Generalized Interlinked Cycle Cover for Index Coding

Chandra Thapa, Lawrence Ong, and Sarah J. Johnson
School of Electrical Engineering and Computer Science, The University of Newcastle, Newcastle, Australia
Email: chandra.thapa@uon.edu.au, lawrence.ong@cantab.net, sarah.johnson@newcastle.edu.au

Abstract—A source coding problem over a noiseless broadcast channel where the source is preformed the contents of the cache of all receivers is an index coding problem. Furthermore, if each receiver is interested in only a unique message, we call this the index coding problem with the unicast message setting. This problem can be represented by a directed graph. The main aim of the index coding problem is to reduce the number of transmission of coded packets. This paper studies these kinds of problems. We propose a simple linear encoding scheme with linear time encoding complexity, which is generated by exploiting index coding present in a digraph. We proved that our scheme is optimal for a class of digraphs with message packets of any length. Moreover, we showed that our scheme can outperform existing techniques, e.g., partial clique cover, local chromatic number, composite coding, interlinked cycle cover.

Index Terms—Index coding problem, unicast, optimal broadcast, linear codes, interlinked cycles.

I. INTRODUCTION

We consider a transmitter broadcasting message packets through a noiseless broadcast channel to multiple receivers, each knowing some message packets a priori, which is known as side information. The side information of receivers can be utilized in order to reduce the number of coded symbols required to be broadcast by the transmitter to all receivers. This is known as the index coding problem, and was introduced by Birk and Kol in 1998 [1]. To date, the index coding problem is an open problem. The main aim of the index coding problem is to find an index code that has the minimum number of coded symbols. When each receiver is interested in only one unique message packet (i.e., unicast), the index coding problems can be modeled by digraphs (directed graphs). This paper considers the unicast message setting.

Linear index codes (scalar and vector linear) [1]–[5] have simpler encoding and decoding process than non-linear index codes. In the literature, optimal scalar linear index codes can be characterized by a graph function called the minrank function [2]. However finding minrank for a general digraph is NP-hard [6], and does not provide much intuition on how to encode a specific side information. Thus, in this paper we use graph-theoretic approach to address specific graph structures.

There are various graph-theoretic approaches such as clique cover [1], partial clique cover [1], cycle cover [2], [3], [7], graph coloring (chromatic number, local chromatic number, and their fractional relaxation) [1], [8], [9], which exploit the graph structure during encoding of the index code to save transmission (i.e., compared to sending uncoded message packets). At ISIT 2015, we will present a new coding scheme exploiting interlinked cycles [10] present in a digraph. This new scheme, called interlinked cycle cover (ICC), generalizes clique cover and cycle cover schemes. We showed that for a class of digraphs, called interlinked cycle cover (ICC) digraphs, ICC can outperform some existing schemes. In an interlinked cycles structure with \( N \) vertices, there exists a set of \( K \) vertices, where each vertex has a directed path to every other vertex of the set. Each of these \( K \) vertices must have out-degree equal to \( K - 1 \), and all remaining vertices (i.e., \( N - K \) vertices) must have out-degree equal to one, and both can have in-degree greater than or equal to one. These conditions on out-degree of vertices make the definition of interlinked cycles rather limiting, which restrict the size of the class of ICC digraphs (for which the ICC scheme is optimal). Furthermore, we were unable to show that the ICC scheme can outperform the composite coding (based on random coding approach which requires infinitely long message packets) proposed by Arbabjolfaei et al. [11].

This paper redefines (and extends) the previous definition of interlinked cycle structure, so that both in-degree and out-degree of any vertex in interlinked cycles, are allowed to be greater than or equal to one. The resultant interlinked cycles structure is called generalized interlinked cycles (GIC). Based on GIC, we proposed a simple encoding scheme called generalized interlinked cycle cover (GICC) which can achieve optimal saving (i.e., optimal index codes) for a larger class of digraphs including all ICC digraphs. We also find digraphs where the GICC scheme strictly outperforms the composite coding approach.

A. Our Contributions

1) We generalize the ICC, clique cover, and cycle cover schemes.
2) We characterize a class of digraphs where GICC is optimal (over all codes, including non-linear index codes).
3) We showed GICC scheme can outperform existing techniques (e.g., partial clique cover, local chromatic number, composite coding, interlinked cycle cover).
II. DEFINITIONS

Consider a transmitter that wants to transmit $N$ message packets $X = \{x_1, x_2, \ldots, x_N\}$ to $N$ receivers $\{1, 2, \ldots, N\}$ in a unicast message setting such that each receiver $i$ is requesting a message packet $x_i$. Moreover, each receiver $i$ has side information $S_i \subseteq X \setminus \{x_i\}$. This problem can be described by a digraph $D = (V(D), A(D))$, where $V(D) = \{1, 2, \ldots, N\}$ is a set of vertices representing the $N$ receivers. An arc $(i \to j) \in A(D)$ exists from vertex $i$ to vertex $j$ if and only if receiver $i$ has packet $x_j$ (requested by receiver $j$) as its side information. The side information of vertex $i$ is $S_i \triangleq \{x_j : j \in N^D_1(i)\}$, where $N^D_1(i)$ is the out-neighborhood of $i$ in $D$. For simplicity, we use the term “messages” to refer to message packets in the remainder of this paper.

**Definition 1:** (Valid index code) Suppose $x_i \in \{0, 1\}^t$ for all $i$, for some integer $t \geq 1$, i.e., each message consists of $t$ bits. Given an index coding problem described by $D$, a valid index code $(\mathcal{F}, \{\mathcal{G}_i\})$ is defined as follows:

1. An encoding function for the source, $\mathcal{F} : \{0, 1\}^{nt} \to \{0, 1\}^p$, which maps $X$ to a $p$-bit index for some integer $p$.
2. A decoding function $\mathcal{G}_i$ for every receiver $v_i$, $\mathcal{G}_i : \{0, 1\}^t \times \{0, 1\}^{|S_i|^h} \to \{0, 1\}^t$, that maps the received index $\mathcal{F}(X)$ and its side information $S_i$ to the requested message $x_i$.

The broadcast rate of the $(\mathcal{F}, \{\mathcal{G}_i\})$ index code is the number of transmitted bits per received message bits at each vertex, or equivalently the number of coded packets (of $t$ bits), denoted by $\beta(D) \triangleq \frac{p}{t}$. Thus, the optimal broadcast rate for a given index coding problem $D$ with $t$-bit message is $\beta^*(D) \triangleq \min_{\mathcal{F}} \beta(D)$, and the optimal broadcast rate over all $t$ is defined as $\beta^*(D) \triangleq \inf_{t} \beta^*(D)$.

**Definition 2:** (Path and cycle) A path contains a sequence of vertices (except possibly the first and last) $i, j, k$, and an arc $(i \to j)$ for each consecutive pair of vertices $(i, j)$ for all $i \in \{1, \ldots, M - 1\}$. We represent a path in a digraph $D$ as $P_{i \to j}(D) = \{1, \ldots, M\}$. Here, 1 is the first vertex, $M$ is the last vertex, and all remaining vertices $(2, 3, \ldots, M - 1)$ are internal vertices of the path. A path with the same first and last vertex is a cycle.

III. DISCUSSION OF OUR APPROACH

**A. Description of the GIC structure**

Consider a graph structure with $N$ vertices having the following properties:

1. A set of $K$ vertices, denoted by $V_1$, such that for any ordered pair of its vertices $(i, j)$, $i \neq j$, there is a path from $i$ to $j$ which does not include any other vertex of $V_1$. We call $V_1$ the inner vertex set and without loss of generality, we represent it as $V_1 = \{1, 2, \ldots, K\}$. The vertices of $V_1$ are referred to as inner vertices.

2. Due to the existence of paths in between any $(i, j) \in V_1$, for each vertex $i$, we can always find a directed rooted tree, denoted by $T_i$ where vertex $i$ is the root vertex, and all vertices of $V_1 \setminus \{i\}$ are the leaves (see Fig. 1(a)). The trees may be non-unique.

Denote the union of all selected $K$ trees as $D_K \triangleq \bigcup_i T_i$ (see Fig. 1(b)). If $D_K$ satisfies two conditions (to be defined shortly), we call it a generalized interlinked cycles, or simply GIC. We first define a type of cycles and paths.

**Definition 3:** (I-cycle) A cycle that has only one inner vertex $i \in V_1$ is an I-cycle.

**Definition 4:** (P-path) A path in which only the first and the last vertices are from $V_1$, and they are unique, is a P-path.

Now, the conditions for a $D_K$ to be qualified as a GIC are as follows:

1. **Condition 1:** There is no I-cycle.
2. **Condition 2:** For all ordered pairs of inner vertices $(i, j)$, $i \neq j$, there is only one P-path from $i$ to $j$.

A GIC with $K$ inner vertices is described by a $K$-GIC sub-digraph: $D_K = (V(D_K), A(D_K))$, where $|V(D_K)| = N$ and $V_1 = \{1, 2, \ldots, K\}$.

**B. Code construction for a $K$-GIC sub-digraph**

We propose the following coded symbols (which are scalar linear index code) for a $K$-GIC sub-digraph $D_K$:

1. A coded symbol obtained by the bitwise XOR (denoted by $\oplus$) of messages (each of $t$-bits) requested by all vertices of the inner vertex set $V_1$, i.e.,
   \[
   w_1 \triangleq \bigoplus_{i=1}^{K} x_i. \tag{1}
   \]

2. For each vertex $j \in \{K + 1, K + 2, \ldots, N\}$, a coded symbol obtained by the bitwise XOR of the message requested by $j$ with the messages requested by its out-neighborhood vertices, i.e.,
   \[
   w_j \triangleq x_j \oplus \bigoplus_{q \in N^D_{\text{out}}(j)} x_q. \tag{2}
   \]

The index code constructed for the $K$-GIC sub-digraph is $W \triangleq \{w_1, w_j : K + 1 \leq j \leq N\}$. The calculation of...
the total number of coded symbols, each of \( t \)-bits, in \( W \) is straightforward,

\[
\beta(D_K) = N - K + 1.
\]

**Remark 1:** Encoding \( W \) requires less than or equal to
\[ t \times \left( (K - 1) + \sum_{v \in V(D_K) \setminus V_i} |N^+_{D_K}(v)| \right) \]
bit-wise XOR operations.

Now let us show that all \( N \) vertices of \( D_K \) can decode their respective requested messages from \( W \).

From \( G \), all \( j \in \{ K+1, K+2, \ldots, N \} \) which are non-inner vertices, can decode their requested messages. This is because the coded symbol \( w_j \) is the bitwise XOR of the messages requested by \( j \) and its out-neighborhood, and any \( j \) knows messages requested by all of its out-neighborhood as side information.

For an inner vertex \( i \), rather than analyzing the sub-digraph \( D_K \), we will analyze its tree \( T_i \), and show that it can decode its message from the relevant symbols in \( W \). We are able to consider only tree \( T_i \) because we will show in Lemma 5 later that for a non-inner vertex, the out-neighborhood is same in a tree and in \( D_K \).

Let us take any tree \( T_i \). Assume that it has a height \( H \) where \( 1 \leq H \leq (N-K+1) \). The vertices in \( T_i \) are at various depths, i.e., \( \{ 0, 1, 2, \ldots, H \} \) from the root vertex \( i \). The root vertex \( i \) has depth zero, and any vertex at depth equal to the height of the tree is a leaf vertex.

First of all, in the tree \( T_i \), we compute the bitwise XOR among coded symbols of all non-leaf vertices at depth greater than zero, i.e., \( Z_i = \bigoplus_{j \in V(T_i) \setminus V_i} w_j \). However, in the tree \( T_i \), the message requested by a non-leaf vertex \( p \) at a depth greater than one, appears exactly twice in \( \{ w_j : j \in V(T_i) \setminus V_i \} \):

1) once in \( w_k \), where \( k \) is parent of \( p \) in tree \( T_i \), and
2) once in \( w_p \).

Thus, they cancel out each other while computing \( Z_i \) and a resultant expression which is bitwise XOR of

1) messages requested by all non-leaf vertices at depth one, and
2) messages requested by all leaf vertices at depth greater than one,
in the tree \( T_i \) is obtained. Secondly, we compute \( w_1 \oplus Z_i \) which yields the bitwise XOR of

1) messages requested by all non-leaf vertices at depth one which are in out-neighborhood of \( i \),
2) messages requested by all leaf vertices at depth one which are also in out-neighborhood of \( i \), and
3) message requested by \( i \), i.e., \( x_i \),
in the tree \( T_i \) is obtained. This is because bitwise XOR of messages requested by all leaf vertices at depth greater than one in the tree \( T_i \) is present in both the resultant terms of \( Z_i \) and in \( w_1 \), thereby they cancel out in \( w_1 \oplus Z_i \). Hence, \( w_1 \oplus Z_i \) yields the bitwise XOR of \( x_i \) and \( \{ x_j : j \in N^+_{D_K}(i) \} \). As \( i \) knows all \( \{ x_j : j \in N^+_{D_K}(i) \} \) as the prior side-information, so any inner vertex \( i \) can decode its required message from \( w_1 \oplus Z_i \). The detail is in the mathematical analyses which follows.

In the tree \( T_i \),

\[
Z_i = \bigoplus_{j \in V(T_i) \setminus V_i} w_j
\]

\[
= \bigoplus_{j \in V(T_i) \setminus V_i} \left( x_j \oplus \bigoplus_{q \in N^+_{D_K}(j) \setminus V_i} x_q \right)
\]

\[
= \bigoplus_{j \in V(T_i) \setminus V_i} \left( x_j \oplus \bigoplus_{q \in N^+_{D_K}(j) \setminus V_i} x_q \right)
\]

\[
= X_{V(T_i)} \bigoplus X'_{V(T_i)}.
\]

Where,

\[
X_{V(T_i)} = \bigoplus_{j \in V(T_i) \setminus V_i} \left( x_j \oplus \bigoplus_{q \in N^+_{D_K}(j) \setminus V_i} x_q \right),
\]

\[
X'_{V(T_i)} = \bigoplus_{j \in V(T_i) \setminus V_i} \left( \bigoplus_{q \in N^+_{D_K}(j) \setminus V_i} x_q \right).
\]

Here \( X'_{V(T_i)} \) is bitwise XOR of messages requested by all leaf vertices which branch from all non leaf and non root vertices in \( T_i \). If we expand \( X_{V(T_i)} \) as per the group of vertices according to their depth, then we get,

\[
X_{V(T_i)} = \bigoplus_{j_1 \in N^+_{D_K}(i) \setminus V_i} \left( x_{j_1} \oplus \bigoplus_{q \in N^+_{D_K}(j_1) \setminus V_i} x_q \right)
\]

\[
= \bigoplus_{j_1 \in N^+_{D_K}(i) \setminus V_i} \left( x_{j_1} \oplus \bigoplus_{q \in N^+_{D_K}(j_1) \setminus V_i} x_q \right) + \bigoplus_{j_2 \in N^+_{D_K}(j_1) \setminus V_i} \left( x_{j_2} \oplus \bigoplus_{q \in N^+_{D_K}(j_2) \setminus V_i} x_q \right) + \ldots + \bigoplus_{j_{H-2} \in N^+_{D_K}(j_{H-3}) \setminus V_i} \left( x_{j_{H-2}} \oplus \bigoplus_{q \in N^+_{D_K}(j_{H-2}) \setminus V_i} x_q \right)
\]

\[
= \bigoplus_{j \in N^+_{D_K}(i) \setminus V_i} \left( x_{j} \right)
\]

Note that the intermediate terms in \( X_{V(T_i)} \) cancel out (we have use the same color to indicate terms that cancel each other).
Now, in the tree $T_i$, 

$$w_1 \oplus z_i = w_1 \oplus x_{V(T_i)} \oplus x'_{V(T_i)} = x_i \oplus \bigoplus_{j \in V_i \setminus \{i\}} x_j \oplus x_k \oplus \bigoplus_{q \in N_i(D_K) \setminus \{i\}} x_q$$

$$= x_i \oplus \bigoplus_{j \in V_i \setminus \{i\}} \bigoplus_{k \in N_i(D_K) \setminus \{i\}} x_j \oplus k \oplus \bigoplus_{q \in N_i(D_K) \setminus \{i\}} x_q$$

$$= x_i \oplus \bigoplus_{j \in V_i \setminus \{i\}} x_j \oplus \bigoplus_{k \in N_i(D_K) \setminus \{i\}} x_k \oplus k \oplus \bigoplus_{q \in N_i(D_K) \setminus \{i\}} x_q$$

Any inner vertex $i$ can decode its required message from (6).

IV. RESULTS

Definition 5: (Generalized interlinked cycle cover (GICC) scheme) For any digraph, the GICC scheme finds a set of disjoint GIC sub-digraphs. It then (a) codes each of these GIC sub-digraphs using code construction described in Section III.B, and (b) sends uncoded messages requested by all remaining vertices (i.e., vertices which are not in any of these disjoint GIC sub-digraphs).

We now present a main result of this paper. It is best expressed in terms of saved packets defined as follows:

Definition 6: (Saved packets) The number of packets saved (i.e., $N - \beta(D)$) by transmitting coded symbols (coded packets) rather than transmitting uncoded message packets, is saved packets or simply savings.

Theorem 1: For any digraph $D$, a valid index code of length $\beta_{GICC}(D) = N - \sum_{i=1}^{s} (K_i - 1)$ can be achieved by using the GICC scheme, where $(K_i - 1)$ is the saving in each disjoint $K_i$-GIC sub-digraphs, and $s$ is the number of disjoint GIC sub-digraphs in $D$.

Remark 2: Just as for the cycle cover and the clique cover schemes, the GIC sub-digraphs found by the GICC scheme are not unique. So, finding the best $\beta_{GICC}(D)$ involves optimizing over all choices of disjoint GIC sub-digraphs in $D$.

A. Proof of Theorem

A $K$-GIC sub-digraph $D_K$, has some properties captured in the following lemmas, which we will use to prove Theorem.

Here we consider $T_i$ and $T_j$ as any two distinct trees present in $D_K$.

Lemma 1: If a vertex $v \in V(T_i)$, $v \in V(T_j)$, and $v \notin V_i$, then the set of leaf vertices that fan out from the common vertex $v$ in each tree is a subset of $V_i \setminus \{i,j\}$.

Proof: In a tree $T_i$ (see Fig. 2), for any vertex $v \in V(T_i)$ and $v \notin V_i$, let $L_T(v)$ be a set of leaf vertices that fan out from the vertex $v$. If the vertex $j \in L_T(v)$, then there exists a path from vertex $v$ to $j$ in the tree $T_i$. However in the tree $T_j$, there is a path from vertex $j$ to $v$. Thus in the sub-digraph $D_K$, we obtain a path from vertex $v$ to $j$ (via $T_i$) and vice versa (via $T_j$). As a result, an I-cycle containing $j$ is present. This contradicts the condition 1 (i.e., no I-cycle) for $D_K$. Hence $j \notin L_T(v)$. In other words $L_T(v) \subseteq V_i \setminus \{i, j\}$. Similarly, $L_{T_j}(v) \subseteq V_i \setminus \{i, j\}$.

Lemma 2: If a vertex $v \in V(T_i)$, $v \in V(T_j)$, and $v \notin V_i$, then the out-neighborhood of the vertex $v$ is same in both of the trees, i.e., $N_T^+(v) = N_T^-(v)$.

Proof: Here the proof is done by contradiction. Let us suppose that $N_T^+(v) \neq N_T^-(v)$.

For this proof we refer to Fig. 2. This proof has two parts. In the first part, we prove $L_{T_i}(v) = L_{T_j}(v)$, and then prove $N_T^+(v) = N_T^+(v)$ in the second part.

From Lemma 1, $L_{T_i}(v)$ is a subset of $V_i \setminus \{i, j\}$. Now pick a vertex $c$ belongs to $V_i \setminus \{i, j\}$ such that $c \in L_{T_i}(v)$ but $c \notin L_{T_j}(v)$ (such $c$ exists since we suppose that $L_{T_i}(v) \neq L_{T_j}(v)$). In tree $T_i$, there exists a directed path from the vertex $i$ which includes the vertex $v$, and ends at the leaf vertex $c$. Let this path be $P_{i \rightarrow c}(T_i)$. Similarly, in tree $T_j$, there exists a directed path from the vertex $j$, which doesn’t include the vertex $v$ (since $c \notin L_{T_j}(v)$), and ends at the leaf vertex $c$. Let this path be $P_{j \rightarrow c}(T_j)$. However, in the digraph $D_K$, we can also obtain a directed path from the vertex $j$ which passes through the vertex $v$ (via $T_j$), and ends at the leaf vertex $c$ (via $T_i$). Let this path be $P_{j \rightarrow c}(D_K)$. The paths $P_{j \rightarrow c}(T_j)$ and $P_{j \rightarrow c}(D_K)$ are different which indicates the existence of multiple P-paths from the vertex $j$ to $c$ in $D_K$, thus contradict the condition 2 for a $D_K$. Consequently, $L_{T_i}(v) = L_{T_j}(v)$.

Now we pick a vertex $b$ such that, without loss of generality, $b \in N_T^+(v)$ but $b \notin N_T^+(v)$ (such $b$ exists since we assumed that $N_T^+(v) \neq N_T^+(v)$). Furthermore, we have two cases for $b$, which are (case 1) $b \in L_{T_j}(v)$, and (case 2) $b \notin L_{T_j}(v)$. Case 1 is addressed in the first part of this proof. On the other hand, for case 2, we pick a leaf vertex $d \in L_{T_j}(b)$ such that

1As $D_K = \bigcup_i T_i$, a path present in any $T_i$ also present in $D_K$. 

Fig. 2. Directed rooted trees $T_i$ and $T_j$ with roots $i$ and $j$ respectively, and a vertex $v$ in common. Here, we have used straight arrow to indicate an arc, and curly arrow to indicate a path.
there exists a path (see Fig. 4) that starts from \( v \) followed by \( b \),
and ends at \( d \), i.e., \( \langle v, b, \ldots, d \rangle \) exists in \( T_i \). A path \( \langle j, \ldots, v \rangle \)
exists in \( T_j \). Thus a path \( \langle j, \ldots, v, b, \ldots, d \rangle \) exists in \( D_K \).
From the first part of the proof, we have \( L_{T_i}(v) = L_{T_j}(v) \), so \( d \in L_{T_j}(v) \).
Now in \( T_j \), there exists a path from \( j \) to \( d \), which
includes vertex \( v \) followed by a vertex \( e \) such that \( e \in N^+_j(v) \)
and \( e \neq b \) (as \( b \not\in N^+_j(v) \)), and the path ends at \( d \), i.e., \( \langle j, \ldots, v, e, \ldots, d \rangle \) which is different from \( \langle j, \ldots, v, b, \ldots, d \rangle \).
So multiple \( P \)-paths are observed at \( d \) from \( j \). This contradicts condition 2 for a \( D_K \).
Consequently, \( N^+_i(v) = N^+_j(v) \).

**Lemma 3**: If a vertex \( v \in V(T_i) \) such that \( v \notin V_i \), then \( N^+_i(v) = N^+_D(v) \).

**Proof**: For any \( v \in V(T_i) \) from Lemma 2 \( N^+_i(v) = N^+_I(v) \) for all \( \{ j : v \in T_j \} \). Since \( D_K = \bigcup_{i} T_i \), vertex \( v \)
must have the same out-neighborhood in \( D_K \) as well.

Now we prove a lemma on the number of savings achieved by the \( GICC \) scheme for a \( D_K \).

**Lemma 4**: Consider a \( K \)-GIC sub-digraph \( D_K \), with each message of \( t \)-bits, for any \( K \geq 1 \) and any \( t \geq 1 \). The total number savings achieved by the \( GICC \) scheme is \( K - 1 \), i.e.,
\( N - \beta_{GICC}(D_K) = K - 1 \).

**Proof**: Subtracting \( \beta_{GICC}(D_K) \) of (3) from \( N \) we get
\( N - \beta_{GICC}(D_K) = N - (N - K + 1) = K - 1 \).

**Proof of Theorem 7**: For any digraph \( D \) containing \( \psi \)
disjoint \( GIC \) sub-digraphs, a saving of \( K - 1 \) is obtained in each \( D_{K_i} \) (from Lemma 3), where \( i \in \{1, 2, \ldots, \psi \} \). Now, the total saving is the summation of savings in all disjoint \( GIC \) sub-digraphs, i.e., \( \sum_{i=1}^{\psi} K_i = 1 \). Hence, \( \beta_{GICC}(D) = N - \sum_{i=1}^{\psi} K_i(1) \).

**B. GIC sub-digraphs include ICC sub-digraphs**

**Theorem 2**: The GIC sub-digraph includes the ICC sub-digraph as a special case.

**Proof**: An ICC sub-digraph \( D_{ICC} \), is defined as follows [10]:

1. It has \( k \) disjoint paths, \( P_i \triangleq \langle v^i_1, v^i_2, \ldots, v^i_{n_i} \rangle \), for each \( i \in \{1, 2, \ldots, k\} \), where each \( P_i \) has \( n_i \) number of vertices.
2. For any distinct \( i, j \in \{1, 2, \ldots, k\} \), there is a path from \( v^i_{n_i} \in V(P_i) \) to some \( v^j \in V(P_j) \), denoted as \( \langle v^i_{n_i}, v^j \rangle \). Denote the sub-path \( P_{i,j} \triangleq \langle v^i_{n_i}, v_{ij}, v^j \rangle \), where each \( P_{i,j} \) has \( n_{ij} \) number of vertices.
3. The set of vertices in all \( P_i \) and in all \( P_{i,j} \) are mutually disjoint.
4. Each first vertex \( v^j \) of \( P_j \) has at least one in-degree.

Select \( V_i = \{ v^i_1, v^i_2, \ldots, v^i_{n_i} \} \). Point 2 guarantees that, for any ordered pair \( (i, j) \), \( i \neq j \) and \( i, j \in \{1, 2, \ldots, k\} \), there exists a path from \( v^i_{n_i} \) to \( v^j \). Now we will show that the ICC sub-digraph with the chosen \( V_i \), satisfies the two conditions for it to be a GIC sub-digraph.

By construction of an ICC digraph, we have the following: For any \( v^i_{n_i} \in V_i \), all paths from \( v^i_{n_i} \) must go through some \( v^j \in V(P_j) \) for some \( j \), and then \( v^i_{n_j} \) before returning to \( v^i_{n_i} \). Therefore, there is no \( I \)-cycle.

Note that all vertices of \( V(D_{ICC}) \setminus V_i \) each has out-degree one. Thus it follows from points 1 and 2 that there is only one \( P \)-path between any \( i, j \in V_i \).

Since there is only one \( P \)-path from \( i \) to \( j \) for any \( i, j \in V_i \), every \( P_{i,i} \), which is part of the \( P \)-path from \( i \) to \( j \), must be a sub-digraph of tree \( T_i \). For every \( j \in V_i \), point 4 dictates that the entire \( P_j \) must be part of the \( P \)-path from some \( i \) to \( j \), and \( P_j \) is hence a sub-digraph of a tree \( T_i \). We have included all vertices and arcs in the construction of the trees in \( D_{GIC} \). Consequently, \( D_{ICC} = D_{GIC} \).

**C. The GICC scheme includes the ICC scheme as a special case**

**Theorem 3**: The GICC scheme includes the ICC scheme.

**Proof**: In an ICC sub-digraph \( D_{ICC} \), select \( V_i = \{ v^i_1, v^i_2, \ldots, v^i_{n_i} \} \). Now the coded symbols \( w^i = \bigoplus_{j=1}^{n_i} x^i_{n_j} \) of the ICC is the same as the coded symbol \( w^i \) of the GICC. Now for the remaining vertices (of set \( V(D_{ICC}) \setminus V_i \)), the coded symbols are simply bitwise XOR of messages requested by the respective vertex and its out-neighborhood, which are same as \( w_j, j \in V(D_K) \setminus V_i \) of the GICC.

**D. The GICC scheme includes the cycle cover and clique cover schemes as special cases**

**Corollary 1**: The GICC scheme includes the cycle cover and the clique cover schemes as special cases.

**Proof**: The ICC scheme includes the cycle cover scheme and the clique cover scheme as its special cases [10]. From Theorem 2, the GICC scheme includes ICC scheme as its special case.

**E. The GICC scheme is optimal for a class of digraphs**

Here, at first, we prove one lemma that will help to prove optimality of our proposed GICC scheme.

**Lemma 5**: In a \( K \)-GIC sub-digraph \( D_K \), if there is no cycle containing only vertices in \( V(D_K) \setminus V_1 \) (i.e., non-inner vertices), then any cycle in \( D_K \) must include at least two inner vertices.

**Proof**: It follows directly from the property of \( D_K \) that a cycle can not be formed by including only one inner vertex because these kind of cycles are \( I \)-cycles.

**Definition 7** (Maximum acyclic induced subgraph (MAIS)) For a digraph \( D \), an induced acyclic sub-digraph with the largest number of vertices is called the maximum acyclic induced subgraph (MAIS). We denote the order of an MAIS by \( \text{MAIS}(D) \).

It has been shown [2] that for any \( D \) and \( t \),
\[
\text{MAIS}(D) \leq \beta^*(D) \leq \beta^t(D) \leq \beta_0(D).
\] (7)

**Theorem 4**: For any \( t \geq 1 \), the scalar linear index code given by GICC scheme is optimal for two classes of digraphs: The digraph is a \( K \)-GIC (referred to as a \( K \)-GIC digraph) with

- (Case 1) no cycle among the non-inner vertices, or
• (Case 2) \( M \geq 1 \) disjoint cycles among non-inner vertices, and we can group the inner vertex set \( V_1 \) in to \( M + 1 \) sub-sets such that each of them forms a disjoint GIC sub-digraph of case 1, and such GIC sub-digraphs are also disjoint from the \( M \) cycles among non-inner vertices.

For these digraphs, \( \beta_{\text{GICC}}(D) = \beta^*(D) = \beta^*_i(D) \).

Proof: We will show that the MAIS lower bound (7) is tight for all \( t \). Denote the digraph by \( D_K \). For \( K = 1 \), the digraph contains only one vertex, and MAIS\((D_1) = 1 \). For \( K \geq 2 \), we have the following:

(Case 1) From Lemma 5, every cycle must include at least two inner vertices, thus if we remove \( K - 1 \) inner vertices, then the digraph \( D_K \) becomes acyclic. Thus,

\[
\text{MAIS}(D_K) \geq (N - K + 1). \tag{8}
\]

Now if we remove any \( K - 2 \) vertices, there still exist at least two inner vertices which can form a cycle by Lemma 5. Thus, least possible number of removal of vertices is \( K - 1 \), i.e.,

\[
\text{MAIS}(D_K) \leq (N - K + 1). \tag{9}
\]

Now from (8) and (9), we get MAIS\((D_K) = N - K + 1 \). From Theorem 1, \( \beta_{\text{GICC}}(D_K) = N - K + 1 \). Hence \( \beta_{\text{GICC}}(D_K) = \beta^*(D_K) = \beta^*_i(D_K) = N - K + 1 \).

(Case 2) A \( D_K \) can be viewed in two ways. The first way is considering the whole \( D_K \) as a \( K \)-GIC digraph. The second way is considering induced sub-digraphs of \( D_K \) which consists of the following:

1) \( M \) total disjoint cycles consisting of \( N_A \) (0 \( N_A < N - K \)) non-inner vertices (if \( N_A = 0 \) or 1, then \( M = 0 \), which is case 1).
2) \( M + 1 \) disjoint GIC sub-digraphs each of \( K_i \) inner vertices in such a way that \( \sum_{i=1}^{M+1} K_i = K \). Note that each GIC sub-digraph is also disjoint from all \( M \) cycles among non-inner vertices.
3) Total remaining of \( N_B \) (\( N_B = N - N_A - K \)) non-inner vertices (which are not included in \( M \) cycles among non-inner vertices, and \( M + 1 \) GIC sub-digraphs).

Now we will show that both ways of looking at \( D_K \) are equivalent in the sense of the index code length generated from our proposed scheme, both equal to MAIS\((D_K)\). We prefer the second way of viewing \( D_K \) for our proof since it makes the process easier to find the MAIS lower bound.

From Theorem 1 the number of coded symbols,

\[
\beta_{\text{GICC}}(D_K) = N - K + 1. \tag{10}
\]

For the partitioned \( D_K \) (looking in the second way), the total number of coded symbols will be the summation of length of coded symbols for each of \( M \) disjoint cycles (each cycle has saving equal to one), \( M + 1 \) disjoint GIC sub-digraphs (each of GIC sub-digraphs has savings equal to \( K - 1 \)), and \( N_B \) uncoded symbols for remaining non-inner vertices, i.e.,

\[
\beta(D_K) = (N_A - M) + \sum_{i=1}^{M+1} (N_i - K_i + 1) + N_B
\]

\[
= N_A - M + \sum_{i=1}^{M+1} N_i - K + M + 1 + N_B
\]

\[
= N - K + 1. \tag{11}
\]

From (10) and (11), \( \beta_{\text{GICC}}(D_K) = \beta(D_K) \).

Now for \( D_K \) (looking in our second way), if we remove one vertex from each of \( M \) cycles among non-inner vertices (\( M \) removal in total), and remove \( K_i - 1 \) from each of \( M + 1 \) GIC sub-digraphs \((\sum_{i=1}^{M+1} (K_i - 1) = K - M - 1)\), i.e., total removal of \( K - 1 \), then the digraph becomes acyclic. Thus,

\[
\text{MAIS}(D_K) \geq (N - K + 1). \tag{12}
\]

To calculate a lower bound to MAIS\((D_K)\), if we remove any \( K - 2 \) vertices, then there can still exists a cycle. This is because we must remove \( M \) vertices each one from \( M \) disjoint cycles formed among non-inner vertices (cycle remains if less than \( M \) removal), and we can not remove less than \( K_i - 1 \) inner vertices in each of \( M + 1 \) disjoint GIC sub-digraphs, otherwise there can remain a pair of inner vertices in any \( K_i \)-GIC sub-digraph which can form a cycle by lemma 5. Therefore, if we remove total of \((M + \sum_{i=1}^{M+1} (K_i - 1) - 1) = K - 2\), then the \( D_K \) is not acyclic. Thus,

\[
\text{MAIS}(D_K) \leq (N - K + 1). \tag{13}
\]

From (12) and (13), we get MAIS\((D_K) = N - K + 1 \), and considering (10), we get \( \beta_{\text{GICC}}(D_K) = \text{MAIS}(D_K) \). Thus \( \beta_{\text{GICC}}(D_K) = \beta^*(D_K) = \beta^*_i(D_K) = N - K + 1 \).

Conjecture 1: For any \( t \geq 1 \), the scalar index codes given by GICC scheme is optimal for any GIC digraph \( D_K \).

If MAIS\((D_K) < \beta_{\text{GICC}}(D_K)\), we conjecture that we can always find disjoint GIC sub-digraphs such that the summation of all number of coded symbols required for each GIC sub-digraph is equal to MAIS\((D_K)\).

V. COMPARISON WITH EXISTING TECHNIQUES

A. The GICC scheme can outperform existing techniques including composite coding and interlinked cycle cover schemes

The GIC digraph shown in Fig. 3(a) has \( K = 4 \) and represented as \( D_4 \). The coded symbols from the GICC scheme
is of length $\beta_{GICC}(D_4) = 3$, which are \{1 \oplus 2 \oplus 3 \oplus 4, 5 \oplus 2 \oplus 3, 6 \oplus 3 \oplus 4\}. The lower bound for the digraph $D_4$, by existing techniques are; (a) composite coding \cite{11}, $\beta_{CC}(D_4) = 3.5$, (b) local chromatic number \cite{9}, $\beta_{LC}(D_4) = 4$, (c) fractional partial clique cover \cite{12}, $\beta_{FPCC}(D_4) = 4$, (d) interlinked cycle cover \cite{10}, $\beta_{ICC}(D_4) = 4$, (e) clique cover \cite{1}, $\beta_{CL}(D_4) = 5$, (f) cycle cover \cite{2}, \cite{4}, \cite{7}, $\beta_{CY}(D_4) = 4$. All of these lower bounds are strictly greater than $\beta_{GICC}(D_4)$. Moreover, maximum distance separable (MDS) codes are used by some of the existing schemes (e.g., partial clique cover, local chromatic number approach), which requires $t$ to be sufficiently long. Composite coding technique requires $t$ to be infinitely long. The GIC scheme is valid for any $t$.

B. Performance improvement of the GIC scheme over existing schemes

We outline a class of GIC digraphs with $K$ number of inner vertices, and $(K - 1)$ number of non-inner vertices, so that $N = 3K - 2$. Without loss of generality, say $V_1 = \{1, 2, \ldots, K\}$, and the non-inner vertex set is $\{K + 1, \ldots, N\}$. Now for $i \in \{2, \ldots, K - 1\}$, every vertex $i$ knows messages requested by vertices $K + i$ and $3K - i$. The vertex $K + i$ knows all messages requested by vertices of $\{i + 1, \ldots, K\}$, and the vertex $3K - i$ knows all messages requested by vertices of $\{1, \ldots, i - 1\}$. The first inner vertex (i.e., 1) knows message requested by vertex $K + 1$, and $K + 1$ knows all messages requested by $V_1 \setminus \{1\}$. Similarly the last inner vertex (i.e., $K$) knows message requested by vertex $2K$, and $2K$ knows all messages requested by $V_1 \setminus \{K\}$.

For $K > 3$, a digraph $D_K$ of this class, is not an ICC digraph, and its complement digraph (underlying graph is a complete graph) has chromatic number, $\chi(D_K) = N = 3K - 2 = \beta_{CL}(D_K)$, and local chromatic number, $\chi_L(D_K) = N - 1 = 3K - 3 = \beta_{LC}(D_K)$. Also, the digraph $D_K$ has $\beta_{GICC}(D_K) = \frac{2N + 1}{3} = 2K - 1$, and $\beta_{CC}(D_K) = \beta_{GICC}(D_K) + \left\lceil \frac{K}{3} \right\rceil - 1 = 2K + \left\lceil \frac{K}{3} \right\rceil - 2$. The gap, $\beta_{CL}(D_K) - \beta_{GICC}(D_K) = K - 2$, so for $K > 2$, the gap grows linearly with $K$. Similarly, $\beta_{CL}(D_K) - \beta_{ICC}(D_K) = K - 1$. The digraph in Fig. 3(b) belongs to this class of digraphs with $K = 4$.

VI. ACKNOWLEDGMENT

Authors would like to acknowledge Fatemeh Arbabjolfaei, University of California, San Diego for providing composite coding lower bound for the digraph in Fig. 3(a).

REFERENCES

[1] Y. Birk and T. Kol, “Informed-source coding-on-demand (ISCOD) over broadcast channels,” in Proc. IEEE INFOCOM 98, vol. 3, San Francisco, CA, Mar. 1998, pp. 1257–1264.

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

[3] L. Ong, “Linear codes are optimal for index-coding instances with five or fewer receivers,” in Proc. IEEE International Symposium on Information Theory(ISIT), June 2014, pp. 491–495.

[4] M. A. R. Chaudhry, Z. Asad, A. Sprintson, and M. Langberg, “On the complementary index coding problem,” in Proc. IEEE International Symposium on Information Theory(ISIT), July 2011, pp. 224–248.