On the Capacity Region for Secure Index Coding

Yuxin Liu*, Badri N. Vellambi†, Young-Han Kim‡, and Parastoo Sadeghi*
*Research School of Engineering, Australian National University, {yuxin.liu, parastoo.sadeghi}@anu.edu.au
†Department of Electrical Engineering and Computer Science, University of Cincinnati, badri.vellambi@uc.edu
‡Department of Electrical and Computer Engineering, University of California, San Diego, yhk@ucsd.edu

***Abstract*—We study the index coding problem in the presence of an eavesdropper, where the aim is to communicate without allowing the eavesdropper to learn any single message aside from the messages it may already know as side information. We establish an outer bound on the underlying secure capacity region of the index coding problem, which includes polymatroidal and security constraints, as well as the set of additional decoding constraints for legitimate receivers. We then propose a secure variant of the composite coding scheme, which yields an inner bound on the secure capacity region of the index coding problem. For the achievability of secure composite coding, a secret key with vanishingly small rate may be needed to ensure that each legitimate receiver who wants the same message as the eavesdropper, knows at least two more messages than the eavesdropper. For all securely feasible index coding problems with four or fewer messages, our numerical results establish the secure index coding capacity region.**

I. Introduction

Index coding is a canonical problem in network information theory with close connections to many important problems such as network coding [1] and distributed storage [2]. Index coding aims to find the optimal broadcast rate and optimal coding schemes for broadcasting n unique messages from a server to n receivers with (possibly differing) side information at each receiver [3]. Characterizing the capacity region of a general index coding problem remains elusive. This paper is concerned with a class of index coding problems where there is, in addition to n legitimate receivers, an eavesdropper who may have side information about some messages and wants to obtain the rest. We aim to characterize inner and outer bounds on the secure index coding capacity region under the restricted security requirement that there is no leakage of information about any single message that is unknown to the eavesdropper.

The secure variant of the index coding problem was first studied in [4], where the conditions for a linear code to be a valid secure index code were investigated. Later in [5], non-linear secure index codes that use side information as secret keys were proposed. The connection between secure network coding and secure index coding (analogous to the relationship between non-secure versions [1]) was developed in [6]. In [7], the authors studied the minimum key length to achieve perfect secrecy where the eavesdropper has no additional side information, but it must not learn any information whatsoever about the messages (namely, zero mutual information). The private index coding problem with linear codes was studied in [8] where the aim is to allow legitimate receivers to only learn about messages they want, but nothing of other unknown messages. Finally, [9], [10] considered the case in which the identity of the demanded message and the side information of each receiver should be kept private from other receivers.

In this paper, we examine the fundamental limits of using side information as the main protection mechanism to effect security in the index coding problem. After introducing the system model and problem setup in Section II, Section III presents a newly developed outer bound on the secure index coding capacity region. Section IV presents an achievable rate region using a secure random coding scheme for index coding. The proposed scheme is based on the existing composite coding scheme [11], [12]. For all securely feasible index coding problems with $n \leq 4$ messages, inner and outer bounds match, yielding the corresponding secure capacity regions. However, we note that for the achievability of the secure composite coding scheme, a secret key with vanishingly small rate may be needed so that each legitimate receiver who wants the same message as the eavesdropper, knows at least two more messages than the eavesdropper.

II. System Model

Throughout the paper, we let $[n] \triangleq \{1, 2, \ldots, n\}$ and use $2^{[n]}$ to denote the power set of $[n]$. The aim of the index coding problem depicted in Figure 1 is to devise a coding scheme to allow a server to communicate n independent and uniformly distributed messages, $\mathbf{X}_i \in \{0,1\}^{t_i}$, $i \in [n]$, to their corresponding receivers over a noiseless broadcast link with unit capacity in the presence of an eavesdropper e. Each receiver has prior knowledge of the realization $\mathbf{x}_{A_i}$ of $\mathbf{X}_{A_i}$, where $A_i \subseteq [n]\backslash\{i\}$. The set $B_i \triangleq [n] \setminus (A_i \cup \{i\})$ denotes the set of *interfering* messages at receiver i. The eavesdropper has access to $\mathbf{X}_{A_e}, A_e \subset [n]$. The encoder should be designed to prevent the eavesdropper from learning any single message $\mathbf{X}_j$, $j \in A_e^c = [n]\backslash A_e$.

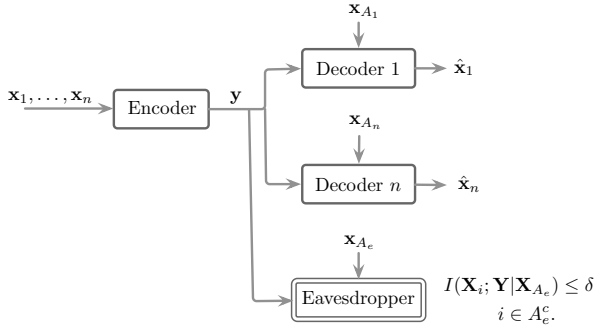

Figure 1. Problem setup for secure index coding.

To compactly represent a non-secure index coding problem we use $(i|A_i), i \in [n]$, to indicate that legitimate receiver i has

messages $\mathbf{x}_{A_i}$ and wants to decode message $\mathbf{x}_i$. With a slight abuse of notation, $(e|A_e)$ denotes what the eavesdropper has, but note that it wants to learn all other messages. A $(\mathbf{t}, r) = ((t_1, \ldots, t_n), r)$ index code is defined by:

- One encoder at the server $\phi : \prod_{i=1}^{n} \{0,1\}^{t_i} \to \{0,1\}^r$ that uses all messages to generate the transmitted codeword $\mathbf{Y} \triangleq \phi(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ of length r bits.
- For each legitimate receiver $i \in [n]$, a decoder function $\psi_i : \{0,1\}^r \times \prod_{k \in A_i} \{0,1\}^{t_k} \to \{0,1\}^{t_i}$ that takes the received sequence $\mathbf{Y}$ together with the side information at receiver i and maps them to $\hat{\mathbf{X}}_i \triangleq \psi_i(\mathbf{Y}, \mathbf{X}_{A_i})$.

A rate tuple $(R_1, \ldots, R_n)$ is said to be *securely* achievable if for every $\delta, \epsilon > 0$, there exists a $(\mathbf{t}, r)$ index code such that the following three constraints are met:

$$\text{Rate:} \quad R_i \leq \tfrac{t_i}{r}, \quad i \in [n]; \tag{1}$$

$$\text{Decoding:} \quad \mathsf{P}\{(\hat{\mathbf{X}}_1, \ldots, \hat{\mathbf{X}}_n) \neq (\mathbf{X}_1, \ldots, \mathbf{X}_n)\} \leq \epsilon; \tag{2}$$

$$\text{Security:} \quad I(\mathbf{X}_i; \mathbf{Y} | \mathbf{X}_{A_e}) \leq \delta, \quad i \in A_e^c. \tag{3}$$

We define the secure capacity region $\mathscr{C}^{\mathrm{s}}$ as the closure of the set of all securely achievable rate tuples. Note that for a sequence of codes operating at a securely achievable rate tuple, the decoding condition along with Fano's inequality ensure that at receiver i, $\lim_{\epsilon \to 0} \frac{1}{r} H(\mathbf{X}_i | \mathbf{Y}, \mathbf{X}_{A_i}) = 0$.

Note that an index coding problem is not securely feasible when there is there is no securely achievable rate tuple. This happens when $A_i \subseteq A_e$ for some $i \in A_e^c$. That is, when the side information of the eavesdropper is equally strong or stronger than that of some receiver. Otherwise, the secure index coding problem is said to be *securely feasible*.

III. Polymatroidal Outer Bound

Theorem 1 (Secure Outer Bound): Any securely achievable rate tuple for the index coding problem $(i|A_i)$, $i \in [n]$, and $(e|A_e)$ must lie in $\mathscr{R}_{\mathrm{g}}^{\mathrm{s}}$ that consists of all rate tuples satisfying

$$R_i = g(B_i \cup \{i\}) - g(B_i), \quad i \in [n], \tag{4}$$

for some set function $g : 2^{[n]} \to [0,1]$ such that for any $J \subseteq [n]$ and $i, k \notin J$,

$$g(\emptyset) = 0, \tag{5}$$

$$g([n]) \leq 1, \tag{6}$$

$$g(J) \leq g(J \cup \{i\}), \tag{7}$$

$$g(J) + g(J \cup \{i,k\}) \leq g(J \cup \{i\}) + g(J \cup \{k\}), \tag{8}$$

$$g(B_i \cup \{i\}) - g(B_i) = g(\{i\}), \tag{9}$$

and additionally for $i \in A_e^c$,

$$g(A_e^c \setminus \{i\}) \geq g(A_e^c). \tag{10}$$

The proof is given in in the extended version of this work [13]. First, we note that the rate constraint (4) and polymatroidal constraints (5)-(8) appeared in [11] to form an outer bound on the non-secure index coding capacity region[1], which is known to be tight for all index coding problems with $n \leq 5$ messages. Constraint (9) captures *additional decoding conditions* for each legitimate receiver $i \in [n]$, since

$$\lim_{\epsilon \to 0} \frac{1}{r} H(\mathbf{X}_i | \mathbf{Y}, \mathbf{X}_{A_i}) = \lim_{\epsilon \to 0} \frac{1}{r} H(\mathbf{X}_i | \mathbf{Y}, \mathbf{X}_{A_i}, \mathbf{X}_C) = 0,$$

for any $C \subseteq B_i$. Due to the submodularity constraint (8), it suffices to write the additional decoding condition for $C = B_i$ alone. See [13] for more details. We note that the same constraint appeared in a similar outer bound to the non-secure index coding capacity region in [14]. Finally, (10) captures the security constraint (3) with $\delta = 0$.

An *explicit* outer bound to the index coding capacity region is derived from Theorem 1 by the means of Fourier-Motzkin elimination (FME) [15] by eliminating $g(J), J \subseteq [n]$.

Example 1: Consider the *non-secure* index coding problem $(1|-), (2|3), (3|2)$ in the absence of the eavesdropper. Invoking Theorem 1 without (10) and eliminating variables $g(J), J \subseteq [n]$, via FME yields

$$R_1 + R_2 \leq 1, \quad R_1 + R_3 \leq 1.$$

Example 2: Consider the index coding problem $(1|-), (2|3), (3|2)$, with the eavesdropper $(e|1)$. The outer bound given by Theorem 1 is

$$R_2 = R_3, \quad R_1 + R_3 \leq 1.$$

In Example 2, the security requirement imposes the equality $R_2 = R_3$. Since the eavesdropper knows $\mathbf{x}_1$, it is not possible to protect $\mathbf{x}_2$ or $\mathbf{x}_3$ using $\mathbf{x}_1$. The only solution to guarantee security is to protect $\mathbf{x}_2$ with $\mathbf{x}_3$ and vice versa at the same rate. This is achieved using a simple linear code, $\mathbf{x}_1$ (of length t_1), $\mathbf{x}_2 \oplus \mathbf{x}_3$ (of length t_2), illustrated below.

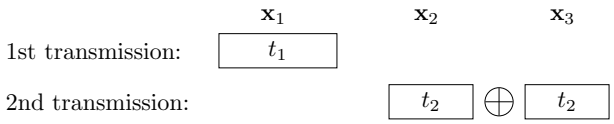

Figure 2. A linear code achieves the capacity region for Example 2.

We summarize a few important observations.

1) The outer bound on the non-secure capacity region of the index coding problem is at least as large as that on the secure capacity region.
2) In some secure index coding problems, it is possible for a stronger receiver (with more side information) to have its rate bounded by that of a weaker receiver (with less side information). For the problem $(1|3), (2|1,3), (3|1), (e|-)$, the outer bound on the secure capacity region stipulates $R_2 \leq R_3$, while $A_3 = \{3\} \subset \{1,3\} = A_2$.
3) The additional decoding constraint (9) in Theorem 1 is essential for deriving a tighter outer bound for some index coding problems. For example, if we exclude (9), the outer bound for the problem $(1|3), (2|3), (3|2), (e|-)$ is $R_1 + R_2 \leq 1, R_2 = R_3$. However, with (9) included, the outer bound is

$$R_1 + R_2 \leq 1, \quad R_2 = R_3, \quad R_1 \leq R_3.$$

[1] In [11], inequalities in (4) and equality in (6) were used. This is immaterial to the non-secure capacity region outer bound. See Remark 4 at the end of this section for its impact on the secure counterpart.

4) One can derive another outer bound, similar to Theorem 1, where for $i \in [n]$ the equality in (4) is replaced by inequality relation $\leq$ and the inequality relation $\leq$ in (6) is replaced by equality. However, doing so does not yield the condition $R_2 = R_3$ in Example 2.

IV. Secure Composite Coding Inner Bound

Before proposing the secure composite coding scheme, we first recap the original composite coding scheme established in [11]. For ease of exposition, the scheme is described for a fixed decoding configuration, which is a tuple of subsets of messages $\mathbf{D} = (D_i, i \in [n])$ such that for each $i \in [n]$, $D_i \subseteq [n] \backslash A_i$ and $i \in D_i$. Let $r \in \mathbb{N}$. Let for each $i \in [n]$, $t_i = \lceil rR_i \rceil$. Denote $s_K = \lceil rS_K \rceil$, where $S_K \in [0,1]$ is the rate of composite index for subset K. By convention, $S_\emptyset = 0$.

Codebook generation: (1) For each $K \subseteq [n]$ and $\mathbf{x}_K$, a corresponding composite index $W_K(\mathbf{x}_K)$ is drawn uniformly at random from $[2^{s_K}]$. **(2)** For every tuple $(w_K, K \in 2^{[n]})$, the codeword to be transmitted, $\mathbf{Y}((w_K, K \in 2^{[n]}))$, is drawn uniformly at random from $[2^r]$. The random codebooks (message-to-composite indices and composite indices-to-codeword maps) are revealed to all parties.

Encoding: To communicate a realization $\mathbf{x}_{[n]}$, the transmitter sends $\mathbf{Y}((W_K(\mathbf{x}_K), K \in 2^{[n]}))$.

Decoding: Upon receiving the realization $\mathbf{y}$: **(1)** Each legitimate receiver i finds the unique tuple of composite indices $(\hat{w}_K, K \in 2^{[n]})$ such that $\mathbf{y} = \mathbf{Y}((\hat{w}_K, K \in 2^{[n]}))$, and declares an error if a unique tuple is not found. **(2)** Assuming composite index tuple decoding is successful, receiver i finds the unique message tuple $\hat{\mathbf{x}}$ such that $w_K = W_K(\hat{\mathbf{x}}_K)$, $K \subseteq D_i \cup A_i$. An error occurs if a unique tuple is not found.

The following result from [11] quantifies the constraints on the message rates and composite index rates for successful decoding (in the non-secure setting).

Proposition 1: A rate tuple $(R_i, i \in [n])$ is achievable for the index coding problem $(i|A_i)$, $i \in [n]$, if for each $i \in [n]$:

$$\sum_{J \not\subseteq A_i} S_J < 1, \tag{11}$$

$$\sum_{i \in K} R_i < \sum_{J \subseteq D_i \cup A_i, J \cap K \neq \emptyset} S_J, \quad K \subseteq D_i. \tag{12}$$

Now we move on to develop the *secure* composite coding scheme. Recall the security condition

$$I(\mathbf{X}_i; \mathbf{Y} | \mathbf{X}_{A_e}) < \delta, \quad i \in A_e^c. \tag{13}$$

Using the chain rule for mutual information and the independence between different messages we have

$$I(\mathbf{X}_i; \mathbf{Y}, \mathbf{X}_{A_e}) < \delta, \quad i \in A_e^c. \tag{14}$$

Since the eavesdropper can generate all composite indices $\{\mathbf{w}_J : J \subseteq A_e\}$ from $\mathbf{X}_{A_e}$, it will be useful to define $T = \{K : K \subseteq [n], K \not\subseteq A_e\}$. Then for any $Q \subseteq T$, $P_Q = \bigcup_{J \in Q} J \backslash A_e$ is the set of messages from Q that are unknown to the eavesdropper. We assume that the eavesdropper learns the codebook and is also able to decode all the composite indices in the first step of decoding. Condition (14) becomes:

$$I(\mathbf{X}_i; \{W_K : K \in T\}, \mathbf{X}_{A_e}) < \delta, \quad i \in A_e^c. \tag{15}$$

Applying Theorem 1 from [16] and Lemma 2.7 from [17], we obtain the following random-coding based achievable rate region. The proof is in Appendix A for the more general secure enhanced composite coding scheme in Proposition 2.

Theorem 2: A rate tuple $(R_i, i \in [n])$ is securely achievable for the index coding problem $(i|A_i), i \in [n], (e|A_e)$ if

$$\sum_{J \not\subseteq A_i} S_J < 1, \quad i \in [n], \tag{16}$$

$$\sum_{i \in K} R_i < \sum_{J \subseteq D_i \cup A_i, J \cap K \neq \emptyset} S_J, \quad K \subseteq D_i, i \in [n], \tag{17}$$

$$\sum_{\substack{K \subseteq P_Q \cup A_e \\ K \not\subseteq A_e}} S_K < \sum_{j \in (P_Q \backslash \{i\})} R_j, \quad Q \subseteq T, i \in A_e^c. \tag{18}$$

When Theorem 2 gives an inequality of the form $S_J < 0$, we set $S_J = 0$. For each index coding problem with $n \leq 5$ messages, a single *natural* decoding configuration $\underline{\mathbf{D}}$ [18] was shown to be sufficient to achieve the non-secure capacity region. We will also use the natural decoding configuration in this paper, which will be sufficient to achieve the secure capacity region for all index coding problems with $n \leq 4$ messages. However, more than one decoding configuration might be necessary for larger problems. Secure composite coding with multiple decoding configurations is given below.

1) Secure Enhanced Composite Coding Scheme: Following similar lines as [12], let Δ be the set of all decoding configurations, i.e, $\Delta = \{\mathbf{D} : D_i \subseteq [n] \backslash A_i, i \in D_i\}$.

Let $r \in \mathbb{N}$. Let for each $\mathbf{D} \in \Delta$ and $i \in [n]$, $t_i(\mathbf{D}) = \lceil rR_i(\mathbf{D}) \rceil$, where $R_i(\mathbf{D})$ is the rate of message i communicated via decoding configuration $\mathbf{D}$. Let $\mathbf{X}_i(\mathbf{D}) \in [2^{t_i(\mathbf{D})}]$ be the part of message i communicated via decoding configuration $\mathbf{D}$. For each $K \subseteq [n]$ and $\mathbf{D} \in \Delta$, let $S_K(\mathbf{D}) \in [0,1]$. Denote $s_K(\mathbf{D}) = \lceil rS_K(\mathbf{D}) \rceil$, $K \subseteq [n]$, where $S_K(\mathbf{D})$ is the rate of composite index for subset K and decoding configuration $\mathbf{D}$. By convention, $S_\emptyset(\mathbf{D}) = 0$ for each $\mathbf{D} \in \Delta$.

Codebook generation: (1) For each $K \subseteq [n]$, $\mathbf{D} \in \Delta$, and $\mathbf{x}_K(\mathbf{D})$, a corresponding composite index $W_{K,\mathbf{D}}(\mathbf{x}_K(\mathbf{D}))$ is drawn uniformly at random from $[2^{s_K(\mathbf{D})}]$. **(2)** For every tuple $(w_{K,\mathbf{D}}, (K, \mathbf{D}) \in 2^{[n]} \times \Delta)$, the codeword to be transmitted, $\mathbf{Y}((w_{K,\mathbf{D}}, (K, \mathbf{D}) \in 2^{[n]} \times \Delta))$, is drawn uniformly at random from $[2^r]$. The codebooks are revealed to all parties.

Encoding: To communicate a realization $\mathbf{x}_{[n]}$, the transmitter sends $\mathbf{Y}((W_{K,\mathbf{D}}(\mathbf{x}_K(\mathbf{D})), (K, \mathbf{D}) \in 2^{[n]} \times \Delta))$.

Decoding: Upon receiving the realization $\mathbf{y}$: **(1)** Each legitimate receiver i finds the unique tuple of composite indices $(\hat{w}_{K,\mathbf{D}}, (K, \mathbf{D}) \in 2^{[n]} \times \Delta)$ such that $\mathbf{y} = \mathbf{Y}((\hat{w}_{K,\mathbf{D}}, (K, \mathbf{D}) \in 2^{[n]} \times \Delta))$, and declares an error if a unique tuple is not found. **(2)** Assuming step **(1)** is successful, for each $\mathbf{D} \in \Delta$, receiver i finds the unique message tuple $\hat{\mathbf{x}}_{D_i}(\mathbf{D})$ such that $w_{K,\mathbf{D}} = W_{K,\mathbf{D}}(\hat{\mathbf{x}}_K)$, for all $K \subseteq D_i \cup A_i$. An error occurs if a unique tuple is not found.

Proposition 2: A rate tuple $(R_i, i \in [n])$ is securely achievable for the index coding problem $(i|A_i), i \in [n], (e|A_e)$ if

$$R_i = \sum_{\mathbf{D} \in \Delta} R_i(\mathbf{D}), \quad i \in [n], \tag{19}$$

$$\sum_{\mathbf{D}\in\Delta}\sum_{J\not\subseteq A_i} S_J(\mathbf{D}) < 1, \quad i\in[n], \tag{20}$$

$$\sum_{i\in K} R_i(\mathbf{D}) < \sum_{\substack{J\subseteq D_i\cup A_i\\ J\cap K\neq\emptyset}} S_J(\mathbf{D}), \quad \mathbf{D}\in\Delta, K\subseteq D_i, i\in[n], \tag{21}$$

$$\sum_{\substack{K\subseteq P_Q\cup A_e\\ K\not\subseteq A_e}} S_K(\mathbf{D}) < \sum_{j\in(P_Q\setminus\{i\})} R_j(\mathbf{D}),\ \mathbf{D}\in\Delta, Q\subseteq T, i\in A_e^c. \tag{22}$$

Theorem 2 is recovered from Proposition 2 by setting $S_K(\mathbf{D})$, $K\in 2^{[n]}$ and $R_j(\mathbf{D})$, $j\in[n]$ to zero for all, but one $\mathbf{D}$.

A. Secure Composite Coding with a Secret Key

Theorem 2 and Proposition 2 may generate conflicting constraints for some index coding problems as shown below.

Example 3: Consider the same setting as in Example 1. Set $D_1=\{1\}$, $D_2=\{1,2\}$, and $D_3=\{1,3\}$. The set of active inequalities generated by Theorem 2 is

$$\begin{aligned}
&S_1+S_2+S_{12}+S_3+S_{13}+S_{23}+S_{123}<1,\\
&R_1<S_1,\\
&R_2<S_2+S_{12}+S_{23}+S_{123},\\
&R_3<S_3+S_{13}+S_{23}+S_{123},\\
&R_1+R_2<S_1+S_2+S_{12}+S_{13}+S_{23}+S_{123},\\
&R_1+R_3<S_1+S_{12}+S_3+S_{13}+S_{23}+S_{123},\\
&S_2+S_{12}+S_{13}+S_{23}+S_{123}<R_2,\\
&S_{12}+S_3+S_{13}+S_{23}+S_{123}<R_3,\\
&S_2=0,\ S_3=0,\ S_{12}=0,\ S_{13}=0.
\end{aligned}$$

Clearly, there are conflicting constraints for R_2 and R_3.

For $n=3$ messages, there are the total of 20 securely feasible index coding problems. Of these, only the problem $(1|2,3),(2|1,3),(3|1,2),(e|-)$ does not have conflicting inequalities. For $n=4$ messages, 43 out of 833 securely feasible index coding problems have no conflicting inequalities. For all such non-conflicting cases, the secure composite coding inner bound matches the secure polymatroidal outer bound, thereby establishing the corresponding secure capacity region. In each of these problems, each receiver who wants the same message as the eavesdropper, knows at least two more messages as side information than the eavesdropper. I.e., $\forall\, i\in A_e^c, |A_i\backslash A_e|\geq 2$. We show a resolution can be obtained for the conflicting cases by using a secret key of arbitrarily small rate shared between the server and legitimate receivers.

Assume there is an independent secret key $\mathbf{M}$ at rate ζ_M, shared between the transmitter and all the legitimate receivers. For each $K\subseteq[n]$, $\mathbf{x}_K(\mathbf{D})\cup\mathbf{M}$ is mapped into a composite index $W_{K,M,\mathbf{D}},(\mathbf{x}_K(\mathbf{D})\cup\mathbf{M})$ drawn uniformly at random from $[2^{s_{K,M}(\mathbf{D})}]$. The second step of codebook generation, encoding and decoding are the same as before.

Theorem 3: A rate tuple $(R_i, i\in[n])$ is securely achievable for the index coding problem $(i|A_i), i\in[n], (e|A_e)$ if

$$R_i=\sum_{\mathbf{D}\in\Delta} R_i(\mathbf{D}), \quad i\in[n], \tag{23}$$

$$\sum_{\mathbf{D}\in\Delta}\sum_{J\not\subseteq A_i} S_{J,M}(\mathbf{D}) < 1, \quad i\in[n], \tag{24}$$

$$\sum_{i\in K} R_i(\mathbf{D}) < \sum_{J\subseteq D_i\cup A_i, J\cap K\neq\emptyset} S_{J,M}(\mathbf{D}),\ \mathbf{D}\in\Delta, K\subseteq D_i, i\in[n], \tag{25}$$

$$\sum_{\substack{K\subseteq P_Q\cup A_e\\ K\not\subseteq A_e}} S_{K,M}(\mathbf{D}) < \sum_{j\in(P_Q\setminus\{i\})} R_j(\mathbf{D})+\zeta_M, \tag{26}$$

$$\mathbf{D}\in\Delta, Q\subseteq T, i\in A_e^c.$$

For the secure index coding problem described in Example 3, the secure achievable rate region with a secret key becomes

$$\begin{aligned}
&R_3-\zeta_M<R_2<R_3+\zeta_M,\\
&R_1+R_2<1, \quad R_1+R_3<1.
\end{aligned}$$

which matches the polymatroidal outer bound as $\zeta_M\to 0$.

We now summarize our key observations.

1) When $\forall\, i\in A_e^c, |A_i\backslash A_e|\geq 2$, the secure composite coding of Theorem 2 directly achieves the capacity region. For $n=3$, there is 1 such problem out of 20 securely feasible problems. For $n=4$, there are 43 such problems out of 833 securely feasible problems.
2) For the remaining cases, conflicting inequalities can be resolved using a secret key of vanishingly small rate. The secret key acts as the second message, unknown to the eavesdropper, to ensure $\forall\, i\in A_e^c, |A_i\backslash A_e|\geq 2$.

Table I lists the secure capacity region for 20 securely feasible index coding problems with $n=3$ messages.

REFERENCES

[1] M. Effros, S. El Rouayheb, and M. Langberg, "An equivalence between network coding and index coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2478–2487, May 2015.

[2] K. Shanmugam and A. G. Dimakis, "Bounding multiple unicasts through index coding and locally repairable codes," in *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, Jul. 2014, pp. 296–300.

[3] Y. Birk and T. Kol, "Informed-source coding-on-demand (ISCOD) over broadcast channels," in *IEEE INFOCOM*, Mar. 1998, pp. 1257–1264.

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

[5] L. Ong, B. N. Vellambi, P. L. Yeoh, J. Kliewer, and J. Yuan, "Secure index coding: Existence and construction," in *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, July 2016, pp. 2834–2838.

[6] L. Ong, J. Kliewer, and B. N. Vellambi, "Secure network-index code equivalence: Extension to non-zero error and leakage," in *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, June 2018, pp. 841–845.

[7] M. M. Mojahedian, M. R. Aref, and A. Gohari, "Perfectly secure index coding," *IEEE Trans. Inf. Theory*, vol. 63, pp. 7382–7395, Nov 2017.

[8] V. Narayanan, J. Ravi, V. K. Mishra, B. K. Dey, N. Karamchandani, and V. M. Prabhakaran, "Private index coding," in *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, 2018, pp. 596–600.

[9] M. Karmoose, L. Song, M. Cardone, and C. Fragouli, "Private broadcasting: An index coding approach," in *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, June 2017, pp. 2543–2547.

[10] ——, "Privacy in index coding: Improved bounds and coding schemes," in *IEEE Int. Symp. on Information Theory (ISIT)*, 2018, pp. 831–835.

[11] F. Arbabjolfaei, B. Bandemer, Y.-H. Kim, E. Şaşoğlu, and L. Wang, "On the capacity region for index coding," in *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, 2013, pp. 962–966.

[12] Y. Liu, P. Sadeghi, F. Arbabjolfaei, and Y.-H. Kim, "On the capacity for distributed index coding," in *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, 2017, pp. 3055–3059.

[13] L. Yuxin, B. N. Vellambi, Y.-H. Kim, and P. Sadeghi. On the capacity region for secure index coding. [Online]. Available: http://arxiv.org/abs/1809.03615

[14] A. Blasiak, R. Kleinberg, and E. Lubetzky, "Lexicographic products and the power of non-linear network coding," in *IEEE Annual Symp. on Foundations of Computer Science*, Oct 2011, pp. 609–618.

[15] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge: Cambridge University Press, 2011.

[16] M. H. Yassaee, M. R. Aref, and A. Gohari, "Achievability proof via output statistics of random binning," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6760–6786, Nov 2014.

[17] I. Csiszar and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[18] Y. Liu, P. Sadeghi, F. Arbabjolfaei, and Y.-H. Kim, "Simplified composite coding for index coding," in *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, 2018, pp. 456–460.

Table I
SECURE CAPACITY REGION OUTER BOUNDS FOR ALL SECURELY FEASIBLE INDEX CODING PROBLEMS WITH $n = 3$ MESSAGES

Receiver and Eavesdropper Information	Outer Bounds	Receiver and Eavesdropper Information	Outer Bounds
$(1\|-)$ $(2\|3)$ $(3\|2),(e\|1)$	$R_2 = R_3$ $R_1 + R_3 \le 1$	$(1\|2,3)$ $(2\|1)$ $(3\|-),(e\|3)$	$R_2 + R_3 \le 1$ $R_1 = R_2$
$(1\|3)$ $(2\|3)$ $(3\|2),(e\|-)$	$R_1 + R_2 \le 1$ $R_2 = R_3$ $R_1 \le R_3$	$(1\|3)$ $(2\|3)$ $(3\|2),(e\|1)$	$R_1 + R_2 \le 1$ $R_2 = R_3$
$(1\|3)$ $(2\|1)$ $(3\|2),(e\|-)$	$R_1 + R_2 \le 1$ $R_1 = R_2 = R_3$	$(1\|2,3)$ $(2\|1,3)$ $(3\|-),(e\|3)$	$R_1 = R_2$ $R_2 + R_3 \le 1$
$(1\|3)$ $(2\|3)$ $(3\|1,2),(e\|-)$	$R_1 + R_2 \le 1$ $R_1 \le R_3$ $R_2 \le R_3$ $R_3 \le R_1 + R_2$	$(1\|3)$ $(2\|3)$ $(3\|1,2),(e\|1)$	$R_2 = R_3$ $R_1 + R_3 \le 1$
$(1\|3)$ $(2\|3)$ $(3\|1,2),(e\|2)$	$R_1 = R_3$ $R_2 + R_3 \le 1$	$(1\|3)$ $(2\|1)$ $(3\|1,2),(e\|-)$	$R_1 = R_3$ $R_2 + R_3 \le 1$ $R_2 \le R_1$
$(1\|3)$ $(2\|1)$ $(3\|1,2),(e\|2)$	$R_1 = R_3$ $R_2 + R_3 \le 1$	$(1\|3)$ $(2\|1,3)$ $(3\|1),(e\|-)$	$R_1 + R_2 \le 1$ $R_1 = R_3$ $R_2 \le R_3$
$(1\|3)$ $(2\|1,3)$ $(3\|1),(e\|2)$	$R_1 + R_2 \le 1$ $R_1 = R_3$	$(1\|2,3)$ $(2\|1,3)$ $(3\|1),(e\|-)$	$R_3 \le R_1$ $R_2 \le R_1$ $R_2 + R_3 \le 1$ $R_1 \le R_2 + R_3$
$(1\|2,3)$ $(2\|1,3)$ $(3\|1),(e\|2)$	$R_1 = R_3$ $R_2 + R_3 \le 1$	$(1\|2,3)$ $(2\|1,3)$ $(3\|1),(e\|3)$	$R_1 = R_2$ $R_2 + R_3 \le 1$
$(1\|2,3)$ $(2\|1,3)$ $(3\|1,2),(e\|-)$	$R_1 \le 1$ $R_2 \le 1$ $R_3 \le 1$ $R_1 \le R_2 + R_3$ $R_2 \le R_1 + R_3$ $R_3 \le R_1 + R_2$	$(1\|2,3)$ $(2\|1,3)$ $(3\|1,2),(e\|1)$	$R_1 \le 1$ $R_2 \le 1$ $R_2 = R_3$
$(1\|2,3)$ $(2\|1,3)$ $(3\|1,2),(e\|2)$	$R_2 \le 1$ $R_3 \le 1$ $R_1 = R_3$	$(1\|2,3)$ $(2\|1,3)$ $(3\|1,2),(e\|3)$	$R_1 \le 1$ $R_3 \le 1$ $R_1 = R_2$

APPENDIX A
PROOF OF PROPOSITION 2

Observe that the first three rate constraints of Proposition 2 follow from the achievability proof of [12]. We are done if we show that (22) implies for each $\mathbf{D} \in \Delta$,

$$I(\mathbf{X}_i(\mathbf{D}); (W_{K,\mathbf{D}} : K \in T)|\mathbf{X}_{A_e}(\mathbf{D})) = o(1) \qquad (27)$$

as $r \to \infty$. Since different parts of messages are independently encoded, the above would also ensure that

$$I(\mathbf{X}_i; (W_{K,\mathbf{D}} : K \in T, \mathbf{D} \in \Delta)|\mathbf{X}_{A_e}) = o(1) \text{ as } r \to \infty.$$

To show that (27) holds under (22), we pick $\mathbf{D} \in \Delta$ and focus on only the part of message conveyed via this choice of decoding configuration. For the remainder of proof, we drop the reference to $\mathbf{D}$ remembering that $\mathbf{X}_j$, R_j, S_K stand for $\mathbf{X}_j(\mathbf{D})$, $R_j(\mathbf{D})$, $S_K(\mathbf{D})$, respectively.

Recall that $\mathbf{X}_j \in [2^{rR_j}]$, $j \in [n]$, are independent and uniformly distributed. Let $Z_i = \{i\} \cup A_e$ for some $i \in A_e^c$. Set $T = \{K : K \subseteq [n], K \not\subseteq A_e\}$ and $P_Q = \bigcup_{J\in Q} J \backslash A_e$, for any $Q \subseteq T$. Define $R_{Z_i} = R_i + \sum_{j\in A_e} R_j$ and $R_K = \sum_{j\in K} R_j$, $K \in T$. We have $\mathbf{X}_{Z_i} = (\mathbf{X}_i, \mathbf{X}_{A_e}) \in [2^{rR_i}] \times \prod_{j\in A_e}[2^{rR_j}] = [2^{rR_{Z_i}}]$ and for each $K \in T$, $\mathbf{X}_K = (x_i, i \in K) \in \prod_{j\in K}[2^{rR_j}] = [2^{rR_K}]$. Then $((\mathbf{X}_K, K \in T), \mathbf{X}_{Z_i})$ is a well-defined discrete memoryless correlated source.

For each $K \in T$, the random mapping $b_K : [2^{rR_K}] \to [2^{rS_K}]$ uniformly and independently maps each sequence $\mathbf{x}_K$ to $w_K \in [2^{rS_K}]$. Denote by B_K and W_K the random encoding and the random bin index corresponding to b_K and w_K, respectively. Let $w_T = (w_K, K \in T)$, $W_T = (W_K, K \in T)$, $b_T = (b_K, K \in T)$ and $B_T = (B_K, K \in T)$. For $Q \subseteq T$

$$\sum_{J\in Q} S_J = \sum_{\substack{K\subseteq P_Q \cup A_e \\ K\not\subseteq A_e}} S_K, \qquad (28)$$

and for each $Q \subseteq T$ and Z_i,

$$H(\mathbf{X}_{\bigcup_{J\in Q}}|\mathbf{X}_{Z_i}) = H(\mathbf{X}_{\bigcup_{J\in Q}}|\mathbf{X}_i \cup \mathbf{X}_{A_e}) = \sum_{j\in(P_Q\backslash\{i\})} R_j.$$

Using our notation, Theorem 1 in [16] can be restated as:

Theorem 4: If for each $Q \subseteq T$,

$$\sum_{\substack{K\subseteq P_Q \cup A_e \\ K\not\subseteq A_e}} S_K < \sum_{j\in(P_Q\backslash\{i\})} R_j. \qquad (29)$$

then

$$\mathbb{E}_{B_T} \left\| p_{\mathbf{X}_{Z_i}W_T} - p_{\mathbf{X}_{Z_i}} \prod_{K\in T} p^U_{[2^{rS_K}]} \right\|_1 < o(2^{-\beta r}), \qquad (30)$$

where: (1) The outer expectation is only over the choice of random binning; (2) $p_{\mathbf{X}_{Z_i}W_T}$ is the joint pmf induced by a particular random binning; (3) $p^U_{\mathscr{A}}$ is the uniform distribution over set $\mathscr{A}$, and (4) $\beta > 0$ is the rate of convergence. Since (30) exponentially converges to 0 as r goes to infinity [16], there exists a sequence of binning schemes $\{b^*_{T,r} : r \in \mathbb{N}\}$ such that the sequence of joint pmfs $\{p^{*(r)}_{\mathbf{X}_{Z_i}W_T} : r \in \mathbb{N}\}$ induced by the sequence of binning schemes satisfies

$$\left\| p^{*(r)}_{\mathbf{X}_{Z_i}W_T} - p_{\mathbf{X}_{Z_i}} \prod_{K\in T} p^U_{[2^{rS_K}]} \right\|_1 < o(2^{-\beta r}), \qquad (31)$$

Now, we use Lemma 2.7 from [17]:

Lemma 1: If p_1 and p_2 are two distributions over a finite set $\mathbf{X}$ such that $\sum_{\mathbf{x}\in\mathbf{X}} \|p_1(\mathbf{x}) - p_2(\mathbf{x})\|_1 \le \theta \le \frac{1}{2}$, then

$$\|H(p_1) - H(p_2)\|_1 \le -\theta \log \frac{\theta}{|\mathbf{X}|}. \qquad (32)$$

Recall $|\mathbf{X}| = r(\sum_{j\in\{i\}\cup A_e} R_j + \sum_{K\in T} S_K) = \alpha r$. Thus, for the sequence of binning schemes $\{b^*_{T,r} : r \in \mathbb{N}\}$:

$$I(\mathbf{X}_i \cup \mathbf{X}_{A_e}; (W_K, K \in T)) < o(\alpha r 2^{-\beta r})$$
$$\Rightarrow \quad I(\mathbf{X}_i; (W_K, K \in T)|\mathbf{X}_{A_e}) < o(\alpha r 2^{-\beta r}).$$

Therefore, the security condition (15) holds as $r \to \infty$.