Secure Network and Index Coding Equivalence: The Last Piece of the Puzzle

Lawrence Ong
The University of Newcastle
Email: lawrence.ong@newcastle.edu.au

Badri N. Vellambi
University of Cincinnati
Email: badri.vellambi@uc.edu

***Abstract*—An equivalence was shown between network coding and index coding. The equivalence allows for a network code for any given network-coding instance to be translated to an index code for a suitably constructed index-coding instance, and vice versa. The equivalence also holds for the opposite direction. A secure version of the equivalence in the presence of eavesdroppers was proven for the case where there is no decoding error and no information leakage to the eavesdroppers. For the case of non-zero decoding error and non-zero leakage, three out of the four directions required for an equivalence were proven. This paper proves the last direction, thereby completing the equivalence between secure network coding and secure index coding.**

I. Introduction

Index coding [1] studies single-hop noiseless broadcast communications, where a single sender transmit data to multiple receivers. Each receiver knows some messages and demands a certain message. This simple setup has been shown to be equivalent to some other problems, for example, distributed storage [2] and guessing games on directed graph [3].

Another non-trivial equivalence is between index coding and network coding [4]. Some equivalence results have already established between multiple-unicast network coding and multiple-multicast network coding [5] and secure network coding [6]. An equivalence between network coding and index coding is intriguing as it reduces a multi-hop communication problem of network coding to a single-hop setting.

An equivalence between index coding and network coding was first established for linear codes [7] and then for non-linear codes in general [8]. This is a *code* equivalence: any network-coding instance can be efficiently mapped to a suitably constructed index-coding instance, such that a network code for the former can be translated to an index code for the latter (and vice versa), preserving the message sizes and the probability of decoding error. This is also true in the other direction that maps any index-coding instance to a suitably constructed network-coding instance. The proof essentially requires four code translations, two for each direction of the mapping.

An extension to this equivalence is to bridge secure index coding [9] and secure network coding [10]. In the secure version, in addition to decodability of the receivers, transmission codes must also meet security constraints: to prevent passive eavesdroppers listening to (part of) the transmissions from getting information about specified messages.

Toward establishing this secure equivalence, a mapping of the eavesdroppers was proposed by Ong et al. [11]. This mapping preserves the mapping of receivers previously established for the non-secure equivalence [8], and specifies how eavesdroppers in one instance are mapped to the other. Using this mapping, it was shown [11] that the code translation proposed for the non-secure case also satisfy the security constraints, if the decoding for all receivers is perfect, and the information leakage to the eavesdroppers is zero.

For the case of non-zero decoding error and non-zero leakage, three of the four code translation results have been obtained [12]. The last direction yet to be proven must translate an index code to a network code. The difficulty in proving this direction stems from the use of the code translation for the non-secure case, in which a variable in the network code needs to be pre-determined. This variable is translated from the broadcast function (a function of all messages) of the index code. However, none of the nodes can execute this function for the network code.

For the non-secure equivalence, a good candidate can be pre-selected for this variable. However, for the secure case, the search for a value that simultaneously guarantees both decoding and security constraints was unsuccessful.

In this paper, we propose a new way to pre-determine the value of this variable. Instead of fixing one value, we cycle through different values for the network code according to their probability of occurrence in the index code. We prove that this method guarantees both the decoding and the security constraints to be independent of the message size. Together with the existing results [12], this paper completes an equivalence between secure network and index coding.

II. Problem Definition and Notation

For a set $S = \{s_1, s_2, \dots s_{|S|}\}$, let $X_S := (X_{s_1}, X_{s_2}, \dots, X_{s_{|S|}})$. Consider a directed graph $G = (V, E)$ with a node set V and an edge set E. We will refer edges as links subsequently. For an edge $e = (u \to v) \in E$, where $u, v \in V$, its tail is $\texttt{tail}(e) := u$, and its head is $\texttt{head}(e) := v$. For any node $v \in V$, the set of incoming edges is denoted by $\texttt{in}(v) := \{e \in E : \texttt{head}(e) = v\}$, and the set of outgoing edges by $\texttt{out}(v) := \{e \in E : \texttt{tail}(e) = v\}$. Define $\mathbb{R}_0^+ := [0, \infty)$, $\mathbb{Z}^+ := \{1, 2, \dots\}$, and $[a] := \{1, 2, \dots, a\}$, for any $a \in \mathbb{Z}^+$.

A. Secure network coding

1) Network-coding instances: A secure network-coding instance [10], denoted by $\mathbb{N} = (G, C, W)$, is defined as follows:

- $G = (V, E)$ is an acyclic directed graph with a node set V and an link set E. Each link $e \in E$ has a capacity $c_e \in \mathbb{R}_0^+$, where $\texttt{tail}(e)$ can send a message $X_e \in [2^{\lfloor c_e n \rfloor}]$ to $\texttt{head}(e)$ without any error over $n \in \mathbb{Z}^+$ link uses.
- $C = (S, O, D)$ is the connection requirement. S contains message indices, where the messages are $\{X_s : s \in S\}$. $O(s) \in V$ is the originating node for message X_s. $D(s) \subsetneq V$ is the set of nodes that require message X_s.
- $W = \big((A_z, B_z) : z \in Z\big)$ defines the eavesdroppers. Z contains the eavesdropper indices. Each eavesdropper $z \in Z$ observes messages X_{B_z} on links $B_z \subseteq E$ and tries to reconstruct messages X_{A_z} for some $A_z \subseteq S$.

2) Deterministic network codes: Consider n uses of each link. Let the messages $\{X_s : s \in S\}$ be mutually independent, and each X_s be uniformly distributed over $[M_s]$ for some $M_s \in \mathbb{Z}^+$. A deterministic network code consists of the following:

- An deterministic encoding function e_e for each link $e \in E$, which takes in $(X_{\texttt{in}(\texttt{tail}(e))}, X_{O^{-1}(\texttt{tail}(e))})$ and outputs a message $X_e \in [2^{\lfloor c_e n \rfloor}]$ on link e. Here, as an abuse of notation, we let $O^{-1}(v)$ be the indices of the messages that originate from node v.
- A decoding function d_v for each node $v \in V$, which takes in $(X_{\texttt{in}(v)}, X_{O^{-1}(v)})$ and outputs an estimate $X^{(v)}_{D^{-1}(v)}$ of the messages $X_{D^{-1}(v)}$ that it requires. Here $D^{-1}(v)$ is the set of indices of messages whose destinations include v.

Here, n is referred to as the blocklength of the code.

3) Randomised network codes: Consider a random key K_v for each node v, which is independently and uniformly distributed over $[\kappa_v]$ for some $\kappa_v \in \mathbb{Z}^+$. A randomised network code is similar to a deterministic network code, except that each edge encoding function e_e is a deterministic function that maps $(X_{\texttt{in}(\texttt{tail}(e))}, X_{O^{-1}(\texttt{tail}(e))}, K_{\texttt{tail}(e)})$ to X_e.

4) Decodability: A network code is said to have a probability of error of at most $\epsilon \in \mathbb{R}_0^+$ if and only if (iff)

$$1 - P_{\mathrm{e}} := \Pr\left\{X^{(v)}_{D^{-1}(v)} = X_{D^{-1}(v)} \text{ for all } v \in V\right\} \geq 1 - \epsilon. \quad (1)$$

5) Leakage: A code network code is said to be have at most $\eta \in \mathbb{R}_0^+$ leakage iff

$$I(X_{A_z}; X_{B_z}) \leq \eta, \qquad \text{for all } z \in Z. \quad (2)$$

6) Secure network-coding feasibility: A secure network-coding instance $\mathbb{N}$ is said to be (M_S, ϵ, η, n)-feasible iff there exists a network code of blocklength n that has at most ϵ probability of error and η leakage. Further, $\left(\frac{\log_2 M_s}{n} : s \in S\right)$ is commonly referred to as the message rate tuple.

B. Secure index coding

1) Index-coding instances: A secure index-coding instance [9], denoted by $\mathbb{I} = (\widehat{S}, \widehat{T}, \{(\widehat{W}_t, \widehat{H}_t) : t \in \widehat{T}\}, \widehat{W})$, is defined as follows:

- $\widehat{S}$ is the message index set.
- $\widehat{T}$ is the receiver index set.
- $\widehat{W}_t \subseteq \widehat{S}$ is the subset of indices of messages required by receiver $t \in \widehat{T}$.
- $\widehat{H}_t \subsetneq \widehat{S}$ is the subset of indices of messages known a priori (known as side information) to receiver $t \in \widehat{T}$.
- $\widehat{W} = ((\widehat{A}_z, \widehat{B}_z) : z \in \widehat{Z})$ defines the eavesdroppers. Each eavesdropper $z \in \widehat{Z}$ has access to the codeword broadcast by the sender and messages indexed by $\widehat{B}_z \subsetneq \widehat{S}$, and attempts to reconstruct messages indexed by $\widehat{A}_z \subseteq \widehat{S}$.

2) Deterministic index codes: Let the messages $\{\widehat{X}_s : s \in \widehat{S}\}$ be mutually independent, and each $\widehat{X}_s$ be uniformly distributed over $[\widehat{M}_s]$ for some $\widehat{M}_s \in \mathbb{Z}^+$. A deterministic index code consists of the following:

- A deterministic encoding (or broadcast) function for the sender: $\widehat{X}_{\mathrm{b}} = \widehat{\mathsf{e}}(\widehat{X}_{\widehat{S}}) \in [2^{\widehat{n}}]$, for some $\widehat{n} \in \mathbb{Z}^+$.
- A decoding function $\widehat{\mathsf{d}}_t$ for each receiver $t \in \widehat{T}$ that takes in $(\widehat{X}_{\mathrm{b}}, \widehat{X}_{\widehat{H}_t})$, and outputs an estimate $\widehat{X}^{(t)}_{\widehat{W}_t}$ of the messages $\widehat{X}_{\widehat{W}_t}$ that t requires.

Here $\widehat{n}$ is referred to as the blocklength.

3) Randomised index codes: A randomised index code is similar to deterministic index codes except that the sender's encoding function takes in an independent random key $\widehat{Z} \in [\widehat{\kappa}]$ in addition to $\widehat{X}_{\widehat{S}}$, for some $\widehat{\kappa} \in \mathbb{Z}^+$.

4) Decodability: As with network coding, an index code has a probability of error of at most $\epsilon \in \mathbb{R}_0^+$ iff

$$1 - \widehat{P}_{\mathrm{e}} := \Pr\left\{\widehat{X}^{(t)}_{\widehat{W}_t} = \widehat{X}_{\widehat{W}_t} \text{ for all } t \in \widehat{T}\right\} \geq 1 - \epsilon. \quad (3)$$

5) Leakage: An index code has at most $\eta \in \mathbb{R}_0^+$ leakage iff

$$I(\widehat{X}_{\widehat{A}_z}; \widehat{X}_{\mathrm{b}}, \widehat{X}_{\widehat{B}_z}) \leq \eta, \qquad \text{for all } z \in \widehat{Z}. \quad (4)$$

6) Secure index-coding feasibility: A secure index-coding instance $\mathbb{I}$ is said to be $(\widehat{M}_{\widehat{S}}, \epsilon, \eta, \widehat{n})$-feasible iff there exists an index code of blocklength $\widehat{n}$ with at most ϵ probability of error and η leakage. Here, $\left(\frac{\log_2 \widehat{M}_s}{\widehat{n}} : s \in \widehat{S}\right)$ is the message rate tuple.

C. Problem Statement

As mentioned earlier, an equivalence between secure network coding and secure index coding has been proven except for one direction. Here, we state the direction that is left to be proven: Any secure network-coding instance $\mathbb{N}$ can be mapped to a secure index-coding instance $\mathbb{I}$, such that a code for $\mathbb{I}$ can be translated to a code for $\mathbb{N}$ with comparable ϵ and η.

We first present the mapping of problem instances, followed by the translation of codes.

III. Instance Mapping

We restate the mapping [12] from $\mathbb{N}$ to $\mathbb{I}$, for the paper to be self-contained.

A. Network-to-index coding mapping

Consider a secure network-coding instance $\mathbb{N}$. Let the message indices be $S = [k]$ and the node indices be $V = [\ell]$. Without loss of generality, each message is requested by at least one destination. Otherwise, it can be removed without affecting the decodability and security of any code.

$\mathbb{N} = (G, C, W)$ is first mapped to an *augmented* secure network-coding instance $\mathbb{N}'$. This step converts any possibly randomised secure network code to a deterministic network code. Then, $\mathbb{N}'$ is mapped to $\mathbb{I}$. The details of the mappings are as follows:

1) Augmented secure network coding: $\mathbb{N}' = (G', C', W')$ is the same as the original $\mathbb{N}$, except that it has $|V|$ more messages. These dummy messages takes the role of random keys in $\mathbb{N}$ and are not required by any destination node.

- $G' = (V, E) = G$, with the same link capacities c_E.
- $C' = (S', O', D')$:
  - $S' = [k + \ell]$ consists of the k original messages and ℓ new dummy messages. Each dummy message X'_{k+i}, $i \in [\ell]$, is independently and uniformly distributed over $[\kappa_i]$. That means, message alphabet sizes are $M'_i = M_i$ for each $i \in [k]$, and $M'_{k+i} = \kappa_i$ for each $i \in [\ell]$.
  - $O'(i) = O(i)$ for $i \in [k]$, and $O'(k + i) = i$ for $i \in [\ell]$.
  - $D'(i) = D(i)$ for $i \in [k]$, and $D'(k + i) = \emptyset$ for $i \in [\ell]$.
- $W' = ((A_z, B_z) : z \in Z) = W$.

By construction, $X'_{[k+\ell]}$ and $(X_{[k]}, K_{[\ell]})$ have the same distribution. So, there is a bijective map between deterministic or randomised network codes for $\mathbb{N}$ and deterministic network codes for $\mathbb{N}'$.

Denote the set of nodes in $\mathbb{N}'$ (and also $\mathbb{N}$) that are the destinations for some source messages by $U := \{j \in [\ell] : j \in D'(i) \text{ for some } i \in [k]\}$.

2) Network-to-index coding mapping: Next, we map $\mathbb{N}'$ to a secure index-coding instance $\mathbb{I}$, which consists of $|E|$ more messages and $|E|$ more receivers compared to $\mathbb{N}'$.

- $\widehat{S} = [k + \ell] \cup E$ consists of $[k + \ell] \cup E$ independent messages, where we set $\widehat{M}_i = M'_i$ for each $i \in [k + \ell]$, and $\widehat{M}_e = 2^{\lfloor c_e n \rfloor}$ for each $e \in E$.
- $\widehat{T} = \{\widehat{t}_i\}_{i \in U} \cup \{\widehat{t}_e\}_{e \in E}$ consists of a receiver for each destination node in $\mathbb{N}'$ and one for each link in $\mathbb{N}'$.
- For each $\widehat{t}_e \in \widehat{T}$ where $e \in E$, we set $\widehat{H}_{\widehat{t}_e} = \mathtt{in}(\mathtt{tail}(e)) \cup O'^{-1}(\mathtt{tail}(e))$, and $\widehat{W}_{\widehat{t}_e} = \{e\}$.
- For each $\widehat{t}_i \in \widehat{T}$ where $i \in U$, we set $\widehat{H}_{\widehat{t}_i} = \mathtt{in}(i) \cup O'^{-1}(i)$, and $\widehat{W}_{\widehat{t}_i} = D'^{-1}(i)$.
- $\widehat{W} = ((A_z, B_z) : z \in Z) = W' = W$, that is, for each $z \in \widehat{Z}$, $\widehat{B}_z = B_z$, and $\widehat{A}_z = A_z$.

Note that, by construction, for every $e \in E$, the message $\widehat{X}_e$ is wanted by receiver $\widehat{t}_e$.

IV. Main Result

For $a \in [0, 1]$, let $H_\mathrm{b}(a) := -a \log_2 a - (1 - a) \log_2(1 - a)$ be the binary entropy function. This paper proves the last code translation direction for an equivalence between secure network coding and secure index coding:

Theorem 1: Consider a secure network-coding instance $\mathbb{N}$ and the mapped index-coding instance $\mathbb{I}$ according to Section III. For any $n \in \mathbb{Z}^+$, $M_{[k+\ell]} \in (\mathbb{Z}^+)^{(k+\ell)}$, and $\epsilon, \eta \in \mathbb{R}_0^+$, if $\mathbb{I}$ is $(M_{[k+\ell]\cup E}, \epsilon, \eta, \widehat{n})$-feasible with a deterministic index code, then $\mathbb{N}$ is $(M_{[k]}, 2\epsilon, \eta + 2H_\mathrm{b}(\epsilon), n)$-feasible, where $\widehat{n} = \sum_{e \in E} \lfloor c_e n \rfloor$, and $M_e = 2^{\lfloor c_e n \rfloor}$ for each $e \in E$.

Proof: See Section VI. ∎

This result significantly improves upon that by Ong et al. [12] as the error probability and leakage constraints of the translated code are now independent of n. This means we can now also map message rates achievable in the Shannon sense, which require diminishing error and leakage as the message sizes increase.

For the network-to-index instance mapping considered here, the other direction of code translation was proven as follows [12]: For any $n \in \mathbb{Z}^+$, $M_{[k]} \in (\mathbb{Z}^+)^k$, and $\epsilon, \eta \in \mathbb{R}_0^+$, if $\mathbb{N}$ is $(M_{[k]}, \epsilon, \eta, n)$-feasible, then $\mathbb{I}$ is $(M_{[k+\ell]\cup E}, \epsilon, \eta, \widehat{n})$-feasible with a deterministic index code, where $\widehat{n} = \sum_{e \in E} \lfloor c_e n \rfloor$, and $M_e = 2^{\lfloor c_e n \rfloor}$, for each $e \in E$.

Remark 1: For both directions of code translation, deterministic index codes suffice.

V. An Important Property of the Broadcast Message

For Theorem 1, we only need to consider deterministic (index) codes for $\mathbb{I}$. That means the broadcast message $\widehat{X}_\mathrm{b}$ is a deterministic function of all messages $\widehat{X}_{[k+\ell]\cup E}$. Central to the proof of the equivalence is the following property of the broadcast message:

Proposition 1: Fix any broadcast message $\widehat{x}_\mathrm{b} \in [2^{\widehat{n}}]$ and any realisation $\widehat{x}_{[k+\ell]}$. If all receivers $\widehat{T}$ can decode their requested messages correctly, then there can be at most one realisation $\widehat{x}_E$ for which $\widehat{\mathsf{e}}(\widehat{x}_{[k+\ell]}, \widehat{x}_E) = \widehat{x}_\mathrm{b}$.

Proposition 1 was proven for a slightly different network-to-index coding mapping [8], which includes an additional receiver $\widehat{t}_\mathrm{all}$ in $\mathbb{I}$, where $\widehat{H}_{\widehat{t}_\mathrm{all}} = [k + \ell]$ and $\widehat{W}_{\widehat{t}_\mathrm{all}} = E$. We will show that the proposition remains true even without $\widehat{t}_\mathrm{all}$, by taking into account that the graph G (which was defined for $\mathbb{N}$ from which $\mathbb{I}$ has been mapped) is acyclic.

Proof of Proposition 1: Fix any $\widehat{x}_{[k+\ell]}$ and $\widehat{x}_\mathrm{b}$. The decoding function of each receiver $\widehat{t}_e$, $e \in E$, is $\widehat{\mathsf{d}}_{\widehat{t}_e}(\widehat{x}_\mathrm{b}, \widehat{x}_{\mathtt{in}(\mathtt{tail}(e)) \cup O'^{-1}(\mathtt{tail}(e))})$, where $\mathtt{in}(\mathtt{tail}(e)) \subsetneq E$ are in the upstream of e, and $O'^{-1}(\mathtt{tail}(e)) \subseteq [k + \ell]$.

Since the decoding for all receivers are correct, $\widehat{\mathsf{d}}_{\widehat{t}_e}(\cdot) = \widehat{x}_e$, for all $e \in E$. As G is acyclic, by considering decoding functions $\widehat{\mathsf{d}}_{\widehat{t}_e}(\cdot)$ starting from *root* nodes, that is, links e where $\mathtt{in}(\mathtt{tail}(e)) = \emptyset$, and traversing the links in the directions of the links, all *link messages* $\widehat{x}_E$ are completely determined by $\widehat{x}_{[k+\ell]}$ and $\widehat{x}_\mathrm{b}$. ∎

VI. Proof of Theorem 1 (From Index Codes to Network Codes)

The code translation from $\mathbb{N}'$ to $\mathbb{N}$ is straightforward. By substituting each $X'_{[k+i]}$, $i \in [\ell]$, with a random key K_i, we conclude that if $\mathbb{N}'$ is $((M_{[k]}, \kappa_{[\ell]}), \epsilon, \eta, n)$-feasible using a deterministic code, then $\mathbb{N}$ is $(M_{[k]}, \epsilon, \eta, n)$-feasible using a randomised code. So, we only need to show the result from deterministic index codes for $\mathbb{I}$ to deterministic network codes for $\mathbb{N}'$.

A. Index codes

Consider an index code that is composed of the following:

- Sender's broadcast (encoding) function, $\widehat{x}_{\mathrm{b}} = \widehat{\mathsf{e}}(\widehat{x}_{[k+\ell]\cup E})$.
- Receiver $\hat{t}$'s decoding function, $\widehat{\mathsf{d}}_{\hat{t}}(\widehat{x}_{\mathrm{b}}, \widehat{x}_{\widehat{H}_{\hat{t}}})$, where
  - $\widehat{H}_{\hat{t}_e} = \mathtt{in}(\mathtt{tail}(e)) \cup O'^{-1}(\mathtt{tail}(e))$, for $e \in E$.
  - $\widehat{H}_{\hat{t}_i} = \mathtt{in}(i) \cup O'^{-1}(i)$, for $i \in U$. Recall that U is the set of destination nodes in $\mathbb{N}$ (and $\mathbb{N}'$).

B. Translated network codes

The translated network code is as follows [13]:

- An encoding function for each link $e \in E$ such that $\mathsf{e}_e(x_{\widehat{H}_{\hat{t}_e}}) = \widehat{\mathsf{d}}_{\hat{t}_e}(\sigma, x_{\widehat{H}_{\hat{t}_e}})$,
- A decoding function for each destination $i \in U$ such that $\mathsf{d}_i(x_{\mathtt{in}(i)\cup O^{-1}(i)}) = \widehat{\mathsf{d}}_{\hat{t}_i}(\sigma, x_{\mathtt{in}(i)\cup O^{-1}(i)})$,

for some $\sigma \in 2^{\widehat{n}}$.

C. The choice of σ for the network code

Suppose in $\mathbb{I}$ that a message realisation $\widehat{x}_{[k+\ell]\cup E}$ results in correct decoding. From the proof of Proposition 1, we know that given $\widehat{x}_{[k+\ell]}$ and $\widehat{\mathsf{e}}(x_{[k+\ell]\cup E})$, a sequence of receiver decoding functions can collectively recover $\widehat{x}_E$. So, using above network-code translation, we have following observation:

Observation 1: Suppose that $\widehat{x}_{[k+\ell]\cup E}$ results in correct decoding in $\mathbb{I}$. We use the translated network code in $\mathbb{N}'$. For the message realisation $x_{[k+\ell]} = \widehat{x}_{[k+\ell]}$ in $\mathbb{N}'$, if $\sigma = \widehat{\mathsf{e}}(\widehat{x}_{[k+\ell]\cup E})$ *had been chosen* for the network code, then each edge $e \in E$ will send $x_e = \widehat{x}_e$ (since its encoding function is derived from the decoding function in $\mathbb{I}$), and the decoding of all receivers in $\mathbb{N}'$ will be correct (since decoding in $\mathbb{I}$ is all correct).

D. Issues in choosing σ

Note that requiring the nodes of the augmented secure network-coding problem $\mathbb{N}'$ to compute an appropriate σ results in a circular argument, since σ is a function of all messages, including the ones the nodes have to decode.

For a non-secure network-index code equivalence, a solution was obtained by showing the existence of a "good" candidate [13] of σ, using which one can bound the decoding probability of error of the network code from above by ϵ.

For a secure equivalence, attempts to pre-select a good candidate of σ were unsuccessful [12]. Although a good candidate that preserves the decoding-error probability exists, finding one that simultaneously preserves both the decoding-error probability and leakage remains elusive.

E. A proposed solution: cycling σ

In this paper, we propose to *cycle* σ over different values. To simplify notation, let

- $m := \prod_1^{k+\ell} M_i$ denote the total number of message realisations of $\widehat{X}_{[k+\ell]}$.
- $d := 2^{\widehat{n}} = 2^{\sum_{e\in E}\lfloor c_e n\rfloor}$ denote the total number of message realisations of $\widehat{X}_E$.

Recall that $\widehat{\mathsf{e}}(x_{[k+\ell]\cup E}) \in 2^{\widehat{n}}$. Figure 1 shows the broadcast message $\widehat{\mathsf{e}}(x_{[k+\ell]\cup E})$ for each message realisation $(\widehat{x}_{[k+\ell]}, \widehat{x}_E)$.

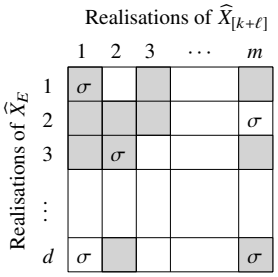

Fig. 1. A table showing the value of $\widehat{\mathsf{e}}(\widehat{x}_{[k+\ell]\cup E}) \in 2^{\widehat{n}}$ for each message realisation $(\widehat{x}_{[k+\ell]}, \widehat{x}_E)$. Shaded cells indicate message realisations that result in correct decoding for all receivers.

Note that, due to Proposition 1, $\widehat{\mathsf{e}}(x_{[k+\ell]\cup E})$ corresponding to all shaded cells in any column must be distinct.

Let T_σ be the number of realisations $\widehat{x}_{[k+\ell]\cup E}$ that result in $\widehat{\mathsf{e}}(\widehat{x}_{[k+\ell]\cup E}) = \sigma$, and N_σ be the number of realisations $\widehat{x}_{[k+\ell]\cup E}$ that results in correct decoding for all receivers and $\widehat{\mathsf{e}}(x_{[k+\ell]\cup E}) = \sigma$. In Figure 1, T_σ is the total number of cells labelled as σ, and N_σ, the total number of shaded cells labelled as σ. Define $\bar{\epsilon}$ as the fraction of unshaded cells. As the messages are uniformly distributed, $\bar{\epsilon}$ is also the probability of decoding error in $\mathbb{I}$. It is easy to see the following:

$$N_\sigma \le T_\sigma \le N_\sigma + \bar{\epsilon} md, \tag{5}$$

$$\sum_\sigma T_\sigma = md, \tag{6}$$

$$\sum_\sigma N_\sigma = (1-\bar{\epsilon})md, \tag{7}$$

$$\sum_\sigma \frac{1}{d}\frac{N_\sigma}{m} = 1-\bar{\epsilon}. \tag{8}$$

Now, for the translated network code, we consider multiple rounds of message transmissions. The choice of σ in each round is pre-determined and announced to all nodes prior to the transmissions. The fraction of rounds for which a particular σ is selected to be proportional to T_σ, which is

$$p_{\widehat{X}_{\mathrm{b}}}(\sigma) = \frac{T_\sigma}{md}. \tag{9}$$

F. Analysis of probability of decoding error for $\mathbb{N}'$

From Observation 1, if σ is chosen for the network code, for any message realisation $x_{[k+\ell]}$ such that the column $\widehat{x}_{[k+\ell]} = x_{[k+\ell]}$ in Figure 1 contains σ in a shaded cell, decoding in $\mathbb{N}'$ is correct. From Proposition 1, each σ can appear at most once over the shaded cells in each column. So, fixing σ, the probability of correct decoding in $\mathbb{N}'$ is at least $\frac{N_\sigma}{m}$.

By cycling through different σ with frequency $p_{\widehat{X}_{\mathrm{b}}}(\sigma)$, the probability of correct decoding in $\mathbb{N}'$ is

$$1 - P'_{\mathrm{e}} \ge \sum_\sigma p_{\widehat{X}_{\mathrm{b}}}(\sigma)\frac{N_\sigma}{m} \tag{10a}$$

$$\overset{(9)}{=} \sum_\sigma \frac{T_\sigma}{md}\frac{N_\sigma}{m} \overset{(5)}{\ge} \sum_\sigma \frac{N_\sigma}{md}\frac{N_\sigma}{m} \tag{10b}$$

$$= (1-\bar{\epsilon})\sum_\sigma \frac{N_\sigma}{(1-\bar{\epsilon})md}\frac{N_\sigma}{m} \tag{10c}$$

$$= (1-\bar{\epsilon}) \sum_{\sigma} p'(\sigma) \frac{N_\sigma}{m} \tag{10d}$$

$$\geq (1-\bar{\epsilon})(1-\bar{\epsilon}) \tag{10e}$$

$$\geq 1 - 2\bar{\epsilon} \tag{10f}$$

$$\geq 1 - 2\epsilon, \tag{10g}$$

where in (10d), we have define $p'(\sigma) := \frac{N_\sigma}{(1-\bar{\epsilon})md}$, which is a valid probability mass function since it follows from (7) that $\sum_\sigma p'(\sigma) = 1$ and $0 \leq p'(\sigma) \leq 1$. (10e) is obtained by comparing (8) and (10d), and noting that in (8), all $\frac{N_\sigma}{m}$ is multiplied by the same $\frac{1}{d}$ (uniform distribution), while in (10d), a larger $\frac{N_\sigma}{m}$ is multiplied by a larger $p'(\sigma)$. (10g) follows from the fact that the probability of decoding error in $\mathbb{I}$ is $\widehat{P}_\mathrm{e} = \bar{\epsilon} \leq \epsilon$. This means the translated network code has a probability of decoding error $P'_\mathrm{e} \leq 2\epsilon$.

G. Analysis of leakage for $\mathbb{N}'$

Define a new random variable $\widehat{C}$, such that $\widehat{C} = 1$ if the decoding is correct in $\mathbb{I}$, and $\widehat{C} = 0$ otherwise. From the identity $I(\mathtt{P};\mathtt{Q}|\mathtt{X}) + I(\mathtt{P};\mathtt{R}|\mathtt{Q},\mathtt{X}) = I(\mathtt{P};\mathtt{R}|\mathtt{X}) + I(\mathtt{P};\mathtt{Q}|\mathtt{R},\mathtt{X})$, we get

$$I(\mathtt{P};\mathtt{Q}|\mathtt{X}) = I(\mathtt{P};\mathtt{Q}|\mathtt{R},\mathtt{X}) + I(\mathtt{P};\mathtt{R}|\mathtt{X}) - I(\mathtt{P};\mathtt{R}|\mathtt{Q},\mathtt{X}) \tag{11}$$

$$\geq I(\mathtt{P};\mathtt{Q}|\mathtt{R},\mathtt{X}) - I(\mathtt{P};\mathtt{R}|\mathtt{Q},\mathtt{X}) \tag{12}$$

$$\geq I(\mathtt{P};\mathtt{Q}|\mathtt{R},\mathtt{X}) - H(\mathtt{R}). \tag{13}$$

Similarly,

$$I(\mathtt{P};\mathtt{Q}|\mathtt{R},\mathtt{X}) \geq I(\mathtt{P};\mathtt{Q}|\mathtt{X}) - H(\mathtt{R}) \tag{14}$$

Consider eavesdropper $z \in Z$. Recall that $\widehat{B}_z = B_z := B$ and $\widehat{A}_z = A_z := A$, where we drop the subscripts for simplicity. Starting with an index code that has at most η leakage,

$$\eta \geq I(\widehat{X}_A; \widehat{X}_\mathrm{b}, \widehat{X}_B) \geq I(\widehat{X}_A; \widehat{X}_B | \widehat{X}_\mathrm{b}) \tag{15a}$$

$$\geq I(\widehat{X}_A; \widehat{X}_B | \widehat{X}_\mathrm{b}, \widehat{C}) - H(\widehat{C}), \tag{15b}$$

where (15b) follows from (13). Rearranging terms, we obtain

$$\eta + H(\widehat{C}) \geq \sum_\sigma p_{\widehat{X}_\mathrm{b},\widehat{C}}(\sigma,1) I(\widehat{X}_A; \widehat{X}_B | \widehat{X}_\mathrm{b} = \sigma, \widehat{C} = 1) + \sum_\sigma p_{\widehat{X}_\mathrm{b},\widehat{C}}(\sigma,0) I(\widehat{X}_A; \widehat{X}_B | \widehat{X}_\mathrm{b} = \sigma, \widehat{C} = 0). \tag{16}$$

From Proposition 1, if a message realisation $(\widehat{x}_{[k+\ell]}, \widehat{x}_E)$ results in correct decoding in $\mathbb{I}$ (that is, it maps to a shaded cell in Figure 1), then $\widehat{x}_E = \phi_\sigma(\widehat{x}_{[k+\ell]})$ for some deterministic function ϕ_σ, where $\sigma = \widehat{\mathsf{e}}(\widehat{x}_{[k+\ell]}, \widehat{x}_E)$. Now, suppose that $\sigma = \widehat{\mathsf{e}}(\widehat{x}_{[k+\ell]}, \widehat{x}_E)$ is chosen for the network code. From Observation 1, if $x_{[k+\ell]} = \widehat{x}_{[k+\ell]}$ is transmitted, then $x_E = \phi_\sigma(x_{[k+\ell]})$. Then, for all a and b,

$$p^\sigma_{X_E|X_{[k+\ell]},C}(b|a,1) = p_{\widehat{X}_E|\widehat{X}_{[k+\ell]},\widehat{X}_\mathrm{b},\widehat{C}}(b|a,\sigma,1), \tag{17}$$

where the superscript σ indicates that σ is used for the network code. Note that $C \in \{0,1\}$ is a random variable for $\mathbb{N}'$, where $C = 1$ indicates that σ appears in a shaded cell in the column $\widehat{x}_{[k+\ell]} = x_{[k+\ell]}$ in Figure 1. So, $C = 1$ implies that the decoding in $\mathbb{N}'$ is correct (but the reverse is not always true). It follows that

$$H_\sigma(C) = H(\widehat{C}|\widehat{X}_\mathrm{b} = \sigma), \tag{18}$$

where $H_\sigma(C)$ is evaluated by fixing σ for the network code.

Considering the fraction of columns that contain σ in a shaded cell in Figure 1, we have that $p^\sigma_{X_{[k+\ell]}|C}(a|1) = p_{\widehat{X}_{[k+\ell]}|\widehat{X}_\mathrm{b},\widehat{C}}(a|\sigma,1)$. Combining this with (17), we get the following for all a and b:

$$p^\sigma_{X_{[k+\ell]},X_E|C}(a,b|1) = p_{\widehat{X}_{[k+\ell]},\widehat{X}_E|\widehat{X}_\mathrm{b},\widehat{C}}(a,b|\sigma,1), \tag{19}$$

which then necessitates that

$$I(\widehat{X}_A; \widehat{X}_B | \widehat{X}_\mathrm{b} = \sigma, \widehat{C} = 1) = I_\sigma(X_A; X_B | C = 1), \tag{20}$$

where $I_\sigma()$ is calculated when σ is used for the network code.

For the second mutual-information term in (16), we note that when there are decoding errors in $\mathbb{I}$ (that is, $\widehat{C} = 0$) and the broadcast message is $\sigma = \widehat{\mathsf{e}}(x_{[k+\ell]\cup E})$, one or more receiver $\widehat{t}_e$ decodes its required message $\widehat{x}_e$ wrongly to some $\widehat{x}_e^{(\widehat{t}_e)} \neq \widehat{x}_e$.

Translating this to the network code, suppose that $C = 0$, and $\sigma = \widehat{\mathsf{e}}(\widehat{x}_{[k+\ell]\cup E})$ for some $\widehat{x}_{[k+\ell]\cup E}$ has been chosen. Now, if the source message realisation is $x_{[k+\ell]} = \widehat{x}_{[k+\ell]}$, then at least one node $\mathtt{tail}(e)$ sends a *wrong* link message $x_e \neq \widehat{x}_e$, which is the *noisy* version of $\widehat{x}_e$. Subsequent downstream nodes that receive *wrong* incoming link messages add further *noise* to their outgoing link messages when compared to $\widehat{x}_E$. Noting that $A \subseteq [k]$, $B \subseteq E$, and $e \in E$, X_A is a subset of source messages, and X_B is a subset of the links messages, we have

$$I(\widehat{X}_A; \widehat{X}_B | \widehat{X}_\mathrm{b} = \sigma, \widehat{C} = 0) \geq I_\sigma(X_A; X_B | C = 0). \tag{21}$$

Indeed, as far as eavesdroppers are concerned, whether or not the destination nodes in $\mathbb{N}'$ decode correctly is irrelevant here.

Substituting (20) and (21) this into (16), we get

$$\eta + H(\widehat{C}) \geq \sum_\sigma p_{\widehat{X}_\mathrm{b},\widehat{C}}(\sigma,1) I_\sigma(X_A; X_B | C = 1) + \sum_\sigma p_{\widehat{X}_\mathrm{b},\widehat{C}}(\sigma,0) I_\sigma(X_A; X_B | C = 0) \tag{22a}$$

$$= \sum_\sigma p_{\widehat{X}_\mathrm{b}}(\sigma) I_\sigma(X_A; X_B | C) \tag{22b}$$

$$= I(X_A; X_B | C) \tag{22c}$$

$$\geq I(X_A; X_B) - H(C), \tag{22d}$$

where

- (22c) follows from cycling the choice of σ for the network code with fractions according to $p_{\widehat{X}_\mathrm{b}}(\sigma)$; and
- (22d) follows from (14).

Now,

$$H(C) = \sum_\sigma p_{\widehat{X}_\mathrm{b}}(\sigma) H_\sigma(C) \overset{(18)}{=} \sum_\sigma p_{\widehat{X}_\mathrm{b}}(\sigma) H(\widehat{C}|\widehat{X}_\mathrm{b} = \sigma) = H(\widehat{C}|\widehat{X}_\mathrm{b}) \leq H(\widehat{C}).$$

So, the leakage in $\mathbb{N}'$ can be upper-bounded as follows:

$$I(X_A; X_B) \leq \eta + H(\widehat{C}) + H(C) \leq \eta + 2H(\widehat{C}) \leq \eta + 2H_\mathrm{b}(\epsilon),$$

where $H_\mathrm{b}(\cdot)$ is the binary entropy function.

Note that for perfect decoding $\epsilon = 0$ and $I(X_A; X_B) \leq \eta$. This then completes the proof of Theorem 1. ∎

REFERENCES

[1] 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.

[2] K. Shanmugam and A. G. Dimakis, "Bounding multiple unicasts through index coding and locally repairable codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, USA, June 29–July 4 2014, pp. 296–300.

[3] F. Arbabjolfaei and Y.-H. Kim, "Three stories on a two-sided coin: Index coding, locally recoverable distributed storage, and guessing games on graphs," in *Proc. 53rd Allerton Conf. Commun. Control Comput. (Allerton Conf.)*, Monticello, USA, Sept. 29–Oct. 2 2015.

[4] R. W. Yeung, *Information Theory and Network Coding*, 1st ed. Springer, 2008.

[5] R. Dougherty and K. Zeger, "Nonreversibility and equivalent constructions of multiple-unicast networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 1982–1986, Nov. 2006.

[6] W. Huang, T. Ho, M. Langberg, and J. Kliewer, "On secure network coding with uniform wiretap sets," in *Proc. IEEE Int. Symp. on Netw. Coding (NetCod)*, Calgary, Canada, June 7–9 2013.

[7] S. El Rouayheb, A. Sprintson, and C. Georghiades, "On the index coding problem and its relation to network coding and matroid theory," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3187–3195, July 2010.

[8] 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.

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

[10] T. Chan and A. Grant, "Capacity bounds for secure network coding," in *Proc. Australian Commun. Theory Workshop (AusCTW)*, Christchurch, New Zealand, Jan. 30–Feb. 1 2008, pp. 95–100.

[11] L. Ong, B. N. Vellambi, J. Kliewer, and P. L. Yeoh, "An equivalence between secure network and index coding," in *Proc. IEEE Glob. Telecomm. Conf. Workshop Workshop Netw. Coding Theory Appl. (Globecom NetCod)*, Washington, USA, Dec. 4 2016.

[12] L. Ong, J. Kliewer, and B. N. Vellambi, "Secure network-index code equivalence: Extension to non-zero error and leakage," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Vail, USA, June 17–22 2018, pp. 841–845.

[13] M. Effros, S. El Rouayheb, and M. Langberg, "An equivalence between network coding and index coding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, July 7–12 2013, pp. 967–971.