Narayanan, V., Ravi, J., Mishra, V.K., Dey, B.K., Karamchandani, N. y Prabhakaran, V. M. (2018). Private Index Coding. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 596-600.

DOI: https://doi.org/10.1109/ISIT.2018.8437353

Private Index Coding

Varun Narayanan*, Jithin Ravi†, Vivek K. Mishra‡, Bikash K. Dey§, Nikhil Karamchandani§, Vinod M. Prabhakaran*
* Tata Institute of Fundamental Research, Mumbai. Email: varun.narayanan@tifr.res.in, vinodmp@tifr.res.in
† Universidad Carlos III de Madrid, Leganés, Spain. Email: rjithin@gmail.com
‡ Qualcomm, India. Email: vivemish@qti.qualcomm.com
§ Indian Institute of Technology Bombay, Mumbai. Email: {bikash, nikhil}@ee.iitb.ac.in

***Abstract*—We study the problem of index coding under the privacy requirement that receivers do not learn anything more than the messages they already have as side information and the message they want from the server. To achieve this *private* index coding, we consider the use of secret keys that are shared among various subsets of users and the server. We characterize key access structures that allow private index coding. For up to three receivers, we characterize the rate region of transmission and key rates and show that scalar coding is optimal; we also show that scalar linear codes are sub-optimal for four receivers. Furthermore, when no keys are available, we consider a weaker notion of privacy analogous to weak security. Finally, for a different setting in which the server is allowed to send messages exclusively to a subset of users, we study the number of transmissions required to achieve error-free decoding and privacy.**

I. Introduction

Index coding [1] is a fundamental problem in network information theory where a server with N messages X_i, $i \in [N] := \{1, \dots, N\}$ communicates over a noiseless broadcast link to N users. User $i \in [N]$ has a subset of the N messages as side information and wants to obtain X_i. In this work, we consider the additional privacy constraint that each user i learns no additional information about the N messages other than the side information it already has and the message X_i which is intended for it. To achieve this, the encoding and decoding schemes use a collection of keys that are shared by subsets of users and are known to the server. We call this the *private index coding* problem.

Some recent works have studied security aspects of index coding [2]–[6]. Most of these consider security against an eavesdropper who tries to learn some information about the messages by wiretapping the broadcast link from the server to the users. Security against an eavesdropper who has access to a subset of messages was first studied in [2]. The authors obtained the conditions that any linear code should satisfy to achieve decodability as well as secrecy. Security against an eavesdropper without any side information was studied in [3]. Here secrecy is achieved using a key shared between the server and users. Weakly-secure index coding against an eavesdropper with some side information was studied in [4]. Eavesdropper with side information was also studied in [5], where an equivalence between secure index coding and secure network coding was shown. In [6], a different aspect of privacy was studied, where each user wants to hide the identities of its side information messages and requested messages from other users.

In this paper, we first characterize the *key access structures* (*i.e.,* the collections of subsets of users which possess exclusive keys) that make private index coding feasible. We also give conditions under which a linear scheme is a valid private index code. Next, we study the rate region of private index coding which gives the fundamental trade-off between the rate of transmission and rates of keys. We characterize the rate region when the number of users is at most 3 and show that all feasible rates may be achieved using scalar linear coding and time sharing. Further, we give an instance of the problem with 4 users where all the points cannot be achieved using scalar linear coding. In contrast, for index coding, it was shown that scalar linear coding is optimal up to 4 users [7].

We also consider privacy when no keys are shared among the users and the server. It turns out that, under all but trivial cases, the above privacy requirement cannot be met. Hence, we study a weaker notion of privacy where each user must not learn any information about each individual message that it does not have access to or is intended for it (though the user may learn some information about the collection of all such messages). In the context of eavesdropper security, a similar notion of secrecy has been called 1-block security [2] or weak security [4]. In this work, we will call this *weak privacy*. We aim to characterize the index coding instances where we can achieve weak privacy. Finally, we consider a setup where there are no shared keys, but the server can multicast to subsets of users. We are interested in the minimum number of multicasts required to achieve (strong) privacy.

The paper is organized as follows. We describe our private index coding setup in Section II. We give the characterization of a feasible key distribution in Section III. We discuss our results on linear coding in Section IV. Characterizations of rate regions for small networks are discussed in Section V. Our results on weak privacy and private index coding through multicast sessions are provided in Section VI and Section VII respectively.

II. Problem Formulation and Preliminaries

The server possesses N messages, $X_1, \dots, X_N$ and user $i \in [N]$ wants the message X_i. We assume that X_i's are independent and take values uniformly in a field $\mathbb{F}$. We allow block coding, i.e., the server observes n independent copies of each message before transmission. The n independent copies of X_i is denoted by $X_i^{(n)}$. The indices of the messages that are available at user i as side information is represented by $\mathcal{A}_i \subseteq [N] \setminus \{i\}$, then the set of messages possessed by user i is represented as $X_{\mathcal{A}_i}$. In general, for a subset of indices $\mathcal{S} \subseteq [N]$, the set of messages $\{X_i : i \in \mathcal{S}\}$ is represented by $X_{\mathcal{S}}$. Let $\overline{\mathcal{A}}_i$ represent the set $\mathcal{A}_i \cup \{i\}$. Index coding problem can be represented by a directed graph G with vertex set $V = [N]$ and edge set E, where $(i, j) \in E$ if and only if $j \in \mathcal{A}_i$. Complement of graph G, denoted by G^c, has vertex set $V = [N]$ and $(i, j) \in E(G^c)$ if and only if $(i, j) \notin E(G)$.

The privacy requirement we consider is that user i should not obtain any information about $X_{[N]\setminus\overline{\mathcal{A}}_i}$. The server has access to *keys* that are shared among various subsets of users. For $\mathcal{S} \subsetneq [N], \mathcal{S} \neq \emptyset$, the key that is available exclusively to users in $\mathcal{S}$ is represented by $K_{\mathbf{b}}$, where $\mathbf{b} \in \{0, 1\}^N$ is the characteristic vector for the set $\mathcal{S}$, *i.e.,* i^{th} bit in $\mathbf{b}$ is 1 if and only if $i \in \mathcal{S}$. A key $K_{\mathbf{b}}$ is a random variable that is independent of the messages and other keys and is uniformly distributed in the set $\{1, \cdots, |\mathbb{F}|^{nR_{\mathbf{b}}}\}$, where $R_{\mathbf{b}}$ denotes the rate of that key[1]. So the set of all keys is indexed by the set $\{\mathbf{b} : \mathbf{b} \in \{0, 1\}^N\}$. The key $K_{\vec{0}}$ which is not available at any user can be viewed as part of the private randomness at the server. Since we will consider a randomized encoder, without loss of generality we set $R_{\vec{0}}$ to 0. In the extended version, we argue that public randomness can be ignored without loss of generality, hence we take $R_{\vec{1}}$ to be 0.

The *key access structure* of a private index code instance, $\mathcal{B}$, is the set of indices corresponding to keys with non-zero rates, *i.e.,*

$$\mathcal{B} = \{\mathbf{b} : \mathbf{b} \in \{0, 1\}^N \setminus \{\vec{0}, \vec{1}\}, R_{\mathbf{b}} \neq 0\}.$$

For $i \in [N]$, let b_i denote i^{th} bit in $\mathbf{b}$, then $\mathcal{B}_i$ denotes the indices in the key access structure which are available to the user i, i.e., $\mathcal{B}_i = \{\mathbf{b} \in \mathcal{B} : b_i = 1\}$.

A transmission scheme consists of a possibly randomized encoder

$$\phi : \prod_{\mathbf{b}\in\mathcal{B}} [|\mathbb{F}|^{nR_{\mathbf{b}}}] \times \prod_{i\in[N]} \mathbb{F}^n \longrightarrow [|\mathbb{F}|^{nR_M}] \tag{1}$$

that outputs the transmitted message M. The rate of transmission is R_M. Each user i uses a deterministic decoder

$$\psi_i : [|\mathbb{F}|^{nR_M}] \times \prod_{\mathbf{b}\in\mathcal{B}_i} [|\mathbb{F}|^{nR_{\mathbf{b}}}] \times \prod_{j\in\mathcal{A}_i} \mathbb{F}^n \longrightarrow \mathbb{F}^n. \tag{2}$$

The rate of the scheme is the $(2^N - 1)-$tuple of R_M and the key rates $R_{\mathbf{b}}, \mathbf{b} \in \{0, 1\}^N \setminus \{\vec{0}, \vec{1}\}$. A given rate is said to be achievable if for some $n \geq 1$, there exists a transmission scheme with these rates such that

$$H\left(X_i^{(n)} \middle| K_{\mathcal{B}_i}, X_{\mathcal{A}_i}^{(n)}, M\right) = 0 \text{ for all } i \in [N], \tag{3}$$

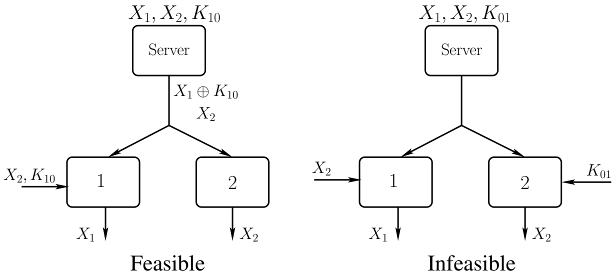
$$I\left(M; X_{[N]\setminus\overline{\mathcal{A}}_i}^{(n)} \middle| K_{\mathcal{B}_i}, X_{\mathcal{A}_i}^{(n)}\right) = 0 \text{ for all } i \in [N]. \tag{4}$$

For an index coding problem represented by graph G, the *rate region* is defined as the closure of the convex hull of all achievable rate tuples, and it is denoted by $\mathcal{R}(G)$.

Definition 1. *A key access structure $\mathcal{B}$ is said to be* feasible *if for some $n \geq 1$ there exists a transmission scheme which achieves* (3) *and* (4) *with $R_{\mathbf{b}} = 0$ for all $\mathbf{b} \notin \mathcal{B}$.*

III. Feasibility of Private Index Coding

In private index coding, achieving (3) and (4) relies on the availability of certain keys among users. Hence, the feasibility of private index coding depends on the key access structure (eg., Fig. 1). The following theorem characterizes the feasible key access structures for a private index coding problem.



Feasible

Infeasible

Fig. 1: Feasible and infeasible key access structures

Theorem 1. *A key access structure $\mathcal{B}$ is feasible if and only if $\forall\, i, j \in [N]$ such that $i \notin \overline{\mathcal{A}}_j$, there exists a $\mathbf{b} \in \mathcal{B}$ such that $b_i = 1, b_j = 0$.*

The full proof of this theorem is provided in the extended version of the paper. An overview follows.

a) Only if part: By the privacy condition (4) at user j,

$$I\left(M \,;\, X_{[N]\setminus\overline{\mathcal{A}}_j}^{(n)} \middle| X_{\mathcal{A}_j}^{(n)}, K_{\mathcal{B}_j}\right) = 0.$$

But, by the decoding condition (3) at user j, $X_j^{(n)}$ is a function of $M, X_{\mathcal{A}_j}^{(n)}, K_{\mathcal{B}_j}$, hence

$$I\left(X_j^{(n)}, M \,;\, X_{[N]\setminus\overline{\mathcal{A}}_j}^{(n)} \middle| X_{\mathcal{A}_j}^{(n)}, K_{\mathcal{B}_j}\right) = 0.$$

Since, $i \notin \overline{\mathcal{A}}_j$, using chain rule,

$$I\left(M \,;\, X_i^{(n)} \middle| X_{[N]\setminus\{i\}}^{(n)}, K_{\mathcal{B}_j}\right) = 0.$$

Using independence of messages and keys, it can be shown from the above that

$$I\left(M \,;\, X_i^{(n)} \middle| X_{[N]\setminus\{i\}}^{(n)}, K_{\mathcal{B}_j \cap \mathcal{B}_i}\right) = 0. \tag{5}$$

By the decoding condition (3) at user i,

$$I\left(M; X_i^{(n)} \middle| X_{[N]\setminus\{i\}}^{(n)}, K_{\mathcal{B}_i}\right) = H\left(X_i^{(n)}\right). \tag{6}$$

[1] Rates and entropies in this paper are expressed in units of $\log |\mathbb{F}|$ bits.

Using the independence of messages and keys, $H(K_{\mathcal{B}_i \setminus \mathcal{B}_j})$ (and, hence, $n \sum_{\mathbf{b} \in \mathcal{B}_i \setminus \mathcal{B}_j} R_{\mathbf{b}}$) is lower bounded by

$$\begin{aligned} & I\left(M; K_{\mathcal{B}_i \setminus \mathcal{B}_j} \middle| X^{(n)}_{[N]\setminus\{i\}}, X_i^{(n)}, K_{\mathcal{B}_j \cap \mathcal{B}_i}\right) \\ & \overset{(a)}{=} I\left(M; K_{\mathcal{B}_i \setminus \mathcal{B}_j}, X_i^{(n)} \middle| X^{(n)}_{[N]\setminus\{i\}}, K_{\mathcal{B}_j \cap \mathcal{B}_i}\right) \\ & \geq I\left(M; X_i^{(n)} \middle| X^{(n)}_{[N]\setminus\{i\}}, K_{\mathcal{B}_i}\right) \overset{(b)}{=} H\left(X_i^{(n)}\right), \end{aligned} \tag{7}$$

where (a) follows from (5) and (b) follows from (6). Hence, $n \sum_{\mathbf{b} \in \mathcal{B}_i \setminus \mathcal{B}_j} R_{\mathbf{b}} \geq H(X_i^{(n)})$. Since $H(X_i^{(n)}) > 0$, $R_{\mathbf{b}}$ is non-zero for some $\mathbf{b} \in \mathcal{B}_i \setminus \mathcal{B}_j$.

b) If part: We can show that the following is a valid linear private index coding scheme. We use N independent copies of each key in the key access structure, *i.e.*, $\forall \mathbf{b} \in \mathcal{B}, K_{\mathbf{b}} \in \mathbb{F}^N$. Let $K^i_{\mathbf{b}}$ be the i^{th} coordinate of the key $K_{\mathbf{b}}$. The transmission M consists of N parts, M_i, $i \in [N]$,

$$M_i = X_i + \sum_{\mathbf{b} \in \mathcal{B}_i} K^i_{\mathbf{b}}. \tag{8}$$

IV. Linear Private Index Codes

In this section we consider linear coding schemes for private index coding. We characterize linear schemes that satisfy the decoding (3) and privacy (4) conditions.

In the context of linear coding, for block length $n \geq 1$, let $\mathbf{X}_i$ denote the row-vector corresponding to $X_i^{(n)}$ and $\mathbf{K}_{\mathbf{b}}$ denote the key uniformly distributed in $\mathbb{F}^{nR_{\mathbf{b}}}$, where $nR_{\mathbf{b}}$ is assumed to be an integer. The linear encoder is of the form

$$M = \sum_{i \in [N]} \mathbf{G}_i \mathbf{X}_i^T + \sum_{\mathbf{b} \in \mathcal{B}} \mathbf{H}_{\mathbf{b}} \mathbf{K}_{\mathbf{b}}^T, \tag{9}$$

where $\mathbf{G}_i \in \mathbb{F}^{r \times n}$ for $i \in [N]$ and $\mathbf{H}_{\mathbf{b}} \in \mathbb{F}^{r \times nR_{\mathbf{b}}}$ for $\mathbf{b} \in \mathcal{B}$. Transmission rate R is said to be achievable if for some $n \geq 1$ there exists a scheme such that $R = r/n$ and it satisfies (3) and (4). If $n = 1$, the scheme is called a *scalar linear code*.

Theorem 2. *A linear encoding scheme is a valid private index coding scheme if and only if it satisfies the following conditions for each $i \in [N]$,*

1) *Let $\mathbf{G}_i = [g_i^1 \ldots g_i^n]$, then for each $1 \leq k \leq n$,*

$$g_i^k \notin \langle \{\mathbf{G}_j\}_{j \notin \overline{\mathcal{A}}_i}, \{\mathbf{H}_{\mathbf{b}}\}_{\mathbf{b} \notin \mathcal{B}_i} \rangle,$$

2) $\langle \{\mathbf{G}_j\}_{j \notin \overline{\mathcal{A}}_i} \rangle \subseteq \langle \{\mathbf{H}_{\mathbf{b}}\}_{\mathbf{b} \notin \mathcal{B}_i} \rangle.$

Here $\langle . \rangle$ denotes the linear span of column vectors. The proof of this theorem is provided in the extended version. The first condition is the decodability condition at user i and is similar to that in index coding. The second condition is the privacy condition against user i.

V. Rate of Private Index Coding

We first observe a simple connection between private index coding and index coding problems with the same side information structure. Given an index coding scheme (specifically, the optimal scheme), we describe a private index code (for a certain key access structure we specify below) with the same transmission rate: Let $K_{\mathbf{b}(i)}$ be a key of the same rate as X_i that is available at user i and all users j such that $i \in \mathcal{A}_j$. Taking $\{X_i + K_{\mathbf{b}(i)}, i \in [N]\}$ as the messages, the index coding scheme can be employed to deliver $X_i + K_{\mathbf{b}(i)}$ to user i, $i \in [N]$; note that user i has access to side-information $\{X_j + K_{\mathbf{b}(j)} : j \in \mathcal{A}_i\}$ as required. Having access to $K_{\mathbf{b}(i)}$, user i can recover X_i. Privacy follows from the fact that $K_{\mathbf{b}(i)}$ is unavailable to any user who should not learn X_i. Thus, the optimal transmission rate of the index coding problem is also achievable in private index coding for a certain key access structure. Clearly, the minimum transmission rate of private index coding cannot be less than that of index coding since setting the keys of the private index code to some arbitrary fixed values gives an index code. Thus we observe:

Observation 1. *For a given side information structure, optimal transmission rates of index coding and private index coding (optimized over key access structures and key rates) are the same.*

The following theorem shows that when $N \leq 3$, rate region of transmission rate and key rates for private index coding can be characterized.

Theorem 3. *For every private index coding instance G with at most 3 users, the rate region $\mathcal{R}(G)$ is achievable using scalar linear codes and time sharing.*

A full proof of this theorem and *the characterization of the rate region* for every private index coding instance with $N \leq 3$ is presented in the extended version of the paper. Here we illustrate the proof method using an example. Consider the graph given in Table I. We first show the necessity of the constraints on the rates. In arguing the "only if" part of Theorem 1 we showed that if $i \notin \overline{\mathcal{A}}_j$, then $\sum_{\mathbf{b} \in \mathcal{B}_i \setminus \mathcal{B}_j} R_{\mathbf{b}} \geq H(X_i^{(n)})/n$. The first three inequalities in the table follow from this using $H(X_i^{(n)})/n = 1, i = 1, 2, 3$. To see the next inequality, note that the transmission rate of a private index code is lower bounded by the rate of the index coding problem for the same side information graph. Hence, R_M is lower bounded by number of vertices in the maximum acyclic induced subgraph [8], which is 2 in this example. To show the next inequality, consider

$$\begin{aligned} & H(M) \geq I(M; X_1^{(n)}, X_2^{(n)}, X_3^{(n)}, K_{001}, \ldots, K_{110}) \\ & = I\left(M; X_2^{(n)}, K_{100}, K_{110}, K_{101}\right) + \\ & \quad I\left(M; X_1^{(n)} \middle| X_2^{(n)}, K_{100}, K_{110}, K_{101}\right) + \\ & \quad I\left(M; K_{001}, K_{011} \middle| X_1^{(n)}, X_2^{(n)}, K_{100}, K_{110}, K_{101}\right) + \\ & \quad I\left(M; X_3^{(n)} \middle| X_1^{(n)}, X_2^{(n)}, K_{100}, K_{110}, K_{101}, K_{001}, K_{011}\right) + \\ & \quad I\left(M; K_{010} \middle| X_{[3]}^{(n)}, K_{100}, K_{110}, K_{101}, K_{010}, K_{011}\right). \end{aligned}$$

We lower bound the terms in red by zero. The decodability condition at user 1 implies that the first term in black is $H(X_1^{(n)})$. Using the independence of keys and messages, the

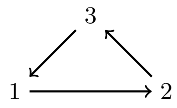
		Vertices	Coding Scheme
	$R_{100}+R_{101}\geq 1$ $R_{010}+R_{110}\geq 1$ $R_{001}+R_{011}\geq 1$ $R_M\geq 2$ $R_M+R_{011}\geq 3$ $R_M+R_{101}\geq 3$ $R_M+R_{110}\geq 3$ $R_{\mathbf{b}}\geq 0, 1\leq \mathbf{b}\leq 6$	$(2,0,0,1,0,1,1)$	$X_1+X_2+K_{101}+K_{110}, X_2+X_3+K_{110}+K_{011}$
		$(3,0,0,1,1,1,0)$	$X_1+K_{101}, X_2+K_{110}, X_3+K_{001}$
		$(3,0,1,0,0,1,1)$	$X_1+K_{101}, X_2+K_{010}, X_3+K_{011}$
		$(3,0,1,0,1,1,0)$	$X_1+K_{101}, X_2+K_{010}, X_3+K_{001}$
		$(3,1,0,1,0,0,1)$	$X_1+K_{100}, X_2+K_{110}, X_3+K_{011}$
		$(3,1,0,1,1,0,0)$	$X_1+K_{100}, X_2+K_{110}, X_3+K_{001}$
		$(3,1,1,0,0,0,1)$	$X_1+K_{100}, X_2+K_{010}, X_3+K_{011}$
		$(3,1,1,0,1,0,0)$	$X_1+K_{100}, X_2+K_{010}, X_3+K_{001}$

TABLE I: Vertices are represented as a tuple $(R_M, R_{100}, R_{010}, R_{110}, R_{001}, R_{101}, R_{011})$

second term in black can be lower bounded by

$$I\left(M; X_3^{(n)}\middle| X_1^{(n)}, K_{101}, K_{001}, K_{011}\right) \overset{(a)}{=} H(X_3^{(n)}),$$

where (a) follows from the decodability condition at user 3. To bound the third term in black, we note that

$$\begin{aligned}&I\left(M; K_{010}, K_{110}\middle| X_{[3]}^{(n)}, K_{100}, K_{101}, K_{010}, K_{110}\right)\\ &\overset{(a)}{\geq} I\left(M; K_{010}, K_{110}\middle| X_{[3]}^{(n)}, K_{101}\right) \overset{(b)}{\geq} H\left(X_2^{(n)}\right),\end{aligned}$$

where (a) can be shown using the independence of messages and keys and (b) follows from (7) with $i=2$ and $j=3$, since $2\notin\overline{\mathcal{A}}_3$. From this, we can show that the third term in black is lower bounded by $H(X_2^{(n)})-H(K_{110})$. Putting all these together we have $H(M)\geq\sum_{i\in[3]}H(X_i^{(n)})-H(K_{110})$ which implies that $R_M\geq 3-R_{110}$, similarly we get the next two inequalities. The table shows that the vertices of the polygon described by these inequalities can be achieved using scalar linear codes. In the extended version, we also show:

Proposition 1. *There is a 4 user private index coding problem where a vector linear code obtains a rate tuple outside the rate region obtained by scalar linear coding and time sharing.*

VI. Weak Privacy

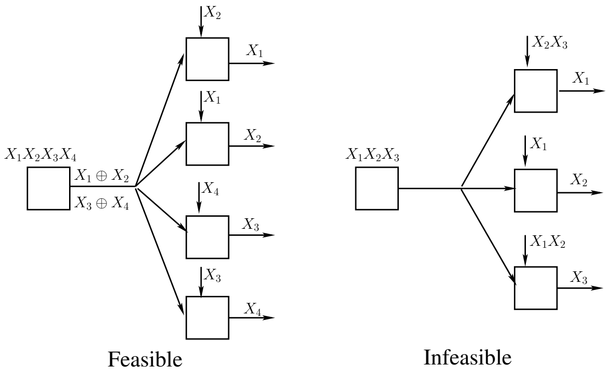
Theorem 1 shows that if the goal is to achieve the privacy required by (4), then, in all but trivial cases, we need to distribute keys among the users. In this section, we consider a model without the extra resource of keys. In the absence of keys, we aim to achieve *weak privacy.* Encoding and decoding are similar to (1) and (2) respectively without any keys. A transmission scheme with broadcast message M with *weak privacy* satisfies

$$H\left(X_i^{(n)}\middle| X_{\mathcal{A}_i}^{(n)}, M\right)=0 \text{ for all } i\in[N], \tag{10}$$

and

$$I\left(M; X_j^{(n)}\middle| X_{\mathcal{A}_i}^{(n)}\right)=0 \text{ for all } i\in[N], j\in[N]\setminus\overline{\mathcal{A}}_i. \tag{11}$$

For example, in weak privacy, if user i does not have X_j and X_k as side information, then the user must not learn anything about X_j or X_k individually, but user i is allowed to get some information about the pair (X_j, X_k).



Feasible

Infeasible

Fig. 2: For the 4 user network, transmitting $X_1\oplus X_2$ and $X_3\oplus X_4$ will suffice. For the 3 user network, decodability at user 3 demands $H(X_3^{(n)}|M, X_1^{(n)}, X_2^{(n)})=0$. But privacy at user 2 implies that $I(M, X_1^{(n)}, X_2^{(n)}; X_3^{(n)})=0 \implies H(X_3^{(n)}|M, X_1^{(n)}, X_2^{(n)})\neq 0$. So there is no scheme for this network.

A. Feasibility of Weak Privacy

Weak privacy is not possible to achieve for all index coding instances. Fig. 2 shows a feasible and an infeasible instance of index coding problems under weak privacy. Next we study the feasibility of index coding under weak privacy. We first give some necessary conditions that the network should satisfy in order to be feasible. We start with a simple subset condition that any pair of nodes should satisfy in order to be feasible. This condition is similar to Theorem 1.

Proposition 2 (Subset Condition)**.** *An index coding problem under weak privacy is feasible only if,* $\forall\, i,j\in[N], i\neq j$ *such that* $i\notin\mathcal{A}_j$*, it holds that* $\mathcal{A}_i\subsetneq\overline{\mathcal{A}}_j$*.*

The proof of Proposition 2 uses the fact that if $\mathcal{A}_i\subseteq\mathcal{A}_j$ and user j does not have X_i as side information, then user j also learns X_i if user i learns X_i from the broadcast message. The subset condition is not sufficient for feasibility. There is no scheme for the network in Fig. 3 (a) though it satisfies Proposition 2. Details are given in the extended version. The condition is extended in the following proposition to get a better necessary condition for feasibility.

Proposition 3. *If* $\exists$ *a user* j *s.t. for any subset* $S\subseteq\overline{\mathcal{A}}_j$ *where* $j\in S$*,* $\exists$ *users* i *and* l *s.t.* $k\in S\setminus\overline{\mathcal{A}}_i$ *and* $j,k\in S\setminus\overline{\mathcal{A}}_l$*, then* *there is no scheme achieving weak privacy.*

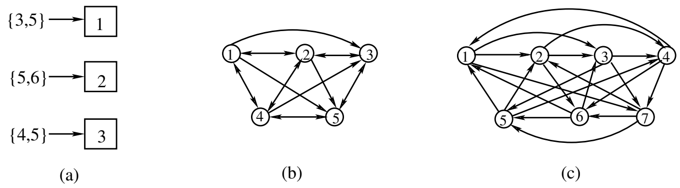


Fig. 3: Examples: (a) infeasible although it satisfies Prop. 2, (b) has a scheme using secure clique cover (Prop. 4) and (c) feasible using a linear code, but has no secure clique cover.

Next we give a sufficient condition to achieve weak privacy. We first give the definition of secure clique cover. A clique cover of a graph is a set of cliques which cover all the vertices.

Definition 2 (Secure clique cover)**.** *A clique cover C_G of G is said to be* secure *if it satisfies the following two conditions:*
1) For any clique $c \in C_G$ with $|c| = k$ and $k \geq 2$, $\nexists$ a vertex $v \in c$ s.t. v has exactly $k-1$ incoming edges from the nodes in $V \setminus c$.
2) For any singleton in C_G, there is an incoming edge from all other vertices to that vertex.

Proposition 4. *For the index coding problem represented by G, weak privacy is achievable if G has a secure clique cover.*

The proof of Proposition 4 is given in the extended version. Observe that $\{\{1,2,4\},\{3,5\}\}$ is a secure clique cover for the graph shown in Fig. 3 (b).

B. Linear Coding for weak privacy

Linear encoding for weak privacy is similar to (9) with only the first term involving $\mathbf{G}_i$s. In Theorem 4, we give the necessary and sufficient condition for the encoding matrix to satisfy for obtaining weak privacy. We use the same notations that we used to describe linear coding in Section IV. The proof of Theorem 4 is along similar lines as that of Theorem 2, hence it is omitted.

Theorem 4. *The matrices $(\mathbf{G}_i)_{i\in[N]}$ is a valid encoding scheme under weak privacy if and only if they satisfy the following conditions for each $i \in [N]$,*

1) $g_i^k \notin \langle \{\mathbf{G}_j\}_{j\notin\overline{\mathcal{A}}_i} \rangle$, for $1 \leq k \leq n$,

2) For $j \notin \overline{\mathcal{A}}_i$, $\langle \mathbf{G}_j \rangle \subseteq \langle \{\mathbf{G}_k\}_{k\notin\overline{\mathcal{A}}_i} \rangle$.

For the index coding problem represented by the graph shown in Fig. 3 (c), there is a linear coding scheme which achieves weak privacy, but there is no secure clique cover. This shows that secure clique cover is not necessary to achieve weak privacy. Details are provided in the extended version.

VII. Privacy Through Multicasts

We consider a model in which there is no shared key between the server and the users. However, the server can multicast to any subset of users. A multicast session is defined as transmitting one element from field $\mathbb{F}$ to a subset of users. A scheme of multicast transmissions is required to satisfy (3) and (4). Given an index coding instance, we are interested in determining the minimum number of multicast sessions required. Let $MS(n)$ denote the minimum number of multicast sessions required for n-instances of the messages. We define $MS^* \triangleq \inf_n \frac{MS(n)}{n}$. Theorem 5 characterizes MS^*. Then we show that the optimal scheme of multicast can be achieved by using a particular key distribution and a transmission scheme for the private index coding problem. We use the following graph theoretic notions to study this problem.

For a graph G, assigning a subset of size b of the set $\{1,\cdots,L\}$ to each node of a graph such that any two adjacent nodes get disjoint sets is called a b-fold coloring. Minimum size of the set $\{1,\cdots,L\}$ required for b-fold coloring is the *b-fold chromatic number*, denoted by $\chi_b(G)$. The *fractional chromatic number $\chi_f(G)$* is defined as

$$\chi_f(G) = \lim_{b\to\infty} \frac{\chi_b(G)}{b} = \inf_b \frac{\chi_b(G)}{b}.$$

Since $\chi_b(G)$ is subadditive, the limit exists.

Theorem 5. *For the index coding problem represented by G, $MS^* = \chi_f(G^c)$.*

Theorem 5 is proved in the extended version. We would also like to note that, given a multicast scheme, there is a natural private index coding scheme in which, corresponding to i^{th} multicast session that sends $M_i \in \mathbb{F}$ to a subset of users $\mathcal{S}_i \subseteq [N]$, there is a secret key $K_{\mathbf{b}_i}$ that is shared by the server and the users in $\mathcal{S}_i$ and a server transmission $M_i + K_{\mathbf{b}_i}$.

Acknowledgments

V. Prabhakaran and N. Karamchandani acknowledge initial discussions with Parathasarathi Panda and Vaishakh Ravi. V. Narayanan was supported by a travel fellowship from the Sarojini Damodaran Foundation. This work was done while Jithin Ravi was at Tata Institute of Fundamental Research and IIT Bombay. He has received funding from ERC grant 714161.

References

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

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

[3] M. M. Mojahedian, A. Gohari, and M. R. Aref, "Perfectly secure index coding," in *ISIT*, Jun. 2015, pp. 1432–1436.

[4] L. Ong, B. N. Vellambi, P. L. Yeoh, J. Kliewer, and J. Yuan, "Secure index coding: Existence and construction," in *ISIT*, Jul. 2016, pp. 2834–2838.

[5] L. Ong, B. N. Vellambi, J. Kliewer, and P. L. Yeoh, "An equivalence between secure network and index coding," in *Globecom Workshops*, Dec. 2016, pp. 1–6.

[6] M. Karmoose, L. Song, M. Cardone, and C. Fragouli, "Private broadcasting: An index coding approach," in *ISIT*, Jun. 2017, pp. 2543–2547.

[7] L. Ong, "Linear codes are optimal for index-coding instances with five or fewer receivers," in *ISIT*, Jun. 2014, pp. 491–495.

[8] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1479–1494, Mar. 2011.