Index Coding With Erroneous Side Information

Jae-Won Kim and Jong-Seon No, *Fellow, IEEE*

***Abstract*— In this paper, new index coding problems are studied, where each receiver has erroneous side information. Although side information is a crucial part of index coding, the existence of erroneous side information has not been considered yet. We study an index code with receivers that have erroneous side information symbols in the error-free broadcast channel, which is called an index code with side information errors (ICSIE). The encoding and decoding procedures of the ICSIE are proposed, based on the syndrome decoding. Then, we derive the bounds on the optimal codelength of the proposed index code with erroneous side information. Furthermore, we introduce a special graph for the proposed index coding problem, called a δ_s-cycle whose properties are similar to those of the cycle in the conventional index coding problem. Properties of the ICSIE are also discussed in the δ_s-cycle and clique. Finally, the proposed ICSIE is generalized to an index code for the scenario having both additive channel errors and side information errors, called a generalized error correcting index code.**

***Index Terms*— δ_s-cycle, error correcting index codes (ECIC), generalized error correcting index codes (GECIC), index codes (IC), index codes with side information errors (ICSIE), side information.**

I. Introduction

INDEX coding has attracted significant attention in various research areas since it was first introduced by Birk and Kol [1]. Due to its relevance to various topics in information theory, lots of research has been done on index coding even though it was first considered for the satellite communication systems. Bar-Yossef *et al.* proved that the optimal codelength of linear index codes is equal to the parameter $minrk$ of the fitting matrix of the side information graph and suggested an optimal construction method for the linear index codes [2]. In addition, it was proved that there is a nonlinear index code which outperforms the optimal linear index codes [3].

It was proved in [4] that every network coding problem can be converted to a corresponding index coding problem, and vice versa, for the given network structure. Furthermore, there was a trial relating the topological interference management (TIM) with index coding [5]. Most index coding problems assume that the sender knows the side information graph and each receiver has some subsets of messages as side information. Recently, Kao *et al.* researched a case where the sender only knows the probability distribution of side information in the receivers [6]. The index coding problems where a form of side information in each receiver is a linear combination of messages were studied [7], [8]. While most of the index coding problems are studied in an error-free broadcast channel, there has been some work on the index codes with channel errors [9]–[12]. In particular, Dau *et al.* introduced error correcting index codes (ECIC) in the erroneous broadcast channel and algebraically analyzed them [9]. There has also been research on capacity analysis and application of index codes with side information over the additive white Gaussian noise (AWGN) channel [10], [11]. Byrne and Calderini [12] studied index coding instances with channel errors and coded side information and extended the results of [9]. As we can see from the previous works, index coding problems have been generalized further and become more realistic.

In this paper, new index coding problems with erroneous side information are presented, where each receiver has erroneous side information. In the conventional index coding, it is assumed that every receiver can exploit its side information directly because there is no side information error. However, there is always a possibility of memory errors in the receivers, which causes side information errors. Since there are some instances where side information is attained by sending messages through the broadcast channel, channel errors also cause erroneous side information. Thus, we have to consider a possibility having erroneous side information in the index coding problem.

There are several applications of index coding with erroneous side information such as TIM. It is known that the interference channels of a receiver in TIM correspond to the messages, which are not the side information of the receiver in index coding. If the interference channels vary due to the moving receivers or are misunderstood as the non-interference channels from the false channel state information in TIM, this corresponds to index coding with erroneous side information. Thus, index coding with erroneous side information can be applied for the effective solutions of TIM.

In this paper, we propose the encoding and decoding procedures of index codes with side information errors (ICSIE), where each receiver has erroneous side information symbols in the error-free broadcast channel. One of the most important parameters in index coding is the codelength. Thus, the bounds on the optimal codelength of the proposed ICSIE are derived and its crucial graph, called a δ_s-cycle, similar to the cycle in the conventional index coding is proposed. We relate the conventional index codes with the ICSIE by using the proposed bound and compare the cycle of the conventional index

Manuscript received June 6, 2016; revised May 17, 2017 and August 22, 2017; accepted September 5, 2017. Date of publication September 26, 2017; date of current version November 20, 2017. This work was supported by the National Research Foundation of Korea (NRF) through the Korea government under Grant NRF-2016R1A2B2012960.

The authors are with the Department of Electrical and Computer Engineering, INMC, Seoul National University, Seoul 08826, South Korea (e-mail: kjw702@ccl.snu.ac.kr; jsno@snu.ac.kr).

Communicated by K. Narayanan, Associate Editor for Coding Techniques.
Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.
Digital Object Identifier 10.1109/TIT.2017.2756663

coding with the δ_s-cycle of the proposed ICSIE. Furthermore, it is shown that there is a similarity between the generator matrix of the ICSIE and the transpose of the parity check matrix of the error correcting code when the corresponding side information graph is a clique. Finally, the ECIC in [9] is generalized by using the proposed ICSIE, which is called a generalized error correcting index code (GECIC). That is, we consider the more general scenario, where both channel errors and side information errors exist.

The paper is organized as follows. The problem formulations for the ICSIE and the GECIC are given and the property of the generator matrix of the GECIC is derived in Section II. In Section III, the encoding and decoding procedures of the ICSIE are proposed. Then, the properties and bounds for the optimal codelength of the ICSIE are derived in Section IV. Many properties of the ECIC in [9] are generalized to those of the GECIC by using the properties of the ICSIE in Section V. Finally, conclusions are presented in Section VI.

II. Problem Formulation and Some Results

A. Notations

Let $\mathbb{F}_q$ be the finite field of size q, where q is a power of prime and $\mathbb{F}_q^* = \mathbb{F}_q \setminus \{0\}$. Let $Z[n] = \{1, 2, \ldots, n\}$ for a positive integer n. For a vector $\mathbf{x} \in \mathbb{F}_q^n$, $\text{wt}(\mathbf{x})$ denotes the Hamming weight of $\mathbf{x}$. Let $\mathbf{x}_D$ be a subvector $(x_{i_1}, x_{i_2}, \ldots, x_{i_{|D|}})$ of a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{F}_q^n$ for a subset $D = \{i_1, i_2, \ldots, i_{|D|}\} \subseteq Z[n]$, where $i_1 < i_2 < \ldots < i_{|D|}$. We also introduce a submatrix A_D of $A \in \mathbb{F}_q{}^{n\times N}$, that is, the matrix consisting of $|D|$ rows of A as

$$A_D = \begin{pmatrix} A_{i_1} \\ A_{i_2} \\ \vdots \\ A_{i_{|D|}} \end{pmatrix}$$

where A_i is the ith row of A.

B. Problem Formulation for Index Coding With Erroneous Side Information

The conventional index coding is explained before introducing index coding with erroneous side information. We consider the index coding problem, where all information packets are elements in $\mathbb{F}_q$ and each receiver just wants one packet. This scenario is considered because any index coding problems can be converted to the problem of the above scenario if the size of packets is fixed. If the certain receiver wants d packets, we can split the receiver into the d receivers, each of which wants to receive one packet with the same side information. In this scenario, we can describe the conventional index coding with side information problem as follows. There are one sender which has n information packets as $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}_q^n$ and m receivers (or users) $R_1, R_2, \ldots, R_m$, having subvectors of $\mathbf{x}$ as side information. Let $\mathcal{X}_i$ be the set of side information indices of the receiver R_i for $i \in Z[m]$. That is, each receiver R_i already knows the subvector $\mathbf{x}_{\mathcal{X}_i}$. Each receiver R_i wants to receive one element in $\mathbf{x}$, called the wanted packet denoted by $x_{f(i)}$ and it is assumed that $\{f(i)\} \cap \mathcal{X}_i = \phi$.

A side information graph shows the wanted packets and side information of all receivers. Fig. 1 shows the directed bipartite

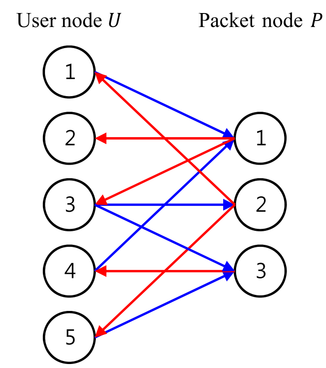

Fig. 1. An example of a directed bipartite side information graph with $n = 3$ and $m = 5$.

side information graph with five receivers and three packets. The edges from an user node to packet nodes represent the user's side information. For example, there are two edges from user 3 to packets 2 and 3 in Fig. 1, which means that user 3 has packets 2 and 3 as side information. An edge from a packet node to an user node represents that the user wants this packet. For example, there is the edge from packet 1 to user 2 in Fig. 1, which means that user 2 wants to receive packet 1. Every packet node should have at least one outgoing edge and every user node should have one incoming edge. It is assumed that the sender knows the side information graph $\mathcal{G}$ and broadcasts a codeword to receivers through the error-free channel in the conventional index coding problem.

Having side information in each receiver is a crucial part of the index coding problem. However, a possibility to have erroneous side information has not been considered yet. Now, we propose a new index coding problem with erroneous side information by changing the side information condition in the conventional index coding problem, that is, each receiver R_i has at most δ_s erroneous side information symbols. In the proposed index coding problem, a sender knows a side information graph $\mathcal{G}$ but does not know which side information is erroneous in each receiver. In addition, each receiver does not know which side information is erroneous. Furthermore, we can also consider additive channel errors in each receiver as in [9]. That is, each receiver receives $\mathbf{y} + \boldsymbol{\epsilon}_i$, where $\mathbf{y}$ is a codeword and $\boldsymbol{\epsilon}_i$ is an additive error vector such that $\mathbf{y}, \boldsymbol{\epsilon}_i \in \mathbb{F}_q^N$ and $\text{wt}(\boldsymbol{\epsilon}_i) \leq \delta_c$.

Definition 1: A generalized error correcting index code with parameters $(\delta_s, \delta_c, \mathcal{G})$ over $\mathbb{F}_q$, denoted by a $(\delta_s, \delta_c, \mathcal{G})$-GECIC is a set of codewords having:

1) An encoding function $E : \mathbb{F}_q^n \to \mathbb{F}_q^N$.
2) A set of decoding functions $D_1, D_2, \ldots, D_m$ such that $D_i : \mathbb{F}_q^N \times \mathbb{F}_q^{|\mathcal{X}_i|} \to \mathbb{F}_q$ satisfying

$$D_i(E(\mathbf{x}) + \boldsymbol{\epsilon}_i, \hat{\mathbf{x}}_{\mathcal{X}_i}) = x_{f(i)}$$

for all $i \in Z[m]$, $\mathbf{x} \in \mathbb{F}_q^n$, $\boldsymbol{\epsilon}_i \in \mathbb{F}_q^N$ with $\text{wt}(\boldsymbol{\epsilon}_i) \leq \delta_c$, and $\text{wt}(\mathbf{x}_{\mathcal{X}_i} - \hat{\mathbf{x}}_{\mathcal{X}_i}) \leq \delta_s$, where $\hat{\mathbf{x}}_{\mathcal{X}_i}$ is the erroneous side information vector of the receiver R_i.

Here, $\text{wt}(\mathbf{x}_{\mathcal{X}_i} - \hat{\mathbf{x}}_{\mathcal{X}_i}) \leq \delta_s$ means that the maximum number of side information errors is δ_s for each receiver. In this paper, we consider a linear index code. That is, $E(\mathbf{x}) = \mathbf{x}G$ for all $\mathbf{x} \in \mathbb{F}_q^n$, where $G \in \mathbb{F}_q^{n\times N}$ is a generator matrix of the index code and N denotes the codelength of the $(\delta_s, \delta_c, \mathcal{G})$-GECIC. Let $N_{\text{opt}}^q(\delta_s, \delta_c, \mathcal{G})$ be the optimal codelength of the $(\delta_s, \delta_c, \mathcal{G})$-GECIC over $\mathbb{F}_q$. If $\delta_c = 0$, a $(\delta_s, 0, \mathcal{G})$-GECIC is called a $(\delta_s, \mathcal{G})$-index code with side information errors, denoted by a $(\delta_s, \mathcal{G})$-ICSIE and the optimal codelength $N_{\text{opt}}^q(\delta_s, \delta_c, \mathcal{G})$ is modified to $N_{\text{opt}}^q(\delta_s, \mathcal{G})$. Similarly, for $\delta_s = 0$, we have $(\delta_c, \mathcal{G})$-ECIC and $N_{\text{opt}}^q(\delta_c, \mathcal{G})$ as in [9].

C. Property of Generator Matrix of GECIC

We find the property of the generator matrix for the proposed $(\delta_s, \delta_c, \mathcal{G})$-GECIC by generalizing [9, Lemma 3.8]. Let $\mathcal{I}(q, \mathcal{G}, \delta_s)$ be the set of vectors defined by

$$\mathcal{I}(q, \mathcal{G}, \delta_s) = \bigcup_{i \in Z[m]} \mathcal{I}_i(q, \mathcal{G}, \delta_s)$$

where $\mathcal{I}_i(q, \mathcal{G}, \delta_s) = \{\mathbf{z} \in \mathbb{F}_q^n : \text{wt}(\mathbf{z}_{\mathcal{X}_i}) \le 2\delta_s, z_{f(i)} \neq 0\}$. The support set of $\mathcal{I}(q, \mathcal{G}, \delta_s)$ is defined as

$$J(\mathcal{G}, \delta_s) = \bigcup_{i \in Z[m]} \left\{\{f(i)\} \cup Y_i \cup I_i : Y_i \subseteq \mathcal{Y}_i, I_i \subseteq \mathcal{X}_i\right\}$$

where $|I_i| \le 2\delta_s$ and $\mathcal{Y}_i = Z[n] \setminus \left(\{f(i)\} \cup \mathcal{X}_i\right)$.

Then, the property of the generator matrix of the $(\delta_s, \delta_c, \mathcal{G})$-GECIC is given in the following theorem.

Theorem 1 (Generalization of [9, Lemma 3.8]): A matrix G is a generator matrix of the $(\delta_s, \delta_c, \mathcal{G})$-GECIC if and only if

$$\text{wt}(\mathbf{z}G) \ge 2\delta_c + 1, \text{ for all } \mathbf{z} \in \mathcal{I}(q, \mathcal{G}, \delta_s).$$

Proof: It will be proved in the similar manner as in [9], that is, we use the similar concepts as Hamming spheres of the classical error correcting codes. Here, we find the set of message vectors which should be distinguished by the received codewords. Let $B(\mathbf{x}, \delta_c)$ be the Hamming sphere of the received codeword $\mathbf{y}$ defined by

$$B(\mathbf{x}, \delta_c) = \{\mathbf{y} \in \mathbb{F}_q^N : \mathbf{y} = \mathbf{x}G + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \in \mathbb{F}_q^N, \text{and } \text{wt}(\boldsymbol{\epsilon}) \le \delta_c\}.$$

First, we prove that the receiver R_i can recover the wanted packet $x_{f(i)}$ if and only if

$$B(\mathbf{x}, \delta_c) \cap B(\mathbf{x}', \delta_c) = \phi \tag{1}$$

for every pair $\mathbf{x}$ and $\mathbf{x}' \in \mathbb{F}_q^n$ such that $x_{f(i)} \neq x'_{f(i)}$ and $\text{wt}(\mathbf{x}_{\mathcal{X}_i} - \mathbf{x}'_{\mathcal{X}_i}) \le 2\delta_s$.

The only difference from [9, Lemma 3.8] is that we replace the condition $\mathbf{x}_{\mathcal{X}_i} = \mathbf{x}'_{\mathcal{X}_i}$ with $\text{wt}(\mathbf{x}_{\mathcal{X}_i} - \mathbf{x}'_{\mathcal{X}_i}) \le 2\delta_s$. The side information vectors $\mathbf{x}_{\mathcal{X}_i}$ and $\mathbf{x}'_{\mathcal{X}_i}$ of the receiver R_i satisfy the inequality $\text{wt}(\mathbf{x}_{\mathcal{X}_i} - \mathbf{x}'_{\mathcal{X}_i}) \le 2\delta_s$ if they can be the same by changing at most δ_s side information symbols, respectively. Now, we prove the above statement as follows.

Necessity: R_i has to recover $x_{f(i)}$ by using side information and the received codeword. Suppose that there are two vectors $\mathbf{x}$ and $\mathbf{x}' \in \mathbb{F}_q^n$ such that $x_{f(i)} \neq x'_{f(i)}$. Since R_i can always recover $x_{f(i)}$, R_i has to distinguish such $\mathbf{x}$ and $\mathbf{x}'$. If $\text{wt}(\mathbf{x}_{\mathcal{X}_i} - \mathbf{x}'_{\mathcal{X}_i}) \ge 2\delta_s + 1$, R_i can distinguish $\mathbf{x}$ and $\mathbf{x}'$ from the side information. However, if $\text{wt}(\mathbf{x}_{\mathcal{X}_i} - \mathbf{x}'_{\mathcal{X}_i}) \le 2\delta_s$, R_i cannot distinguish $\mathbf{x}$ and $\mathbf{x}'$ from the side information and thus the received codewords of $\mathbf{x}$ and $\mathbf{x}'$ should be distinguished, which corresponds to (1).

Sufficiency: For $\mathbf{x}$ and $\mathbf{x}' \in \mathbb{F}_q^n$, consider the cases where $x_{f(i)} \neq x'_{f(i)}$. For these cases, it is easy to note that the sufficiency also holds. Thus, we only have to consider the remaining cases such that $x_{f(i)} = x'_{f(i)}$ to prove the sufficiency. In fact, for such $\mathbf{x}$ and $\mathbf{x}'$, R_i does not need to distinguish $\mathbf{x}$ and $\mathbf{x}'$ because $x_{f(i)} = x'_{f(i)}$ and R_i is only interested in $x_{f(i)}$. Thus, there is no restriction on such $\mathbf{x}$ and $\mathbf{x}'$ for R_i to recover $x_{f(i)}$.

Let $\mathbf{z} = \mathbf{x} - \mathbf{x}'$. Since each receiver R_i has to recover $x_{f(i)}$, (1) should be satisfied for all $i \in Z[m]$. That is, the matrix G corresponds to the generator matrix of the $(\delta_s, \delta_c, \mathcal{G})$-GECIC if and only if $\text{wt}(\mathbf{z}G) \ge 2\delta_c + 1$ for all $\mathbf{z} \in \mathcal{I}(q, \mathcal{G}, \delta_s)$. ■

Here are several remarks regarding the above theorem.

Remark 1: It is obvious that G is a generator matrix of the $(\delta_s, \delta_c, \mathcal{G})$-GECIC if and only if

$$\text{wt}(\Sigma_{i \in K} z_i G_i) \ge 2\delta_c + 1$$

for all $K \in J(\mathcal{G}, \delta_s)$ and all $z_i \in \mathbb{F}_q^*$.

Remark 2: Since the receiver R_i is only interested in $x_{f(i)}$, it is possible to have $B(\mathbf{x}, \delta_c) \cap B(\mathbf{x}', \delta_c) \neq \phi$ for $\text{wt}(\mathbf{x}_{\mathcal{X}_i} - \mathbf{x}'_{\mathcal{X}_i}) \le 2\delta_s$ and $x_{f(i)} = x'_{f(i)}$. It means that R_i does not need to distinguish $\mathbf{x}$ and $\mathbf{x}'$ because $x_{f(i)} = x'_{f(i)}$.

Remark 3: For a $(\delta_s, \mathcal{G})$-ICSIE, the inequality $\text{wt}(\mathbf{z}G) \ge 2\delta_c + 1$ becomes $\mathbf{z}G \neq \mathbf{0}$. If the side information is assumed to be erased, we have

$$\mathcal{I}_i(q, \mathcal{G}, \delta_s) = \{\mathbf{z} \in \mathbb{F}_q^n : \text{wt}(\mathbf{z}_{\mathcal{X}_i}) \le \delta_s, z_{f(i)} \neq 0\}.$$

We provide an example for the aforementioned theorem.

Example 1: Let $q = 2$, $m = n = 4$, $\delta_s = 1$, $f(i) = i$, and $\mathcal{X}_i = Z[4] \backslash \{i\}$ for all $i \in Z[4]$. In general, the uncoded case is the worst case in the error-free channel, that is, the codelength is $n = 4$. However, we can construct a $(\delta_s = 1, \mathcal{G})$-ICSIE with codelength 3. From Theorem 1, it is clear that $\mathcal{I}(q, \mathcal{G}, \delta_s = 1)$ includes all vectors in $\mathbb{F}_2^4$ except $(1, 1, 1, 1)$ and $(0, 0, 0, 0)$. Assume that we have a 4×3 matrix G as

$$G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Then, we have $\mathbf{z}G \neq \mathbf{0}$ for all $\mathbf{z} \in \mathcal{I}(q, \mathcal{G}, \delta_s = 1)$. Thus, the above matrix G is a generator matrix of the $(\delta_s = 1, \mathcal{G})$-ICSIE.

From [9, Lemma 8.2] and Theorem 1, a relation between an ECIC and a GECIC can be deduced. In [9], $\Gamma = \{(m, n, \mathcal{X}, f)\}$ is defined as a set of index coding instances and an ECIC with δ_c is said to be static under the set Γ if it satisfies all instances in Γ. Now, we define Γ_{δ_s} for a side information graph $\mathcal{G}$ as a set of all instances, each of which is constructed by deleting $\min(2\delta_s, |\mathcal{X}_i|)$ outgoing edges from each receiver R_i in $\mathcal{G}$. Then, we have the following lemma.

Lemma 1: For a side information graph $\mathcal{G}$, an ECIC with δ_c static under Γ_{δ_s} is equivalent to a $(\delta_s, \delta_c, \mathcal{G})$-GECIC.

Proof: Let $\mathcal{I}(\Gamma_{\delta_s}) = \cup_{\bar{\mathcal{G}}} \mathcal{I}(q, \bar{\mathcal{G}}, 0)$, where $\bar{\mathcal{G}}$ is constructed by deleting $\min(2\delta_s, |\mathcal{X}_i|)$ outgoing edges from each receiver R_i in $\mathcal{G}$. From [9, Lemma 8.2] and Theorem 1, two problems are equivalent if $\mathcal{I}(\Gamma_{\delta_s}) = \mathcal{I}(q, \mathcal{G}, \delta_s)$. Since each $\bar{\mathcal{G}}$ represents the vector $\mathbf{z} \in \mathcal{I}_i(q, \mathcal{G}, \delta_s)$ by selecting at most $2\delta_s$ non-zero elements of $\mathbf{z}_{\mathcal{X}_i}$, we have $\mathcal{I}(\Gamma_{\delta_s}) = \mathcal{I}(q, \mathcal{G}, \delta_s)$. ■

From Lemma 1, we can further infer that a GECIC is equivalent to an ECIC for a modified index coding instance. First, we modify $\mathcal{G}$ to $\tilde{\mathcal{G}}$, where there are $\binom{|\mathcal{X}_i|}{2\delta_s}$ corresponding receivers of R_i in $\mathcal{G}$ wanting the same $x_{f(i)}$ with $\mathcal{X}_i^* = \mathcal{X}_i \setminus I_i$ such that $I_i \subseteq \mathcal{X}_i$ and $|I_i| = 2\delta_s$ in $\tilde{\mathcal{G}}$. Then, we have the following lemma.

Lemma 2: For a side information graph $\mathcal{G}$, a $(\delta_c, \tilde{\mathcal{G}})$-ECIC is equivalent to a $(\delta_s, \delta_c, \mathcal{G})$-GECIC.

Proof: We can easily notice that a $(\delta_c, \tilde{\mathcal{G}})$-ECIC is equivalent to an ECIC with δ_c static under Γ_{δ_s} and thus to a $(\delta_s, \delta_c, \mathcal{G})$-GECIC from Lemma 1. ■

III. Encoding and Decoding of $(\delta_s, \mathcal{G})$-ICSIE

In this section, we propose the encoding and decoding procedures of the proposed index code with erroneous side information in the error-free broadcast channel, that is, the ICSIE.

A. Encoding Procedure

In general, design of the index codes is to find a generator matrix with the minimum codelength for the given side information graph, called the optimal index codes. In fact, any linearly dependent equations of the message packets can be generated by their minimum set of linearly independent equations. Thus, design of the optimal $(\delta_s, \mathcal{G})$-ICSIE corresponds to finding the minimum number of linearly independent equations of message packets, whose generator matrix satisfies the property in Theorem 1 with no channel error. Thus, we have the following remark for the $(\delta_s, \mathcal{G})$-ICSIE.

Remark 4: If a generator matrix G of the $(\delta_s, \mathcal{G})$-ICSIE has rank less than or equal to the codelength N, the matrix deleting any dependent columns from G can also be its generator matrix. Thus, the generator matrix G_{opt} of the optimal $(\delta_s, \mathcal{G})$-ICSIE should have the rank $N_{\text{opt}}^q(\delta_s, \mathcal{G})$.

We propose the optimal construction method of the $(\delta_s, \mathcal{G})$-ICSIE, which is similar to that of the conventional index code in [2]. First, we generalize two well known definitions, the fitting matrices and their minimum rank for the given side information graph in [2]. From Lemma 2, we can find a generalized fitting matrix for $\mathcal{G}$ through a fitting matrix for $\tilde{\mathcal{G}}$ as in the following definition.

Definition 2: Let $T_i = \{\mathbf{i}_1, \ldots, \mathbf{i}_{\binom{|\mathcal{X}_i|}{2\delta_s}}\}$ for $i \in Z[m]$, where $\mathbf{i}_j$ denotes the set of chosen indices from $\mathcal{X}_i$ with cardinality $2\delta_s$ for $|\mathcal{X}_i| \geq 2\delta_s$ and otherwise, $T_i = \{\mathbf{i}_1\} = \{\mathcal{X}_i\}$. An $n \times \Sigma_{i \in Z[m]} \binom{|\mathcal{X}_i|}{2\delta_s}$ matrix A_g is said to be a generalized fitting matrix for $\mathcal{G}$ if A_g satisfies the followings:

1) $A_g = [A_{ab}^{(i)}]$ consists of m disjoint $n \times \binom{|\mathcal{X}_i|}{2\delta_s}$ submatrices $A^{(i)}$ for $i \in Z[m]$.
2) For $i \in Z[m]$ and $b \in Z\left[\binom{|\mathcal{X}_i|}{2\delta_s}\right]$, $A_{ab}^{(i)} = 0$ for $a \in \mathbf{i}_b$ and $A_{ab}^{(i)}$ can take any value of $\mathbb{F}_q$ for $a \in \mathcal{X}_i \setminus \mathbf{i}_b$.
3) $A_{f(i)b}^{(i)} = 1$ for $b \in Z\left[\binom{|\mathcal{X}_i|}{2\delta_s}\right]$ and $i \in Z[m]$.
4) $A_{ab}^{(i)} = 0$ for $a \in \mathcal{Y}_i$, $b \in Z\left[\binom{|\mathcal{X}_i|}{2\delta_s}\right]$, and $i \in Z[m]$.

Definition 3: $minrk_q(\delta_s, \mathcal{G}) = \min\{rk_q(A_g) : A_g \text{ fits } \mathcal{G}\}$, where $rk_q(A_g)$ denotes the rank of A_g over $\mathbb{F}_q$.

By using the generalized fitting matrices and their minimum rank, the optimal codelength of the proposed ICSIE can be given in the following theorem, which corresponds to generalization of [2, Th. 1] for the conventional index coding with $\delta_s = 0$.

Theorem 2 (Generalization of [2, Th. 1]): $N_{\text{opt}}^q(\delta_s, \mathcal{G}) = minrk_q(\delta_s, \mathcal{G})$.

Proof: Since the $(\delta_s, \mathcal{G})$-ICSIE is equivalent to the conventional index code with $\delta_s = 0$ for a modified side information graph $\tilde{\mathcal{G}}$ from Lemma 2, a generator matrix of the conventional index code for $\tilde{\mathcal{G}}$ satisfies the condition of the $(\delta_s, \mathcal{G})$-ICSIE in Theorem 1 and vice versa. Thus, A_g can be a generator matrix of the $(\delta_s, \mathcal{G})$-ICSIE because A_g is the conventional fitting matrix for $\tilde{\mathcal{G}}$. Since A_g can be a generator matrix of the $(\delta_s, \mathcal{G})$-ICSIE and $N_{\text{opt}}^q(\delta_s = 0, \tilde{\mathcal{G}}) = minrk_q(\delta_s, \mathcal{G})$ from [9, Lemma 3.5], we have $N_{\text{opt}}^q(\delta_s, \mathcal{G}) = minrk_q(\delta_s, \mathcal{G})$. ■

Thus, the optimal generator matrix of the $(\delta_s, \mathcal{G})$-ICSIE can be constructed from the generalized fitting matrix whose rank is $minrk_q(\delta_s, \mathcal{G})$ by removing all dependent columns. There is an example for construction of a generator matrix in Example 1 as follows.

Example 2: Let $q = 2$, $m = n = 4$, $\delta_s = 1$, $f(i) = i$, and $\mathcal{X}_i = Z[4] \setminus \{i\}$ for all $i \in Z[4]$. From Theorem 2, $N_{\text{opt}}^q(\delta_s, \mathcal{G})$ can be found by $minrk_q(\delta_s, \mathcal{G})$. A matrix A_g which fits $\mathcal{G}$ is described as

$$A_g = \begin{pmatrix} A^{(1)} & A^{(2)} & A^{(3)} & A^{(4)} \end{pmatrix}$$

where

$$A^{(1)} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & * \\ 0 & * & 0 \\ * & 0 & 0 \end{pmatrix}, \quad A^{(2)} = \begin{pmatrix} 0 & 0 & * \\ 1 & 1 & 1 \\ 0 & * & 0 \\ * & 0 & 0 \end{pmatrix}$$

$$A^{(3)} = \begin{pmatrix} 0 & 0 & * \\ 0 & * & 0 \\ 1 & 1 & 1 \\ * & 0 & 0 \end{pmatrix}, \quad A^{(4)} = \begin{pmatrix} 0 & 0 & * \\ 0 & * & 0 \\ * & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

and $*$ denotes any value of 0 or 1. In order to minimize the rank of A_g, the value 0 or 1 is selected for $*$ in A_g and the dependent columns are removed. Then, one of the optimal generator matrices G_{opt} is given as

$$G_{\text{opt}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

B. Decoding Procedure

We propose the decoding procedure of the $(\delta_s, \mathcal{G})$-ICSIE similar to that of the ECIC in [9]. That is, we can consider the decoding procedure which is similar to the syndrome decoding of the classical linear error correcting code. First, we find the syndrome related to the side information errors, which is used to find the correct side information. This procedure is different from that of the ECIC because the side information errors exist in the proposed ICSIE.

In order to introduce the decoding procedure, we assume the followings:

1) Each receiver receives a codeword $\mathbf{y} = \mathbf{x}G$ through the error-free channel.

2) The receiver R_i has a side information vector $\hat{\mathbf{x}}_{\mathcal{X}_i}$ for $i \in Z[m]$, where the number of erroneous side information symbols is less than or equal to δ_s.
3) The receiver R_i only wants to recover $x_{f(i)}$ for $i \in Z[m]$.

In addition, we define the following notations:

1) $\tilde{\mathbf{x}}_{\delta_s} = \mathbf{x}_{\mathcal{X}_i} - \hat{\mathbf{x}}_{\mathcal{X}_i}$, where $\mathbf{x}_{\mathcal{X}_i}$ is a correct side information vector of R_i.
2) $H^{(i)}$ is a matrix whose rows form a basis of the dual of $\text{span}(\{G_j\}_{j \in \{f(i)\} \cup \mathcal{Y}_i})$.
3) $H_e^{(i)}$ is a matrix whose rows form a basis of the dual of $\text{span}(\{G_j\}_{j \in \mathcal{Y}_i})$.

Algorithm 1 Decoding Procedure for R_i

Input: $\mathbf{y}$, $\hat{\mathbf{x}}_{\mathcal{X}_i}$, and G

Output: $x_{f(i)}$

Step 1) Compute the syndrome

$$\mathbf{s}_i = H^{(i)}(\mathbf{y} - \hat{\mathbf{x}}_{\mathcal{X}_i} G_{\mathcal{X}_i})^{\mathsf{T}} = H^{(i)}(\tilde{\mathbf{x}}_{\delta_s} G_{\mathcal{X}_i})^{\mathsf{T}}. \quad (2)$$

Step 2) Find one solution $\mathbf{p}_i$ that satisfies $\mathbf{s}_i = H^{(i)} \mathbf{p}_i{}^{\mathsf{T}}$ under the condition that $\mathbf{p}_i$ is a linear combination of rows of $G_{\mathcal{X}_i}$, where the number of linearly combined rows in $G_{\mathcal{X}_i}$ is less than or equal to δ_s.

Step 3) Make the following equation

$$\tilde{\mathbf{y}} = \mathbf{y} - \hat{\mathbf{x}}_{\mathcal{X}_i} G_{\mathcal{X}_i} - \mathbf{p}_i = x_{f(i)} G_{f(i)} + (\mathbf{x}_{\mathcal{Y}_i} - \mathbf{b}) G_{\mathcal{Y}_i} \quad (3)$$

where $\mathbf{b} \in \mathbb{F}_q^{|\mathcal{Y}_i|}$.

Step 4) Find $x_{f(i)}$ by multiplying the matrix $H_e^{(i)\mathsf{T}}$ on both sides of (3).

Remark 5: If we choose $H_e^{(i)}$ so that its rows span the orthogonal complement of $\text{span}(\{G_j\}_{j \in \mathcal{Y}_i})$, then automatically, rows of $H_e^{(i)}$ do not span the orthogonal complement of $\text{span}(G_{f(i)})$ because $G_{f(i)}$ does not belong to $\text{span}(\{G_j\}_{j \in \mathcal{Y}_i})$ in the case of the ICSIE by Theorem 1.

Then, the decoding procedure of the $(\delta_s, \mathcal{G})$-ICSIE for each receiver R_i is described in Algorithm 1. We need the following theorem for $\mathbf{p}_i$ in (3) of the proposed decoding procedure.

Theorem 3: Let η_i be a subset of $\mathcal{X}_i$ with $|\eta_i| \le \delta_s$. Let $\mathbf{p}_i$ be a solution of $\mathbf{s}_i = H^{(i)} \mathbf{p}_i$, where $\mathbf{p}_i \in \text{span}(\{G_j\}_{j \in \eta_i})$. Then, $\mathbf{p}_i$ is given as $\mathbf{p}_i = \tilde{\mathbf{x}}_{\delta_s} G_{\mathcal{X}_i} + \mathbf{k}$, where $\mathbf{k} \in \text{span}(\{G_j\}_{j \in \mathcal{Y}_i})$.

Proof: We can find a solution $\mathbf{p}_i$ for $\mathbf{s}_i = H^{(i)} \mathbf{p}_i^{\mathsf{T}}$, under the condition that $\mathbf{p}_i$ is a linear combination of rows of $G_{\mathcal{X}_i}$, where the number of linearly combined rows in $G_{\mathcal{X}_i}$ is less than or equal to δ_s because there exist at least one such $\mathbf{p}_i$ due to $\tilde{\mathbf{x}}_{\delta_s} G_{\mathcal{X}_i}$. Moreover, if we find a solution $\mathbf{p}_i$ under the condition mentioned above, then we have

$$\mathbf{p}_i = \tilde{\mathbf{x}}_{\delta_s} G_{\mathcal{X}_i} + \mathbf{k}$$

due to the property of the generator matrix. Specifically, from

$$\mathbf{s}_i = H^{(i)}(\tilde{\mathbf{x}}_{\delta_s} G_{\mathcal{X}_i})^{\mathsf{T}} = H^{(i)} \mathbf{p}_i^{\mathsf{T}}$$

we have $H^{(i)}(\tilde{\mathbf{x}}_{\delta_s} G_{\mathcal{X}_i} - \mathbf{p}_i)^{\mathsf{T}} = 0$. Thus, $\tilde{\mathbf{x}}_{\delta_s} G_{\mathcal{X}_i} - \mathbf{p}_i = a G_{f(i)} - \mathbf{b} G_{\mathcal{Y}_i}$ and

$$a G_{f(i)} - \mathbf{b} G_{\mathcal{Y}_i} - \tilde{\mathbf{x}}_{\delta_s} G_{\mathcal{X}_i} + \mathbf{p}_i = \mathbf{0} \quad (4)$$

where $a \in \mathbb{F}_q$ and $\mathbf{b} \in \mathbb{F}_q^{|\mathcal{Y}_i|}$. Then, we can easily check that LHS of (4) is $\mathbf{x}G$ such that $\text{wt}(\mathbf{x}_{\mathcal{X}_i}) \le 2\delta_s$. Since RHS of (4) is zero, a should be zero by Theorem 1. Therefore, $\mathbf{p}_i = \tilde{\mathbf{x}}_{\delta_s} G_{\mathcal{X}_i} + \mathbf{b} G_{\mathcal{Y}_i}$. ■

Using Theorem 3, (3) can be given as $\tilde{\mathbf{y}} = x_{f(i)} G_{f(i)} + (\mathbf{x}_{\mathcal{Y}_i} - \mathbf{b}) G_{\mathcal{Y}_i}$ and from Step 4, we have $\tilde{\mathbf{y}} H_e^{(i)\mathsf{T}} = x_{f(i)} G_{f(i)} H_e^{(i)\mathsf{T}}$ due to Remark 5. Thus, $x_{f(i)}$ can easily be obtained.

Remark 6: An interesting fact of this decoding procedure is that we can decode $x_{f(i)}$ even if we do not know the exact $\tilde{\mathbf{x}}_{\delta_s} G_{\mathcal{X}_i}$ as in the following example.

Example 3: Let $q = 2, m = n = 9, \delta_s = 1$, and $f(i) = i$ for $i \in Z[9]$. Suppose that $\mathcal{X}_i = Z[9] \setminus \{i\}$ for $i \in Z[8]$ and $\mathcal{X}_9 = \{2, 3, 5, 6, 7, 8\}$. It is easy to check that one of the possible generator matrices of the above setting is given as

$$G = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

For the message vector $\mathbf{x} = (1, 1, 1, 1, 0, 0, 0, 0, 1)$, we have the received codeword $\mathbf{y} = \mathbf{x}G = (0, 1, 1, 0, 1, 0)$ in the error-free channel. In this case, we focus on the decoding procedure of the receiver R_9. We assume that $\hat{\mathbf{x}}_{\mathcal{X}_9} = (1, 1, 0, 0, 0, 1)$. That is, the receiver R_9 has erroneous side information $\hat{x}_8$.

The decoding procedure is described as follows:

1) We compute $\mathbf{y} - \hat{\mathbf{x}}_{\mathcal{X}_9} G_{\mathcal{X}_9} = (1, 1, 1, 0, 1, 1)$.
2) Then, we can make $H^{(9)}$ from $G_{\{9\} \cup \mathcal{Y}_9}$ as

$$H^{(9)} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

3) Also, we can make $H_e^{(9)}$ as

$$H_e^{(9)} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

4) Compute the syndrome as

$$\mathbf{s}_9 = H^{(9)}(\mathbf{y} - \hat{\mathbf{x}}_{\mathcal{X}_9} G_{\mathcal{X}_9})^{\mathsf{T}} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

5) Find a solution $\mathbf{p}_9$ for $H^{(9)} \mathbf{p}_9^{\mathsf{T}} = \mathbf{s}_9$ under the decoding condition.

Then, we have $\mathbf{p}_9 = (0, 0, 0, 1, 1, 1)$ and $(0, 0, 1, 1, 1, 0)$. In fact, we need just one of two solutions. Choosing the first solution for $\mathbf{p}_9$ means that the receiver R_9 decides $\hat{x}_8$ as the erroneous side information while choosing the second solution means that $\hat{x}_7$ is decided as the erroneous side information.

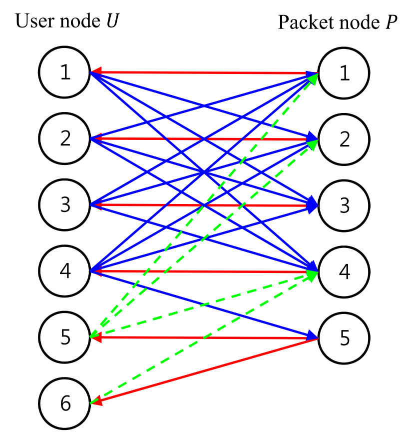

Fig. 2. A side information graph $\mathcal{G}$ in Example 4 for description of a δ_s-cycle.

6) If $\hat{x}_8$ is chosen as the erroneous side information, then (3) in Algorithm 1 becomes

$$\mathbf{y} - \hat{\mathbf{x}}_{\mathcal{X}_9} G_{\mathcal{X}_9} - (0,0,0,1,1,1) = (1,1,1,1,0,0).$$
$$(1,1,1,1,0,0) = x_9 G_9 + (\mathbf{x}_{\mathcal{Y}_i} - \mathbf{b}) G_{\mathcal{Y}_9}.$$

Multiplying $H_e^{(9)\intercal}$ on both sides leads to $(1,0,0,0) = x_9(1,0,0,0)$. Thus, $x_9 = 1$, which is the correct value.

7) If $\hat{x}_7$ is chosen as the erroneous side information, then (3) in Algorithm 1 becomes

$$\mathbf{y} - \hat{\mathbf{x}}_{\mathcal{X}_9} G_{\mathcal{X}_9} - (0,0,1,1,1,0) = (1,1,0,1,0,1).$$
$$(1,1,0,1,0,1) = x_9 G_9 + (\mathbf{x}_{\mathcal{Y}_i} - \mathbf{b}) G_{\mathcal{Y}_9}.$$

Multiplying $H_e^{(9)\intercal}$ on both sides leads to $(1,0,0,0) = x_9(1,0,0,0)$. Thus, $x_9 = 1$, which is also the correct value.

IV. Properties and Bounds for Codelength of $(\delta_s, \mathcal{G})$-ICSIE

In this section, we introduce a new type of graphs, called a δ_s-cycle for encoding of the index codes and derive some bounds for the optimal codelength of the $(\delta_s, \mathcal{G})$-ICSIE.

A. δ_s-Cycle

First, we define a δ_s-cycle in $\mathcal{G}$ and generalize the generalized independent set and the generalized independence number of $\mathcal{G}$ in [9]. Let Φ be the set of subsets of $Z[n]$ defined by

$$\Phi = \{B \subseteq Z[n] \,\big|\, |\mathcal{X}_i \cap B| \geq 2\delta_s + 1 \text{ for all } i \in Z[m] \text{ s.t. } f(i) \in B\}$$

for a side information graph $\mathcal{G}$ of the $(\delta_s, \delta_c, \mathcal{G})$-GECIC.

Definition 4: A subgraph $\mathcal{G}'$ of $\mathcal{G}$ is called a δ_s-cycle if the set of packet node indices of $\mathcal{G}'$ is an element of Φ (say B) and the set of user node indices of $\mathcal{G}'$ consists of $i \in Z[m]$ such that $f(i) \in B$ and its edges consist of the corresponding edges in $\mathcal{G}$. The graph $\mathcal{G}$ is said to be δ_s-acyclic if there is no δ_s-cycle in $\mathcal{G}$.

There is an example for a δ_s-cycle in $\mathcal{G}$ as follows.

Example 4: Let $\delta_s = 1$ and a side information graph $\mathcal{G}$ is given in Fig. 2. Then, a subgraph $\mathcal{G}'$ which consists of users 1 to 4, packets 1 to 4, and corresponding solid edges is a δ_s-cycle because $B = \{1,2,3,4\} \in \Phi$. However, $\mathcal{G}$ is not a δ_s-cycle because R_6 only has x_5 as side information so that $|\mathcal{X}_6 \cap \{1,2,3,4,5\}| = 1 < 3$. Although $|\mathcal{X}_5 \cap \{1,2,3,4,5\}| = 3$, $\mathcal{G}$ cannot be a δ_s-cycle because of R_6.

Definition 5: A set of packet node indices of a δ_s-cycle is called a δ_s-cycle induced set. Two δ_s-cycles are said to be disjoint if their δ_s-cycle induced sets are disjoint.

Definition 6 (Generalization of [9, Definition 4.1]): A subset Q of $Z[n]$ is called a δ_s-generalized independent set in $\mathcal{G}$ if every nonempty subset K of Q belongs to $J(\mathcal{G}, \delta_s)$.

Definition 7 (Generalization of [9, Definition 4.2]): Let $\gamma(\mathcal{G})$ be the largest size of a δ_s-generalized independent set in $\mathcal{G}$, which is called the δ_s-generalized independence number.

In the following lemma, we will show the relationship between a δ_s-generalized independent set and a δ_s-acyclic graph $\mathcal{G}$.

Lemma 3: Let $Z[n]$ be a set of the packet node indices in $\mathcal{G}$. Then, $Z[n]$ is a δ_s-generalized independent set if and only if the side information graph $\mathcal{G}$ is δ_s-acyclic.

Proof:

Necessity: Let $\mathcal{G}'$ be a subgraph of $\mathcal{G}$, where packet nodes of $\mathcal{G}'$ is a subset of packet nodes of $\mathcal{G}$ and the sets of user nodes and edges of $\mathcal{G}'$ are determined by $\mathcal{G}$. Suppose that $\mathcal{G}$ is δ_s-cyclic. Then, we can make a subgraph $\mathcal{G}'$ such that each receiver has side information symbols whose number is larger than or equal to $2\delta_s + 1$. Let $Q \subseteq Z[n]$ be the set of packet node indices of $\mathcal{G}'$. Then, $Q \notin J(\mathcal{G}', \delta_s)$ and Q is not also included in $J(\mathcal{G}, \delta_s)$. It contradicts the assumption and thus $\mathcal{G}$ is δ_s-acyclic.

Sufficiency: If $\mathcal{G}$ is δ_s-acyclic, every subgraph $\mathcal{G}'$ is also δ_s-acyclic. That is, every nonempty subset $Q \subseteq Z[n]$ belongs to $J(\mathcal{G}, \delta_s)$ because there exists at least one receiver, whose number of side information symbols is less than or equal to $2\delta_s$. Specifically, there is at least one receiver R_i in $\mathcal{G}$ whose number of side information symbols is less than or equal to $2\delta_s$ because $\mathcal{G}$ is δ_s-acyclic. Thus, $Z[n]$ belongs to $J(\mathcal{G}, \delta_s)$. In fact, any subset of $Z[n]$ containing $f(i)$ belongs to $J(\mathcal{G}, \delta_s)$. Now, consider the subgraph $\mathcal{G}'$ of $\mathcal{G}$ obtained by removing the packet node $f(i)$. Since $\mathcal{G}'$ is also δ_s-acyclic, there also exists at least one receiver R_j in $\mathcal{G}'$ whose number of side information symbols is less than or equal to $2\delta_s$. It means that any subset of $Z[n] \setminus \{f(i)\}$ containing $f(j)$ belongs to $J(\mathcal{G}, \delta_s)$. Through the similar ways, we can note that $Z[n]$ is a δ_s-generalized independent set. ■

The important theorem for a δ_s-cycle is given as follows.

Theorem 4: $\Phi = \phi$ if and only if $N_{\text{opt}}^q(\delta_s, \mathcal{G}) = n$ for the $(\delta_s, \mathcal{G})$-ICSIE.

Proof:

Necessity: From Lemma 3, there is an equivalence between a δ_s-generalized independent set and a δ_s-acyclic graph. If the set of packet node indices of $\mathcal{G}$, $Z[n]$, is the δ_s-generalized independent set, $\mathcal{I}(q, \mathcal{G}, \delta_s)$ is the set of all vectors in $\mathbb{F}_q^n$ except for the all zero vector. Thus, if $\Phi = \phi$, $Z[n]$ is the δ_s-generalized independent set. Thus, all rows of a generator matrix of the $(\delta_s, \mathcal{G})$-ICSIE should be linearly independent from Theorem 1. Then, we have $N_{\text{opt}}^q(\delta_s, \mathcal{G}) = n$.

Sufficiency: Suppose that $\Phi \neq \phi$. Then, we prove that the codelength can be reduced by at least one. We choose one δ_s-cycle $\mathcal{G}'$ in $\mathcal{G}$ and let $|B|$ be the number of packet nodes in $\mathcal{G}'$. Let $\mathbf{x} \in \mathbb{F}_q^n$ consist of two parts as $\mathbf{x}' \in \mathbb{F}_q^{|B|}$ related with $\mathcal{G}'$ and the remaining packets $\mathbf{x}'' \in \mathbb{F}_q^{n-|B|}$. If we encode $\mathbf{x}' \in \mathbb{F}_q^{|B|}$ as $(x_1' + x_2', x_2' + x_3', \ldots, x_{|B|-1}' + x_{|B|}')$ whose length is $|B| - 1$ and send $\mathbf{x}''$ in the uncoded form, it satisfies the property of the generator matrix in Theorem 1 as follows.

The generator matrix G' of this code except the uncoded part is given as

$$G' = \begin{pmatrix} 1 & & & & & \\ 1 & 1 & & & & \\ & 1 & 1 & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & & & & 1 \\ & & & & & 1 \end{pmatrix}.$$

We can easily find that any $|B| - 1$ rows of G' are linearly independent and $|B|$ rows of G' are linearly dependent. Thus, if there is no vector whose Hamming weight is $|B|$ in $\mathcal{I}(q, \mathcal{G}', \delta_s)$, G' satisfies the property of the generator matrix. Since $|\mathcal{X}_i'| \geq 2\delta_s + 1$ for every user i in $\mathcal{G}'$ and $|B| \geq 2\delta_s + 2$, there is no vector whose Hamming weight is $|B|$ in $\mathcal{I}(q, \mathcal{G}', \delta_s)$. Since there is the generator matrix G' whose codelength is $|B| - 1$, $N_{\text{opt}}^q(\delta_s, \mathcal{G}) \neq n$. It contradicts the assumption and thus $\Phi = \phi$. ■

Remark 7: From Theorem 4, we can conclude that a δ_s-cycle is very crucial for index coding with erroneous side information because the existence of a δ_s-cycle is a necessary and sufficient condition for a possibility to reduce the codelength of the $(\delta_s, \mathcal{G})$-ICSIE.

Next, we compare the δ_s-cycle of the ICSIE with the cycle of the conventional index code in the following remark, where we can find similarities between the δ_s-cycle and the cycle.

Remark 8: The δ_s-cycle of the ICSIE has very similar properties as those of cycles in the conventional index coding problem with respect to the index codelength.

The cycle properties of the conventional index coding problem in [13] are given as:

1) ([13, Th. 1]) If $\mathcal{G}$ is acyclic, $N_{\text{opt}}^q(0, \mathcal{G}) = n$.
2) ([13, Lemma 2]) If a packet node has one outgoing edge, we say that the packet node is a unicast packet node. That is, there is only one user who wants this packet. We can reduce the codelength by the number of disjoint cycles which only consist of unicast packet nodes.
3) ([13, Lemma 3]) If all distinct cycles of $\mathcal{G}$ are edge-disjoint and involve only unicast packets, then $N_{\text{opt}}^q(0, \mathcal{G}) = n - C(\mathcal{G})$, where $C(\mathcal{G})$ is defined as the largest number of edge-disjoint cycles in $\mathcal{G}$ that involve only unicast packets.

We compare properties of the δ_s-cycle in the ICSIE with those of the cycle in the conventional index coding problem [13]. If $\delta_s = 0$, there is at least one cycle in a 0-cycle because every receiver node has at least one outgoing edge. Then, from Theorem 4, the corresponding properties of the δ_s-cycle of the $(\delta_s, \mathcal{G})$-ICSIE are given as follows:

1) $\mathcal{G}$ is δ_s-acyclic if and only if $N_{\text{opt}}^q(\delta_s, \mathcal{G}) = n$.
2) Let β be the maximum number of vertex disjoint δ_s-cycles in $\mathcal{G}$. Then, the optimal codelength can be reduced by the number of disjoint δ_s-cycles, that is, $N_{\text{opt}}^q(\delta_s, \mathcal{G}) \leq n - \beta$. This is given in the proof of the sufficiency of Theorem 4.
3) The last property is given in the corollary below.

Note that the cycle property 1) of the conventional index coding is the sufficient condition but the corresponding property of the ICSIE is the necessary and sufficient condition.

Corollary 1: Suppose that there is no δ_s-cycle in a subgraph $\mathcal{G}'$ constructed by removing one packet node from each of β δ_s-cycles and the corresponding edges. Then, $N_{\text{opt}}^q(\delta_s, \mathcal{G}) = n - \beta$.

Proof: It is trivial that $n - \beta$ can be an upper bound for $N_{\text{opt}}^q(\delta_s, \mathcal{G})$ from the above property 2) and Theorem 4. Thus, it is enough to show that $n - \beta$ can be a lower bound. If we remove one packet node and the corresponding edges from each δ_s-cycle, the optimal codelength for the resulting graph is given as $N_{\text{opt}}^q(\delta_s, \mathcal{G}') = n - \beta$ because $\mathcal{G}'$ is δ_s-acyclic. Also, an index code for $\mathcal{G}$ satisfies the property of the generator matrix for $\mathcal{G}'$ if we just substitute values of the removed packet nodes with 0. Thus, we can say $N_{\text{opt}}^q(\delta_s, \mathcal{G}') \leq N_{\text{opt}}^q(\delta_s, \mathcal{G})$. That is, $n - \beta = N_{\text{opt}}^q(\delta_s, \mathcal{G})$. ■

An example for the previous corollary is provided as follows.

Example 5: Let $q = 2, m = 9, n = 8$, and $\delta_s = 1$. A side information graph $\mathcal{G}$ is shown in Fig. 3(a), where the maximum number of disjoint δ_s-cycles is 2. We denote two δ_s-cycle induced sets as B_1 and B_2, that is, $B_1 = \{1, 2, 3, 4\}$ and $B_2 = \{5, 6, 7, 8\}$. The edges for B_1 and B_2 are represented by solid and dotted lines, respectively. If we remove the packet nodes 1 and 5 and the corresponding edges from two δ_s-cycles, there is no δ_s-cycle in the resulting side information graph as shown in Fig. 3(b). Then, the optimal codelength is 6 from Corollary 1.

Remark 9: When $|\mathcal{X}_i| = 2\delta_s + 1$ for all $i \in Z[m]$, we have another example for Corollary 1.

B. Clique

In this section, we consider side information graphs, where each receiver knows all information except the wanted packet, that is, $\mathcal{X}_i = Z[n] \backslash \{f(i)\}$ for all $i \in Z[m]$. Since each receiver wants one symbol, we can consider these side information graphs as cliques. In the conventional index coding problem, the number of disjoint cliques is used as an upper bound of the optimal codelength and there are many heuristic algorithms to find the codelength based on the number of disjoint cliques. Therefore, we discuss cliques in the $(\delta_s, \mathcal{G})$-ICSIE problem. It is clear that cliques in index coding with erroneous side information are different from those of the conventional index coding, where cliques can be covered by one transmission for any fields.

A set of vectors is said to be $(2\delta_s + 1)$-linearly independent if any $2\delta_s + 1$ vectors in the vector set are linearly independent. Since all vectors whose Hamming weight is less than or equal to $2\delta_s + 1$ belong to $\mathcal{I}(q, \mathcal{G}, \delta_s)$ when $\mathcal{G}$ is a clique, we have the following observation.

Observation 1: In the $(\delta_s, \mathcal{G})$-ICSIE problem, finding the optimal codelength for cliques is equivalent to finding the

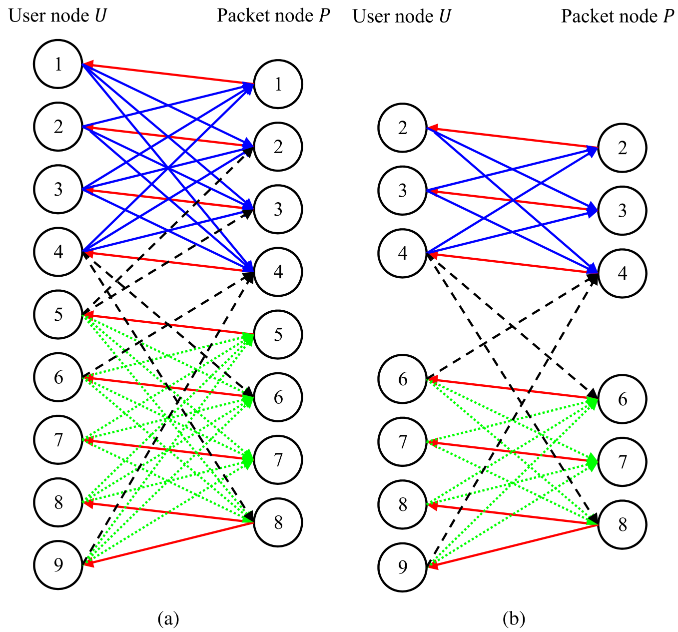

Fig. 3. The side information graphs of Example 5: (a) The side information graph $\mathcal{G}$. (b) The side information graph $\mathcal{G}'$ after removing packet nodes 1 and 5.

minimum dimension of the set of n vectors which is $(2\delta_s+1)$-linearly independent.

In general, $N_{\text{opt}}^q(\delta_s, \mathcal{G})$ goes to $2\delta_s+1$ as the size of the finite field goes to infinity. If the size of a clique is less than or equal to $2\delta_s+1$, it is easy to check that $N_{\text{opt}}^q(\delta_s, \mathcal{G}) = n$. Thus, we consider the clique of size $n > 2\delta_s + 1$.

Theorem 5: There are some special cases of cliques for the $(\delta_s, \mathcal{G})$-ICSIE whose optimal codelength can be found as:

1) When $n = 2\delta_s + 2$, $N_{\text{opt}}^q(\delta_s, \mathcal{G}) = 2\delta_s + 1$ over $\mathbb{F}_q$.
2) When $\delta_s = 1$, $N_{\text{opt}}^q(\delta_s, \mathcal{G})$ is the minimum value of N satisfying $2^{N-1} \geq n$ over $\mathbb{F}_2$.
3) If there are N and n satisfying $N \leq \frac{q+1}{q}(2\delta_s + 1) - 1$ and $n \leq N + 1$ over $\mathbb{F}_q$, $N_{\text{opt}}^q(\delta_s, \mathcal{G})$ is the minimum value of N.
4) When $n = 3\delta_s + 3$, $N_{\text{opt}}^q(\delta_s, \mathcal{G}) = 3\delta_s + 1$ over $\mathbb{F}_2$.

Proof:

1) It is directly proved from Corollary 1 because $|\mathcal{X}_i| = 2\delta_s + 1$ for all $i \in Z[m]$.
2) Let $Ind_q(N, k)$ be the maximal cardinality of the k-linearly independent subset of $\mathbb{F}_q^N$. We have $Ind_2(N, 3) = 2^{N-1}$ from [14], where $N \geq 3$. Thus, we can think N as a codelength and $k = 2\delta_s + 1$. To achieve the codelength N, the number of messages should be less than or equal to $Ind_q(N, k)$.
3) From [15, Th. 2], for $2 \leq k \leq N$, we have $Ind_q(N, k) = N + 1$ if and only if $\frac{q}{q+1}(N + 1) \leq k$.
4) From [14], we have $Ind_2(N, N - m) = N + 2$, where $N = 3m+i, i = 0, 1,$ and $m \geq 2$. Since $N-m = 2\delta_s+1$, we have $i = 1$ and $m = \delta_s$. Then, for $N = 3\delta_s + 1$ and $\delta_s \geq 2$, we have $Ind_2(N, 2\delta_s + 1) = N + 2$. If $\delta_s = 1$, $N_{\text{opt}}^q(\delta_s, \mathcal{G}) = 3\delta_s + 1$ by 2) when $n = 3\delta_s + 3$.

■

Remark 10: Construction of a generator matrix of each case in Theorem 5 is also well defined. Specifically, we can simply attain a generator matrix of case 1) as in proof of Theorem 4

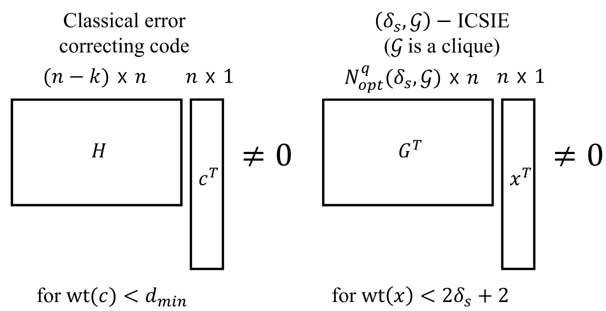

Fig. 4. Comparison between the classical error correcting code and the $(\delta_s, \mathcal{G})$-ICSIE.

and a matrix whose rows consist of any vectors having odd weight in $\mathbb{F}_2^N$ can be a generator matrix in the above case 2). We can find a generator matrix of case 3), which consists of N-tuple unit vectors and the all one vector. A generator matrix of case 4) consists of N-tuple unit vectors and two vectors $(1, ..., 1, 0, ..., 0)$ and $(0, ..., 0, 1, ..., 1)$, where the number of 1 in the two vectors is $2\delta_s + 1$. For all cases, we just select n vectors among them as rows of a generator matrix.

Next, we show how to construct a generator matrix of the $(\delta_s, \mathcal{G})$-ICSIE from a parity check matrix H of an (n, k, d_{min}) classical error correcting code in the following proposition.

Proposition 1: Let $\mathcal{G}$ be the clique of size n and $\bar{H}$ be the matrix having the smallest $n - k$ among H of (n, k, d_{min}) classical error correcting codes with $d_{min} \geq 2\delta_s + 2$. Then, $\bar{H}^\intercal$ becomes the optimal generator matrix of the $(\delta_s, \mathcal{G})$-ICSIE.

Proof: In the $(\delta_s, \mathcal{G})$-ICSIE problem, where a side information graph $\mathcal{G}$ is the clique of size n, it is clear that any $2\delta_s + 1$ row vectors of a generator matrix are linearly independent by Observation 1. That is, $\mathbf{x}G \neq \mathbf{0}$ for any $\mathbf{x}$ such that $\text{wt}(\mathbf{x}) \leq 2\delta_s + 1$. Then, we can easily check that an $n \times (n - k)$ matrix $H^\intercal$ can be a generator matrix of a $(\delta_s, \mathcal{G})$-ICSIE if $d_{min} \geq 2\delta_s+2$ as shown in Fig. 4. Thus, when a side information graph $\mathcal{G}$ is the clique of size n, the optimal codelength of the $(\delta_s, \mathcal{G})$-ICSIE is the minimum value of $n-k$ for an (n, k, d_{min}) classical error correcting code satisfying $d_{min} \geq 2\delta_s + 2$. ∎

Remark 11: One of examples is the Reed-Solomon code when n divides $q-1$. In this case, the optimal codelength of the $(\delta_s, \mathcal{G})$-ICSIE is $n-k = d_{min} - 1 = 2\delta_s + 1$.

Remark 12: Even if $\mathcal{G}$ is not a clique, we can regard the parity check matrix of the classical error correcting code as the transpose of the generator matrix of the $(\delta_s, \mathcal{G})$-ICSIE when d_{min} of the error correcting code is larger than the maximum weight of vectors in $\mathcal{I}(q, \mathcal{G}, \delta_s)$.

C. Lower Bounds for the Optimal Codelength

It is not difficult to check the following corollary and observations for the $(\delta_s, \mathcal{G})$-ICSIE.

Corollary 2: Let $S = \{j \in Z[n] | \exists i \in Z[m] \text{ s.t. } f(i) = j \text{ and } |\mathcal{X}_i| \leq 2\delta_s\}$. Then, we have $N_{\text{opt}}^q(\delta_s, \mathcal{G}) \geq |S| + 1$ for $n > |S|$.

Proof: It is clear that $\mathcal{I}_i(q, \mathcal{G}, \delta_s) = \{\mathbf{z} \in \mathbb{F}_q^n : z_{f(i)} \neq 0\}$, that is, $G_{f(i)}$ does not belong to $\text{span}(\{G_j\}_{j \in Z[n] \setminus f(i)})$ for $f(i) \in S$. Thus, the corollary is obvious. ∎

Remark 13: Thus, having less than or equal to $2\delta_s$ side information symbols is the same as not having side information in index coding with erroneous side information.

Observation 2: For the given $(\delta_s, \mathcal{G})$-ICSIE problem, let $\mathcal{G}'$ be an edge-induced subgraph obtained by deleting some outgoing edges of user nodes of $\mathcal{G}$. Then, $N_{\text{opt}}^q(\delta_s, \mathcal{G}) \leq N_{\text{opt}}^q(\delta_s, \mathcal{G}')$.

Observation 3: If $\delta_s' \leq \delta_s$, $N_{\text{opt}}^q(\delta_s', \mathcal{G}) \leq N_{\text{opt}}^q(\delta_s, \mathcal{G})$.

Now, the relationship of the optimal codelength between the conventional index code and the proposed ICSIE is given in the following theorem.

Theorem 6: Suppose that the $(0, \bar{\mathcal{G}})$-IC problem is constructed by deleting any $\min(2\delta_s, |\mathcal{X}_i|)$ outgoing edges from each receiver R_i in the $(\delta_s, \mathcal{G})$-ICSIE problem. That is, each receiver of $\bar{\mathcal{G}}$ has $\max(0, |\mathcal{X}_i| - 2\delta_s)$ side information symbols and then it becomes the conventional index coding problem. Then, $N_{\text{opt}}^q(0, \bar{\mathcal{G}}) \leq N_{\text{opt}}^q(\delta_s, \mathcal{G})$.

Proof: From Lemma 2, a generator matrix of the $(\delta_s, \mathcal{G})$-ICSIE problem can be a generator matrix of the $(0, \bar{\mathcal{G}})$-IC problem because $\mathcal{I}(q, \bar{\mathcal{G}}, 0) \subseteq \mathcal{I}(q, \mathcal{G}, \delta_s)$. Specifically, for a vector $\mathbf{z}' \in \mathcal{I}_i(q, \bar{\mathcal{G}}, 0)$, $\text{wt}(\mathbf{z}'_{\mathcal{X}_i}) \leq 2\delta_s$ since $\text{wt}(\mathbf{z}'_{\mathcal{X}_i'})$ should be zero, where $\mathcal{X}_i'$ is a set of side information indices of R_i for $\bar{\mathcal{G}}$. Thus, $\mathcal{I}_i(q, \bar{\mathcal{G}}, 0) \subseteq \mathcal{I}_i(q, \mathcal{G}, \delta_s)$. Since it is true for all $i \in Z[m]$, $\mathcal{I}(q, \bar{\mathcal{G}}, 0) \subseteq \mathcal{I}(q, \mathcal{G}, \delta_s)$. ∎

Remark 14: In general, if we reduce δ_s, we can have a lower bound from Observation 3. Similarly, if we delete outgoing edges of user nodes, we can have an upper bound from Observation 2. However, if we reduce δ_s to 0 and delete $\min(2\delta_s, |\mathcal{X}_i|)$ outgoing edges of each receiver R_i, we can have a lower bound as in Theorem 6. Thus, the worst case of the resulting $(0, \bar{\mathcal{G}})$-IC problems can be a lower bound for the corresponding $(\delta_s, \mathcal{G})$-ICSIE problem.

Example 6: Let $q = 2, n = m = 4, \delta_s = 1, f(i) = i$, and $\mathcal{G}$ as shown in Fig. 5(a). Then, we have $N_{\text{opt}}^q(1, \mathcal{G}) = 4$. If we delete two outgoing edges from each receiver, there is a side information graph $\bar{\mathcal{G}}$ as shown in Fig. 5(b). In the conventional index coding problem, $N_{\text{opt}}^q(0, \bar{\mathcal{G}}) = n$ if a side information graph $\bar{\mathcal{G}}$ is acyclic [13]. Since the graph in Fig. 5(b) is acyclic, $N_{\text{opt}}^q(0, \bar{\mathcal{G}}) = 4$. From Theorem 6, $N_{\text{opt}}^q(1, \mathcal{G}) = 4$ because $4 = N_{\text{opt}}^q(0, \bar{\mathcal{G}}) \leq N_{\text{opt}}^q(1, \mathcal{G}) \leq 4 = n$.

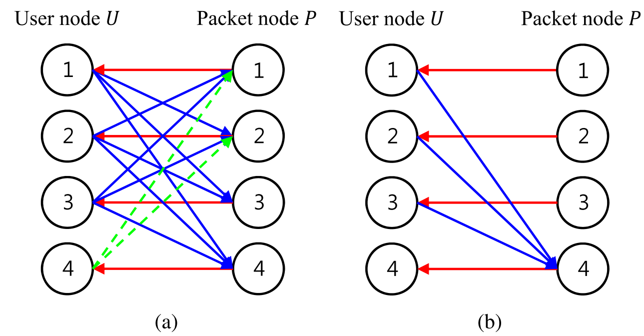

Fig. 5. The bipartite side information graphs of Example 6: (a) The bipartite side information graph $\mathcal{G}$. (b) The bipartite side information graph $\bar{\mathcal{G}}$.

Example 7: Let $q = 2, n = m = 4, \delta_s = 1, f(i) = i$, and $\mathcal{X}_i = Z[4] \setminus \{i\}$ for all $i \in Z[4]$ as in Example 1. If we delete two outgoing edges from each receiver, the corresponding graph has at least one cycle because each receiver has one outgoing edge. In this case, we can reduce the codelength by at least one because all cycles in the graph consist of unicast packets [13]. Then, the worst case of the corresponding graph $\bar{\mathcal{G}}$ has $N_{\text{opt}}^q(0, \bar{\mathcal{G}}) = 3$. Thus, $3 \leq N_{\text{opt}}^q(1, \mathcal{G}) \leq 4$. In fact, we have $3 \leq N_{\text{opt}}^q(1, \mathcal{G}) = 3$ because there is a generator matrix of the $(\delta_s, \mathcal{G})$-ICSIE given by

$$G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

It is easy to derive the following lower bound for the optimal codelength.

Theorem 7: $N_{\text{opt}}^q(\delta_s, \mathcal{G}) \geq \gamma(\mathcal{G})$.

Proof: From the definition of $\gamma(\mathcal{G})$, the corresponding $\gamma(\mathcal{G})$ rows of a generator matrix of the $(\delta_s, \mathcal{G})$-ICSIE should be linearly independent. ∎

V. Generalized Error Correcting Index Codes

We can generalize many properties of the ECIC in [9] by considering the side information errors. That is, we can describe the properties of the $(\delta_s, \delta_c, \mathcal{G})$-GECIC similar to those of the $(\delta_c, \mathcal{G})$-ECIC in [9] by using the properties of the $(\delta_s, \mathcal{G})$-ICSIE. Some notations of the ECIC in [9] are changed for consistency within paper.

Proposition 2: Properties of the $(\delta_c, \mathcal{G})$-ECIC in [9] can be generalized to those of the $(\delta_s, \delta_c, \mathcal{G})$-GECIC as:

1) Generalization of [9, Lemma 3.8];
   Theorem 1
2) Generalization of [9, Proposition 4.6];
   $N_{\text{opt}}^q(\delta_s, \delta_c, \mathcal{G}) \leq l_q[N_{\text{opt}}^q(\delta_s, \mathcal{G}), 2\delta_c + 1]$, where $l_q[a, b]$ denotes the minimum codelength for the dimension a and $d_{min} = b$.
3) Generalization of [9, Th. 5.1];
   $N_{\text{opt}}^q(\delta_s, \delta_c, \mathcal{G}) \geq N_{\text{opt}}^q(\delta_s, \mathcal{G}) + 2\delta_c$.
4) Generalization of the property of $\gamma(\mathcal{G})$ in [9];
   Assume that $m = n$ and $f(i) = i$ for all $i \in Z[n]$ so that the side information graph $\mathcal{G}$ can be represented as

the unipartite form. Then, $\gamma(\mathcal{G}) = \delta_s$-MAIS($\mathcal{G}$), where δ_s-MAIS($\mathcal{G}$) denotes the maximum size of a δ_s-acyclic induced subgraph of $\mathcal{G}$.

5) Generalization of [9, Th. 4.3]; $N_{\text{opt}}^q(\delta_s, \delta_c, \mathcal{G}) \geq l_q[\gamma(\mathcal{G}), 2\delta_c + 1]$.

Proof: From Lemma 2, all generalization except 4) can be easily proved by the same methods as in [9] if we replace the conventional index code with the $(\delta_s, \mathcal{G})$-ICSIE. In the case of 4), we already prove an equivalence between a δ_s-generalized independent set and a δ_s-acyclic graph in Lemma 3. ■

Remark 15: By 4) of Proposition 2, we can think that a δ_s-cycle corresponds to a cycle in the conventional index coding.

Now, we introduce some properties of the $(\delta_s, \delta_c, \mathcal{G})$-GECIC.

Theorem 8: Suppose that the $(\delta_c, \bar{\mathcal{G}})$-ECIC problem is constructed by deleting any $\min(2\delta_s, |\mathcal{X}_i|)$ outgoing edges from each receiver R_i in the $(\delta_s, \delta_c, \mathcal{G})$-GECIC problem. That is, each receiver of $\bar{\mathcal{G}}$ has $\max(0, |\mathcal{X}_i| - 2\delta_s)$ side information symbols and then it becomes the conventional error correcting index coding problem. Then, $N_{\text{opt}}^q(\delta_c, \bar{\mathcal{G}}) \leq N_{\text{opt}}^q(\delta_s, \delta_c, \mathcal{G})$.

Proof: The proof is similar to that of Theorem 6 and thus we omit it. ■

Theorem 9: Let $\hat{\mathcal{G}}$ be an edge-induced subgraph of $\mathcal{G}$, which is obtained by deleting all outgoing edges of all users in $\mathcal{G}$, that is, none of the receivers have any side information in $\hat{\mathcal{G}}$. Then, $N_{\text{opt}}^q(\delta_s, \delta_c, \mathcal{G}) = N_{\text{opt}}^q(\delta_s, \delta_c, \hat{\mathcal{G}})$ if $\Phi = \phi$.

Proof: If $\Phi = \phi$, $Z[n]$ is a δ_s-generalized independent set. Then, $\mathcal{I}(q, \mathcal{G}, \delta_s) = \mathcal{I}(q, \hat{\mathcal{G}}, \delta_s)$ and thus $N_{\text{opt}}^q(\delta_s, \delta_c, \mathcal{G}) = N_{\text{opt}}^q(\delta_s, \delta_c, \hat{\mathcal{G}})$. ■

In Section III-B, the decoding procedure of the $(\delta_s, \mathcal{G})$-ICSIE was introduced in Algorithm 1 and the decoding procedure of the $(\delta_s, \delta_c, \mathcal{G})$-GECIC can also be derived similarly. Since the decoding procedure of the $(\delta_s, \delta_c, \mathcal{G})$-GECIC is similar to Algorithm 1, we just enumerate some differences as follows.

1) Each receiver receives a codeword $\mathbf{y}' = \mathbf{y} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is the error with $\text{wt}(\boldsymbol{\epsilon}) \leq \delta_c$.
2) The input of the decoding procedure $\mathbf{y}$ is changed to $\mathbf{y}'$.
3) $H_e^{(i)}$ is a matrix whose rows form a basis of the dual of $\text{span}(\{G_j\}_{j \in \mathcal{Y}_i})$ and do not form a basis of the dual of $\text{span}(G_{f(i)})$.
4) Equation (2) in Algorithm 1 is changed to $\mathbf{s}_i = H^{(i)}(\mathbf{y}' - \hat{\mathbf{x}}_{\mathcal{X}_i} G_{\mathcal{X}_i})^\intercal$.
5) Step 2) in Algorithm 1: Find a solution having syndrome $\mathbf{s}_i$ in the set $\{\mathbf{p}_i + \hat{\boldsymbol{\epsilon}} | \mathbf{p}_i$ is a linear combination of at most δ_s rows of $G_{\mathcal{X}_i}$ and $\text{wt}(\hat{\boldsymbol{\epsilon}}) \leq \delta_c\}$.
6) Equation (3) in Algorithm 1 is changed to

$$\tilde{\mathbf{y}} = \mathbf{y}' - \hat{\mathbf{x}}_{\mathcal{X}_i} G_{\mathcal{X}_i} - \mathbf{p}_i - \hat{\boldsymbol{\epsilon}} = x_{f(i)} G_{f(i)} + (\mathbf{x}_{\mathcal{Y}_i} - \mathbf{b}) G_{\mathcal{Y}_i}. \quad (5)$$

It is enough to show validity of (5) in order to prove successful decoding of each receiver. We can see the following equations in the decoding procedure of the $(\delta_s, \delta_c, \mathcal{G})$-GECIC as

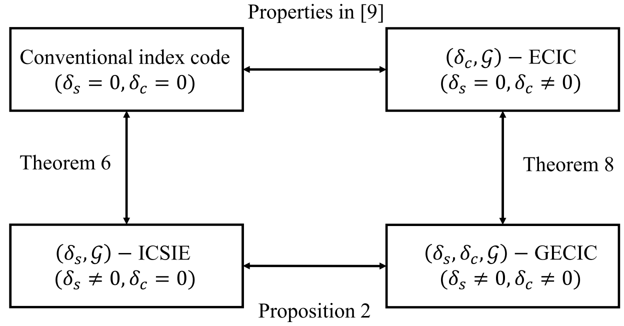

Fig. 6. Relationship of several index coding problems for the optimal codelength.

$$\begin{aligned}\mathbf{s}_i &= H^{(i)}(\mathbf{y}' - \hat{\mathbf{x}}_{\mathcal{X}_i} G_{\mathcal{X}_i})^\intercal = H^{(i)}(\tilde{\mathbf{x}}_{\delta_s} G_{\mathcal{X}_i} + \boldsymbol{\epsilon})^\intercal = H^{(i)}(\mathbf{p}_i + \hat{\boldsymbol{\epsilon}})^\intercal \\ &H^{(i)}(\tilde{\mathbf{x}}_{\delta_s} G_{\mathcal{X}_i} + \boldsymbol{\epsilon} - \mathbf{p}_i - \hat{\boldsymbol{\epsilon}})^\intercal = \mathbf{0} \\ &\tilde{\mathbf{x}}_{\delta_s} G_{\mathcal{X}_i} + \boldsymbol{\epsilon} - \mathbf{p}_i - \hat{\boldsymbol{\epsilon}} = a G_{f(i)} - \mathbf{b} G_{\mathcal{Y}_i}.\end{aligned}$$

Then, by Theorem 1, $a = 0$ and we have

$$-\mathbf{p}_i - \hat{\boldsymbol{\epsilon}} = -\tilde{\mathbf{x}}_{\delta_s} G_{\mathcal{X}_i} - \boldsymbol{\epsilon} - \mathbf{b} G_{\mathcal{Y}_i}.$$

Since $\mathbf{y}' = x_{f(i)} G_{f(i)} + \mathbf{x}_{\mathcal{X}_i} G_{\mathcal{X}_i} + \mathbf{x}_{\mathcal{Y}_i} G_{\mathcal{Y}_i} + \boldsymbol{\epsilon}$, we have (5).

In Fig. 6, we show the relationship between the proposed index codes and several index coding problems, specifically in terms of the optimal codelength.

VI. Conclusions

We generalized the index coding problem, where there is a possibility to have erroneous side information in each receiver. The property of the generator matrix and the decoding procedure of the proposed index codes with erroneous side information were suggested, which are based on the idea of Hamming spheres and the syndrome decoding, respectively.

We also suggested some bounds for the optimal codelength of the $(\delta_s, \mathcal{G})$-ICSIE and showed the relationship between the conventional index coding and index coding with erroneous side information. In addition, we found a δ_s-cycle of the GECIC, which has similar properties as those of a cycle in the conventional index coding. It was also found that the existence of a δ_s-cycle is crucial in the proposed index coding problem. The proposed ICSIE was also analyzed when a side information graph $\mathcal{G}$ is a clique. Through this, it was found that the generator matrix for the $(\delta_s, \mathcal{G})$-ICSIE corresponds to the transpose of the parity check matrix of the classical error correcting code when the related parameters are properly chosen.

Finally, it was shown that the existing bounds and properties for the $(\delta_c, \mathcal{G})$-ECIC can be generalized to those of the $(\delta_s, \delta_c, \mathcal{G})$-GECIC by using the properties of the $(\delta_s, \mathcal{G})$-ICSIE. That is, we can easily derive bounds for the optimal codelength of the $(\delta_s, \delta_c, \mathcal{G})$-GECIC, which is the index code in the more generalized scenario.

VII. Acknowledgment

The authors are so grateful to the anonymous reviewers for their valuable comments.

REFERENCES

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

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

[3] E. Lubetzky and U. Stav, "Nonlinear index coding outperforming the linear optimum," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3544–3551, Aug. 2009.

[4] 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.

[5] S. A. Jafar, "Topological interference management through index coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 529–568, Jan. 2014.

[6] D. T. H. Kao, M. A. Maddah-Aliy, and A. S. Avestimehr, "Blind index coding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 2371–2375.

[7] K. W. Shum, M. Dai, and C. W. Sung, "Broadcasting with coded side information," in *Proc. IEEE 23rd Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2012, pp. 89–94.

[8] N. Lee, A. G. Dimakis, and R. W. Heath, Jr., "Index coding with coded side-information," *IEEE Commun. Lett.*, vol. 19, no. 3, pp. 319–322, Mar. 2015.

[9] S. H. Dau, V. Skachek, and Y. M. Chee, "Error correction for index coding with side information," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1517–1531, Mar. 2013.

[10] L. Natarajan, Y. Hong, and E. Viterbo, "Index codes for the Gaussian broadcast channel using quadrature amplitude modulation," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1291–1294, Aug. 2015.

[11] L. Natarajan, Y. Hong, and E. Viterbo, "Capacity of coded index modulation," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 596–600.

[12] E. Byrne and M. Calderini, "Error correction for index coding with coded side information," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3712–3728, Jun. 2017.

[13] M. J. Neely, A. S. Tehrani, and Z. Zhang, "Dynamic index coding for wireless broadcast networks," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7525–7540, Nov. 2013.

[14] S. B. Damelin, G. Michalski, G. L. Mullen, and D. Stone, "The number of linearly independent binary vectors with applications to the construction of hypercubes and orthogonal arrays, pseudo (t, m, s)-nets and linear codes," *Monatshefte Math.*, vol. 141, pp. 277–288, Apr. 2004.

[15] S. B. Damelin, G. Michalski, and G. L. Mullen, "The cardinality of sets of k-independent vectors over finite fields," *Monatshefte Math.*, vol. 150, no. 4, pp. 289–295, 2007.

Jae-Won Kim received the B.S. degree in electrical and computer engineering from Seoul National University, Seoul, Korea, in 2014, where he is currently pursuing the Ph.D. degree in electrical engineering and computer science. His area of research interests includes error-correcting codes, coding theory, and index coding.

Jong-Seon No (S'80–M'88–SM'10–F'12) received the B.S. and M.S.E.E. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1981 and 1984, respectively, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1988. He was a Senior MTS with Hughes Network Systems from 1988 to 1990. He was an Associate Professor with the Department of Electronic Engineering, Konkuk University, Seoul, from 1990 to 1999. He joined the faculty of the Department of Electrical and Computer Engineering, Seoul National University, in 1999, where he is currently a Professor. His area of research interests includes error-correcting codes, sequences, cryptography, LDPC codes, interference alignment, and wireless communication systems. He was a recipient of the IEEE Information Theory Society Chapter of the Year Award in 2007. From 1996 to 2008, he served as a Founding Chair of the Seoul Chapter of the IEEE Information Theory Society. He was a General Chair of Sequence and Their Applications 2004, Seoul. He also served as a General Co-Chair of the International Symposium on Information Theory and Its Applications 2006 and the International Symposium on Information Theory 2009, Seoul. He has been a Co-Editor-in-Chief of the IEEE JOURNAL OF COMMUNICATIONS AND NETWORKS since 2012.