Index Coding and Error Correction

Son Hoang Dau, Vitaly Skachek, and Yeow Meng Chee
Division of Mathematical Sciences, School of Physical and Mathematical Sciences
Nanyang Technological University, 21 Nanyang Link, Singapore 637371
Emails: {DauS0002, Vitaly.Skachek, YMChee}@ntu.edu.sg

Abstract—A problem of index coding with side information was first considered by Y. Birk and T. Kol (IEEE INFOCOM, 1998). In the present work, a generalization of index coding scheme, where transmitted symbols are subject to errors, is studied. Error-correcting methods for such a scheme, and their parameters, are investigated. In particular, the following question is discussed: given the side information hypergraph of index coding scheme and the maximal number of erroneous symbols \( \delta \), what is the shortest length of a linear index code, such that every receiver is able to recover the required information? This question turns out to be a generalization of the problem of finding a shortest-length error-correcting code with a prescribed error-correcting capability in the classical coding theory.

The Singleton bound and two other bounds, referred to as the \( \alpha \)-bound and the \( \kappa \)-bound, for the optimal length of a linear error-correcting index code (ECIC) are established. For large alphabets, a construction based on concatenation of an optimal index code with an MDS classical code, is shown to attain the Singleton bound. For smaller alphabets, however, this construction may not be optimal. A random construction is also analyzed. It yields another inexplicit bound on the length of an optimal linear ECIC. Finally, the decoding of linear ECIC’s is discussed. The syndrome decoding is shown to output the exact message if the weight of the error vector is less or equal to the error-correcting capability of the corresponding ECIC.

I. INTRODUCTION

A. Background

The problem of Index Coding with Side Information (ICSI) was introduced by Birk and Kol [1]. During the transmission, each client might miss a certain part of the data, due to intermittent reception, limited storage capacity or any other reasons. Via a slow backward channel, the clients let the server know which messages they already have in their possession, and which messages they are interested to receive. The server has to find a way to deliver to each client all the messages he requested, yet spending a minimum number of transmissions. As it was shown in [1], the server can significantly reduce the number of transmissions by coding the messages.

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

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

B. Our contribution

In this work, we generalize the ICSI problem towards a setup with error correction. We extend some known results on index coding to a case where any receiver can correct up to a certain number of errors. The problem of designing such error-correcting index codes (ECIC’s) naturally generalizes the problem of constructing classical error-correcting codes. We establish an upper bound (the \( \kappa \)-bound) and a lower bound (the \( \alpha \)-bound) on the shortest length of a linear ECIC, which is able to correct any error pattern of size up to \( \delta \). We also derive an analog of the Singleton bound, and show that this bound is tight for codes over large alphabets. We also consider random ECIC’s. By analyzing their parameters, we obtain an upper bound on their length. Finally, we discuss the decoding of linear ECIC’s. We show that the syndrome decoding results in a correct result, provided that the number of errors does not exceed the error-correcting capability of the code.

The problem of error correction for NC was studied in several previous works. However, these results are not directly applicable for the ICSI problem. First, the existing works only consider the multicast scenario, while the ICSI problem, however, is a special case of the non-multicast NC problem. Second, the ICSI problem can be modeled by the NC scenario [3], yet, this requires that there are directed edges from particular sources to each sink, which provide the side information. The symbols transmitted on these special edges, unlike for error-correcting NC, are not allowed to be corrupted.

II. PRELIMINARIES

Let \( \mathbb{F}_q \) be the finite field of \( q \) elements, where \( q \) is a power of prime, and \( \mathbb{F}_q^n = \mathbb{F}_q \setminus \{0\} \). Let \( [n] = \{1, 2, \ldots, n\} \). For the vectors \( u, v \in \mathbb{F}_q^n \), we use \( d(u, v) \) to denote the the Hamming distance between \( u \) and \( v \). If \( u \in \mathbb{F}_q^n \) and \( M \subseteq \mathbb{F}_q^n \) is a set of vectors (or a vector subspace), then this notation can be extended to

\[
d(u, M) = \min_{v \in M} d(u, v).
\]

Given \( q, k, \) and \( d \), let \( N_q[k, d] \) denote the length of the shortest linear code over \( \mathbb{F}_q \) which has dimension \( k \) and minimum distance \( d \). The support of a vector \( u \in \mathbb{F}_q^n \) is defined by \( \text{supp}(u) \triangleq \{i \in [n] : u_i \neq 0\} \). The Hamming weight of \( u \) is defined by \( \text{wt}(u) \triangleq |\text{supp}(u)| \). Suppose \( E \subseteq [n] \). We write \( u \prec E \) whenever \( \text{supp}(u) \subseteq E \).

We use \( e_i = (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{F}_q^n \) to denote the unit vector, which has a one at the \( i \)-th position, and zeros
III. ERROR-CORRECTING INDEX CODING WITH SIDE INFORMATION

Index Coding with Side Information problem considers the following communications scenario. There is a unique sender (or source) $S$, who has a vector of messages $x = (x_1, x_2, \ldots, x_n)$ in his possession. There are also $m$ receivers $R_1, R_2, \ldots, R_m$, receiving information from $S$ via a broadcast channel. For each $i \in [m]$, $R_i$ has side information, i.e. $R_i$ owns a subset of messages $\{x_j\}_{j \in X_i}$, where $X_i \subseteq [n]$. Each $R_i$, $i \in [m]$, is interested in receiving the message $x_{f(i)}$ (we say that $R_i$ requires $x_{f(i)}$), where the mapping $f : [m] \rightarrow [n]$ satisfies $f(i) \notin X_i$ for all $i \in [m]$. Hereafter, we use the notation $X' = (X_1, X_2, \ldots, X_m)$. An instance of the ICSI problem is given by a quadruple $(m, n, X', f)$. An instance of the ICSI problem can also be conveniently described by the following directed hypergraph $\mathcal{H}$.

Definition 3.1: Let $(m, n, X', f)$ be an instance of the ICSI problem. The corresponding side information (directed) hypergraph $\mathcal{H} = \mathcal{H}(m, n, X', f)$ is defined by the vertex set $V = [n]$ and the edge set $\mathcal{E}_\mathcal{H}$, where

$$\mathcal{E}_\mathcal{H} = \{(f(i), X_i) : i \in [n]\}.$$ 

We often refer to $(m, n, X', f)$ as an instance of the ICSI problem described by the hypergraph $\mathcal{H}$.

Each side information hypergraph $\mathcal{H} = (V, \mathcal{E}_\mathcal{H})$ can be associated with the directed graph $\mathcal{G}_\mathcal{H} = (V, \mathcal{E})$ in the following way. For each directed edge $(f(i), X_i) \in \mathcal{E}_\mathcal{H}$, there will be $|X_i|$ directed edges $(f(i), v) \in \mathcal{E}$, for $v \in X_i$. When $m = n$ and $f(i) = i$ for all $i \in [m]$, the graph $\mathcal{G}_\mathcal{H}$ is, in fact, the side information graph, defined in [3].

Due to noise, the symbols received by $R_i$, $i \in [m]$, may be subject to errors. Assume that $S$ broadcasts a vector $y \in \mathbb{F}_q^n$. Let $e_i \in \mathbb{F}_q^n$ be the error affecting the information received by $R_i$, $i \in [m]$. Then $R_i$ actually receives the vector $y_i = y + e_i \in \mathbb{F}_q^n$, instead of $y$.

Definition 3.2: Consider an instance of the ICSI problem described by $\mathcal{H} = \mathcal{H}(m, n, X', f)$. A $\delta$-error-correcting index code ($(\delta, \mathcal{H})$-ECIC) over $\mathbb{F}_q$ for this instance is an encoding function

$$\mathcal{E} : \mathbb{F}_q^n \rightarrow \mathbb{F}_q^N,$$

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

$$\mathcal{D}_i : \mathbb{F}_q^N \times \mathbb{F}_q^{\lvert X_i \rvert} \rightarrow \mathbb{F}_q,$$

satisfying

$$\forall x, e_i \in \mathbb{F}_q^n, \text{ wt}(e_i) \leq \delta : \mathcal{D}_i(\mathcal{E}(x) + e_i, x_{X_i}) = x_{f(i)}.$$ 

If $\delta = 0$, we refer to such $\mathcal{E}$ as a non-error-correcting index code, or just $\mathcal{H}$-IC. The parameter $\delta$ is called the length of the index code. In the scheme corresponding to the code $\mathcal{E}$, $S$ broadcasts a vector $\mathcal{E}(x)$ of length $N$ over $\mathbb{F}_q$.

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

$$\forall x \in \mathbb{F}_q^n : \mathcal{E}(x) = xL,$$

where $L$ is an $n \times N$ matrix over $\mathbb{F}_q$. The matrix $L$ is called the matrix corresponding to the index code $\mathcal{E}$, while $\mathcal{E}$ is referred to as the linear index code based on $L$.

Definition 3.4: An optimal linear $(\delta, \mathcal{H})$-ECIC over $\mathbb{F}_q$ is a linear $(\delta, \mathcal{H})$-ECIC over $\mathbb{F}_q$ of the smallest possible length $N_q(\mathcal{H}, \delta)$.

Hereafter, we assume that $X' = (X_i)_{i \in [m]}$ is known to $S$. We also assume that the code $\mathcal{E}$ is known to each receiver $R_i$, $i \in [m]$.

Definition 3.5: Suppose $\mathcal{H} = \mathcal{H}(m, n, X, f)$ corresponds to an instance of the ICSI problem. Then the min-rank of $\mathcal{H}$ over $\mathbb{F}_q$ is defined as

$$\kappa_q(\mathcal{H}) \triangleq \min \{ \text{rank}_{\mathbb{F}_q}(\{v_i + e_{f(i)}\}_{i \in [m]} : v_i \in \mathbb{F}_q^n, v_i \in X_i\}.$$

Observe that $\kappa_q(\mathcal{H})$ generalizes the min-rank over $\mathbb{F}_q$ of the side information graph, which was defined in [3]. More specifically, when $m = n$ and $f(i) = i$ for all $i \in [m]$, $\mathcal{G}_{\mathcal{H}}$ becomes the side information graph, and $\kappa_q(\mathcal{H}) = \text{min-rank}_{\mathbb{F}_q}(\mathcal{G}_{\mathcal{H}})$. The min-rank was shown in [3, 4] to be the smallest number of transmissions in a linear index code.

Lemma 3.1: (3, 11) Consider an instance of the ICSI problem described by $\mathcal{H} = \mathcal{H}(m, n, X, f)$.

1) The matrix $L$ corresponds to a linear $\mathcal{H}$-IC over $\mathbb{F}_q$ if and only if for each $i \in [m]$ there exists $v_i \in \mathbb{F}_q^n$ such that $v_i \in X_i$ and $v_i + e_{f(i)} \in \text{colspan}(L)$.

2) The smallest possible length of a linear $\mathcal{H}$-IC over $\mathbb{F}_q$ is $\kappa_q(\mathcal{H})$. 

IV. BASIC PROPERTIES

We define the set of vectors
\[ \mathcal{I}(q, \mathcal{H}) \triangleq \{ z \in \mathbb{F}_q^n : \exists i \in [m] \text{ s.t. } z_{X_i} = 0, \ z_{f(i)} \neq 0 \} . \]
For all \( i \in [m] \), we also define \( \mathcal{Y}_i \triangleq [n] \setminus \{ f(i) \} \cup X_i \). Then the collection of supports of all vectors in \( \mathcal{I}(q, \mathcal{H}) \) is given by
\[ \mathcal{J}(\mathcal{H}) \triangleq \bigcup_{i \in [m]} \{ f(i) \} \cup Y_i : Y_i \subseteq \mathcal{Y}_i \}. \]

Lemma 4.1: The matrix \( L \) corresponds to a \((\delta, \mathcal{H})\)-ECIC over \( \mathbb{F}_q \) if and only if
\[ \text{wt}(zL) \geq 2\delta + 1 \quad \text{for all } z \in \mathcal{I}(q, \mathcal{H}) . \]
Equivalently, \( L \) corresponds to a \((\delta, \mathcal{H})\)-ECIC over \( \mathbb{F}_q \) if and only if
\[ \text{wt} \left( \sum_{i \in K} z_i L_i \right) \geq 2\delta + 1, \]
for all \( K \in \mathcal{J}(\mathcal{H}) \) and for all choices of \( z_i \in \mathbb{F}_q^* \), \( i \in K \).

Proof: For each \( x \in \mathbb{F}_q^n \), we define
\[ B(x, \delta) = \{ y \in \mathbb{F}_q^N : y = xL + \epsilon, \ \epsilon \in \mathbb{F}_q^n, \ \text{wt}(\epsilon) \leq \delta \} , \]
the set of all vectors resulting from at most \( \delta \) errors in the transmitted vector \( x \) associated with the information vector \( x \).
Then the receiver \( R_i \) can recover \( x_{f(i)} \) correctly if and only if
\[ B(x, \delta) \cap B(x', \delta) = \emptyset , \]
for every pair \( x, x' \in \mathbb{F}_q^n \) satisfying:
\[ x_{X_i} = x'_{X_i} \quad \text{and} \quad x_{f(i)} \neq x'_{f(i)} . \]
(Observe that \( R_i \) is interested only in the bit \( x_{f(i)} \), not in the whole vector \( x \).)

Therefore, \( L \) corresponds to a \((\delta, \mathcal{H})\)-ECIC if and only if the following condition is satisfied: for all \( i \in [m] \) and for all \( x, x' \in \mathbb{F}_q^n \) such that \( x_{X_i} = x'_{X_i} \) and \( x_{f(i)} \neq x'_{f(i)} \), it holds
\[ \forall \epsilon, \epsilon' \in \mathbb{F}_q^N, \ \text{wt}(\epsilon) \leq \delta, \ \text{wt}(\epsilon') \leq \delta : \ xL + \epsilon \neq x'L + \epsilon' . \]
Denote \( z = x' - x \). Then, the condition in (4) can be reformulated as follows: for all \( i \in [n] \) and for all \( z \in \mathbb{F}_q^n \) such that \( z_{X_i} = 0 \) and \( z_{f(i)} \neq 0 \), it holds
\[ \forall \epsilon, \epsilon' \in \mathbb{F}_q^N, \ \text{wt}(\epsilon) \leq \delta, \ \text{wt}(\epsilon') \leq \delta : \ zL \neq \epsilon - \epsilon' . \]
The equivalent condition is that for all \( z \in \mathcal{I}(q, \mathcal{H}) \),
\[ \text{wt}(zL) \geq 2\delta + 1 \, . \]
Inequality (5) follows from this condition in a straight-forward manner.

Corollary 4.1: For all \( i \in [m] \), let
\[ M_i \triangleq \text{span}(\{ L_j : j \in \mathcal{Y}_i \}) . \]
Then, the matrix \( L \) corresponds to a \((\delta, \mathcal{H})\)-ECIC over \( \mathbb{F}_q \) if and only if
\[ \forall i \in [m] : \ \text{d}(L_{f(i)}, M_i) \geq 2\delta + 1 . \]

Example 4.1: Let \( q = 2, m = n = 3 \), and \( f(i) = i \) for \( i \in [3] \). Suppose \( X_1 = \{2,3\}, X_2 = \{1,3\} \), and \( X_3 = \{1,2\} \).
Let
\[ L = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} . \]
Note that \( L \) generates a \([4,3,1]_2 \) code, which has minimum distance one. However, the index code based on \( L \) can still correct one error. Indeed, let \( \mathcal{H} = \mathcal{H}(3,3, X, f) \), we have
\[ \mathcal{I}(2, \mathcal{H}) = \{100, 010, 001\} \, . \]
Since each row of \( L \) has weight at least three, it follows that \( \text{wt}(zL) \geq 3 \) for all \( z \in \mathcal{I}(2, \mathcal{H}) \).
By Lemma 4.1 \( L \) corresponds to a \((1, \mathcal{H})\)-ECIC over \( \mathbb{F}_2 \).

Example 4.2: Assume that \( m = n \) and \( f(i) = i \) for all \( i \in [m] \). Furthermore, suppose that \( X_i = \emptyset \) for all \( i \in [m] \) (i.e. there is no side information available to the receivers). Let \( \mathcal{H} = \mathcal{H}(m, n, X, f) \). Then, \( \mathcal{I}(q, \mathcal{H}) = \mathbb{F}_q^n \setminus \{0\} \). Hence, by Lemma 4.1 the \( n \times N \) matrix \( L \) corresponding to a \((\delta, \mathcal{H})\)-ECIC over \( \mathbb{F}_q \) (for some integer \( \delta \geq 0 \)) is a generating matrix of an \([N, n, \geq 2\delta + 1]_q \) linear code. Thus, the problem of designing an ECIC is reduced to the problem of constructing a classical linear error-correcting code.

V. THE \( \alpha \)-BOUND AND THE \( \kappa \)-BOUND

Let \((m, n, X, f)\) be an instance of the ICSI problem, and let \( \mathcal{H} \) be the corresponding side information hypergraph. Next, we introduce the following definitions for the hypergraph \( \mathcal{H} \).

Definition 5.1: A subset \( H \) of \([n]\) is called a generalized independent set in \( \mathcal{H} \) if every nonempty subset \( K \) of \( H \) belongs to \( \mathcal{J}(\mathcal{H}) \).

Definition 5.2: A generalized independent set of the largest size in \( \mathcal{H} \) is called a maximum generalized independent set. The size of a maximum generalized independent set in \( \mathcal{H} \) is called the generalized independence number, and denoted by \( \alpha(\mathcal{H}) \).

When \( m = n \) and \( f(i) = i \) for all \( i \in [n] \), the generalized independence number of \( \mathcal{H} \) is equal to the maximum size of an acyclic induced subgraph of \( \mathcal{G}_\mathcal{H} \), which was introduced in [3]. In particular, when \( \mathcal{G}_\mathcal{H} \) is symmetric, \( \alpha(\mathcal{H}) \) is the independence number of \( \mathcal{G}_\mathcal{H} \). We omit the proof.

Theorem 5.1 (\( \alpha \)-bound): The length of an optimal linear \((\delta, \mathcal{H})\)-ECIC over \( \mathbb{F}_q \) satisfies
\[ N_q(\mathcal{H}, \delta) \geq N_q[\alpha(\mathcal{H}), 2\delta + 1] \, . \]
Therefore, form a generator matrix of an $H$-IC. Hence, the $\alpha$ code may fail to produce an optimal linear ECIC. Using a computer search, we obtain that $\kappa_q(\alpha(H), 2\delta + 1)$.

The following proposition is based on the fact that concatenation of a $\delta$-error-correcting code with an optimal (non-error-correcting) $H$-IC yields a $\alpha$-bound.

Proposition 5.2 ($\kappa$-bound): The length of an optimal $(\delta, H)$-ECIC over $\mathbb{F}_q$ satisfies
\[ N_q(\delta, H) \leq N_q[\kappa_q(\alpha(H), 2\delta + 1)]. \]

The proof of this proposition is omitted due to lack of space.

Corollary 5.1: The length of an optimal linear $(\delta, H)$-ECIC over $\mathbb{F}_q$ satisfies
\[ N_q[\alpha(H), 2\delta + 1] \leq N_q(\delta, H) \leq N_q[\kappa_q(\alpha(H), 2\delta + 1)]. \]

Example 5.1: Let $q = 2$, $m = n = 5$, $\delta = 2$, and $f(i) = i$ for all $i \in [m]$. Assume
\[ X_1 = \{2, 5\}, \quad X_2 = \{1, 3\}, \quad X_3 = \{2, 4\}, \quad X_4 = \{3, 5\}, \quad X_5 = \{1, 4\}. \]
Let $H = H(5, 5, X, f)$. The side information graph $G_H$ of this instance is a pentagon. It is easy to verify that $\alpha(H) = \alpha(G) = 2$. It follows from Theorem 9 in [5] that $\kappa_2(H) = \text{min-rank}_2(G_H) = 3$. Thus, from (12) we have
\[ N_2[2, 5] = 8 \quad \text{and} \quad N_2[3, 5] = 10. \]
Due to Corollary 5.1 we have
\[ 8 \leq N_2(\alpha(H), 2). \]

Using a computer search, we obtain that $N_2(H, 2) = 9$, and the corresponding optimal scheme is based on
\[ L = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}. \]

It is technical to verify that by Lemma 4.1 $L$ corresponds to $(2, H)$-ECIC. The length of this ECIC lies strictly between the $\alpha$-bound and the $\kappa$-bound.

Remark 5.1: Example 5.1 illustrates that over small alphabets, the concatenation of an optimal linear (non-error-correcting) index code and an optimal linear error-correcting code may fail to produce an optimal linear ECIC. The corollary below shows that for sufficiently large alphabets, a concatenation of a classical MDS error-correcting code with an optimal (non-error-correcting) index code yields an optimal linear ECIC.

Corollary 6.1 (MDS error-correcting index code): For $q \geq 2$, $m = n = 5$, $\delta = 2$, and $f(i) = i$ for all $i \in [m]$. Assume
\[ X_1 = \{2, 5\}, \quad X_2 = \{1, 3\}, \quad X_3 = \{2, 4\}, \quad X_4 = \{3, 5\}, \quad X_5 = \{1, 4\}. \]
Let $H = H(5, 5, X, f)$. The side information graph $G_H$ of this instance is a pentagon. It is easy to verify that $\alpha(H) = \alpha(G) = 2$. It follows from Theorem 9 in [5] that $\kappa_2(H) = \text{min-rank}_2(G_H) = 3$. Thus, from (12) we have
\[ N_2[2, 5] = 8 \quad \text{and} \quad N_2[3, 5] = 10. \]

Due to Corollary 5.1 we have
\[ 8 \leq N_2(\alpha(H), 2). \]

Using a computer search, we obtain that $N_2(H, 2) = 9$, and the corresponding optimal scheme is based on
\[ L = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}. \]

It is technical to verify that by Lemma 4.1 $L$ corresponds to $(2, H)$-ECIC. The length of this ECIC lies strictly between the $\alpha$-bound and the $\kappa$-bound.

Remark 6.1: There exist hypergraph $H$, such that $G_H$ is the (symmetric) odd cycle of length $n$, for which the $\alpha$-bound is at least as good as the Singleton bound.
Remark 7.1: The bound in Theorem 7.1 implies a bound on $\kappa_q(\mathcal{H})$, which is tight for some $\mathcal{H}$. Indeed, fix $\delta = 0$. Take $m = n = 2\ell + 1$ ($\ell \geq 2$), and $f(i) = i$ for all $i \in [n]$. Let $X_1 = [n] \setminus \{1, 2, n\}$ and $X_n = [n] \setminus \{1, n, 1-n\}$. For $2 \leq i \leq n-1$, let $X_i = [n] \setminus \{i-1, i, i+1\}$. Take $\mathcal{H} = \mathcal{H}(n, n, X_1, f)$. Then $\mathcal{G}_\mathcal{H}$ is the complement of the (symmetric directed) odd cycle of length $n$. We have $|X_i| = 2\ell - 2$ for all $i \in [n]$. Then (8) becomes

$$N > 2 + \log_q(2\ell + 1).$$

If $q > 2\ell + 1$ then we obtain $N \geq 3$. Observe that in this case $\kappa_q(\mathcal{H}) = \text{min-rank}_q(\mathcal{G}_\mathcal{H}) = 3$ (see [8] Claim A.1), and thus the bound is tight.

VIII. Syndrome decoding

Consider the $(\delta, \mathcal{H})$-ECIC based on a matrix $L$. Suppose that the receiver $R_i$, $i \in [m]$, receives the vector

$$y_i = xL + \epsilon_i,$$

where $xL$ is the codeword transmitted by $S$, and $\epsilon_i$ is the error pattern affecting this codeword.

In the classical coding theory, the transmitted vector $c$, the received vector $y$, and the error pattern $e$ are related by $y = c + e$. For index coding, however, this is no longer the case. The following theorem shows that, in order to recover the message $x_{f(i)}$ from $y_i$ using (9), it is sufficient to find just one vector from a set of possible error patterns. This set is defined as follows:

$$\mathcal{L}_i(\epsilon_i) = \{\epsilon_i + z : z \in \text{span}(\{L_j\}_{j \in Y_i})\}.$$  

We henceforth refer to the set $\mathcal{L}_i(\epsilon_i)$ as the set of relevant error patterns.

Lemma 8.1: Assume that the receiver $R_i$ receives $y_i$.
1) If $R_i$ knows the message $x_{f(i)}$, then it is able to determine the set $\mathcal{L}_i(\epsilon_i)$.
2) If $R_i$ knows some vector $\hat{e} \in \mathcal{L}_i(\epsilon_i)$, then it is able to determine $x_{f(i)}$.

We now describe a syndrome decoding algorithm for linear error-correcting index codes. We have

$$y_i - x_{X_i}L_{X_i} - \epsilon_i \in \text{span}(\{L_{f(i)}\} \cup \{L_j\}_{j \in Y_i}).$$

Let $\mathcal{G}_i = \text{span}(\{L_{f(i)}\} \cup \{L_j\}_{j \in Y_i})$, and let $H^{(i)}$ be a parity check matrix of $\mathcal{G}_i$. We obtain that

$$H^{(i)}\epsilon_i^T = H^{(i)}(y_i - x_{X_i}L_{X_i})^T.$$  

Let $\beta_i$ be a column vector defined by

$$\beta_i = H^{(i)}(y_i - x_{X_i}L_{X_i})^T.$$  

Observe that each $R_i$ is capable of determining $\beta_i$. This leads us to the formulation of the decoding procedure for $R_i$ in Figure 1.

Theorem 8.2: Let $y_i = xL + \epsilon_i$ be the vector received by $R_i$, and let $\text{wt}(\epsilon_i) \leq \delta$. Assume that the procedure in Figure 1 is applied to $(y_i, x_{X_i}, L)$. Then, its output satisfies $\hat{x}_{f(i)} = x_{f(i)}$.

Remark 8.1: It is not impossible that $\hat{e} \neq \epsilon_i$. However, if $\text{wt}(\epsilon_i) \leq \delta$, it can be shown that $\hat{e} \in \mathcal{L}_i(\epsilon_i)$. Hence, by Lemma 8.1, we have $\hat{x}_{f(i)} = x_{f(i)}$.

IX. Acknowledgements

The authors would like to thank the authors of [4] for providing a preprint of their paper. This work is supported by the National Research Foundation of Singapore (Research Grant NRF-CRP2-2007-03).

REFERENCES

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

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

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

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

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

[6] S. El Rouayheb, M. A. R. Chaudhry, and A. Sprintson, “On the minimum number of transmissions in single-hop wireless coding networks,” in Proc. IEEE Inform. Theory Workshop (ITW), 2007, pp. 120–125.

[7] S. El Rouayheb, A. Sprintson, and C. Georghiades, “On the relation between the index coding and the network coding problems,” in Proc. IEEE Inform. Theory Workshop (ITW), 2007, pp. 161–168.

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

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

[10] R. Koetter and M. Médard, “An algebraic approach to network coding,” IEEE/ACM Trans. Netw., vol. 11, no. 1, pp. 782–795, 2003.

[11] S. H. Dau, V. Skachek, and Y. M. Chee, “Secure index coding with side information,” available online at http://arxiv.org/abs/1011.5566.

[12] M. Grassl, “Bounds on the minimum distance of linear codes and quantum codes,” available online at http://www.codetables.de.