security constraints in I, and vice versa.

REFERENCES

[1] 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*, 52(6), pp. 2825–2830, June 2006.

[2] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," *IEEE Trans. Inf. Theory*, 57(3), pp. 1479–1494, Mar. 2011.

[3] A. Blasiak, R. Kleinberg, and E. Lubetzky, "Broadcasting with side information: Bounding and approximating the broadcast rate," *IEEE Trans. Inf. Theory*, 59(9), pp. 292–298, Sept. 2013.

[4] H. Yu and M. J. Neely, "Duality codes and the integrality gap bound for index coding," *IEEE Trans. Inf. Theory*, 60(11), pp. 7256–7268, Nov. 2014.

[5] F. Arbabjolfaei, B. Bandemer, Y.-H. Kim, E. Şaşoğlu, and L. Wang, "On the capacity region for index coding," in *Proc. ISIT*, 2013, pp. 962–966.

[6] S. H. Dau, V. Skachek, and Y. M. Chee, "On the security of index coding with side information," *IEEE Trans. Inf. Theory*, 58(6), pp. 3975–3988, June 2012.

[7] S. El Rouayheb, A. Sprintson, and C. Georghiades, "On the index coding problem and its relation to network coding and matroid theory," *IEEE Trans. Inf. Theory*, 56(7), pp. 3187–3195, July 2010.

[8] M. Effros, S. El Rouayheb, and M. Langberg, "An equivalence between network coding and index coding," *IEEE Trans. Inf. Theory*, 61(5), pp. 2478–2487, May 2015.

[9] N. Cai and R. W. Yeung, "Secure network coding on wiretap network," *IEEE Trans. Inf. Theory*, 57(1), pp. 424–435, Jan. 2011.

[10] K. Bhattad and K. R. Narayanan, "Weakly secure network coding," in *Proc. Netcod*, 2005.

[11] M. M. Mojahedian, M. R. Aref, and A. Gohari, "Perfectly secure index coding" [Online]. Available: http://arxiv.org/abs/1504.04494v2

[12] K. Shanmugam, A. G. Dimakis, and M. Langberg, "Local graph coloring and index coding," in *Proc. ISIT*, 2013, pp. 1152–1156.

[13] M. J. Neely, A. S. Tehrani, and Z. Zhang, "Dynamic index coding for wireless broadcast networks," in *Proc. INFOCOM*, 2012, pp. 316–324.