This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings
This paper was presented as part of the main Technical Program at IEEE INFOCOM 2010.

Optimal Linear Network Coding Design for Secure Unicast with Multiple Streams

Jin Wang[1,2], Jianping Wang[2], Kejie Lu[3], Bin Xiao[4], Naijie Gu[1]
[1] Department of Computer Science, University of Science and Technology of China
[2] Department of Computer Science, City University of Hong Kong
[3] Department of Electrical and Computer Engineering, University of Puerto Rico at Mayagüez
[4] Department of Computing, Hong Kong Polytechnic University
Email: wangjin3@student.cityu.edu.hk, jianwang@cityu.edu.hk, kejie.lu@upr.edu,
csbxiao@comp.polyu.edu.hk, gunj@ustc.edu.cn

***Abstract*—Linear network coding is a promising technology that can maximize the throughput capacity of communication network. Despite this salient feature, there are still many challenges to be addressed, and security is clearly one of the most important challenges. In this paper, we will address the design of secure linear network coding. Specifically, we will investigate the network coding design that can both satisfy the weakly secure requirements and maximize the transmission data rate of multiple unicast streams between the same source and destination pair, which has not been addressed in the literature. In our study, we first prove that the *secure unicast routing* problem is equivalent to a *constrained link-disjoint path* problem. We then develop efficient algorithm that can find the optimal unicast topology in a polynomial amount of time. Based on the topology, we design *deterministic* linear network code that is weakly secure and can be constructed at the source node. And finally, we investigate the potential of *random* linear code for weakly secure unicast and prove the low bound of the probability that a random linear code is weakly secure.**

Keywords: Network coding, Weakly secure, Throughput Capacity

I. Introduction

Since the pioneer work by Ahlswede *et al.* [1], *network coding* has attracted significant attention because it has the potential to maximize the throughput capacity of a network. Li *et al.* [2] demonstrated that the maximum flow from a source to multiple destinations can be achieved by *linear network coding* with a finite field size. Due to the simplicity of the coding scheme, linear network coding has been widely used in the literature. In this paper, we will address network coding design based on linear network coding.

In addition to maximizing the transmission data rate, *network security* is another key challenge for communication networks. Due to the fundamental importance of network security, many researchers have been working on the security aspects of network coding. In the literature, existing studies on *secure network coding* can be classified according to the attack models.

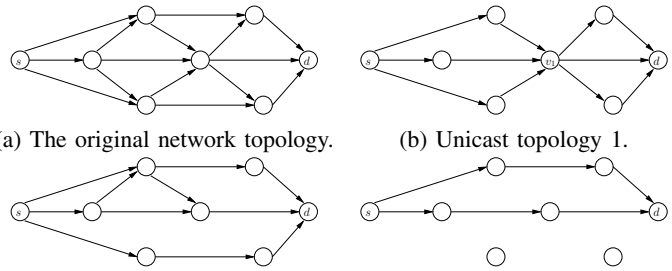
Specifically, *active attacks*, including entropy attack and byzantine modification attack, have been investigated in [3]–[7]. In these attacks, an attacker can alter the messages transmitted in the network, and consequently the receiver cannot recover the original data. Generally, to deal with active attacks, a data verification scheme is necessary to detect and to filter out modified messages in order to provide *integrity* of data transmission.

On the other hand, *passive attacks*, such as wiretapping attack, have been studied in [8]–[11]. In these attacks, an attacker may either wiretap one or more links in the network, or try to acquire as much information as possible from the data that pass through one or more intermediate nodes in the network. For passive attacks, network coding can provide *confidentiality* naturally because network coding tends to utilize more links and nodes to maximize the transmission data rate. For instance, if linear network coding is used, then a message transmitted in the network is a linear combination of several original messages. Therefore, the attacker may not be able to decode the original messages if the number of independent linear combinations acquired by the attacker is smaller than the number of original messages. Clearly, linear network coding can be used as an effective approach to provide secure transmission against passive attacks. In this paper, we will investigate the design of secure linear network coding to deal with passive attacks.

To defend against passive attacks, secure network coding generally can be implemented in two steps: (1) construct a transmission topology; (2) design secure network coding scheme based on the constructed transmission topology. In the literature, most existing work on secure network coding against passive attacks focus on Step 2, i.e., to design secure coding scheme based on a given routing topology [8], or to create a transformation matrix based on a given linear network coding scheme [10].

Although the studies on Step 2 are important, they also have some limitations. Firstly, the transmission topology may determine the maximum rate that can be achieved with certain security requirements. Suppose that a transmission topology has been chosen to support a unicast flow, the messages obtained by every intermediate node should be limited to defend against passive attacks. Consequently, the achievable *secure transmission rate* may be reduced. Secondly, given the attack model and a connection request, the transmission

978-1-4244-5837-0/10/$26.00 ©2010 IEEE



(a) The original network topology. (b) Unicast topology 1.

(c) Unicast topology 2. (d) Unicast topology 3.

Fig. 1. Examples for Unicast Routing Topology.

topology may affect whether a secure transmission is possible in some scenarios.

We now use an example to show the impact of transmission topology on secure transmission rate. We consider the passive attack model used in [10] and [11] in which the network consists entirely of "nice but curious" intermediate nodes. These nodes may try to acquire as much information as possible from the data that pass through them. We also assume that the intermediate nodes do not cooperate with each other. Fig. 1(a) shows a directed acyclic graph, in which each link has a unit capacity. In this example, there are only one source node s and one destination node d in the network, where the maximal flow between the source and destination is 3. Fig. 1(b)-(d) give three different unicast routing topologies, which lead to different secure transmission rates:

- In Fig. 1(b), all packets from s to d pass through the intermediate node v_1, which means that no matter how we design network coding, node v_1 can decode all the messages as the destination node d does.
- In Fig. 1(c), the maximum number of packets passing through each intermediate node simultaneously is at most 2. Given such a routing topology, we can construct linear network coding which will be explained later) with secure transmission rate of 3.
- In Fig. 1(d), the maximum number of packets passing through each intermediate node simultaneously is at most 1. Given such a routing topology, we can construct linear network coding with secure transmission rate of 2.

The example above has well demonstrated that the routing topology has a great impact on the maximum secure transmission rate under certain security requirements. In this paper, we focus on the *secure unicast routing* problem, in which we will integrate the topology formation and secure network coding design together. The objective of the problem is to design optimal linear network coding scheme that can satisfy the *weakly secure*[1] requirements and maximize the transmission rate of multiple streams that share the same source and destination nodes, which is a common scenario in communication networks.

To the best of the authors' knowledge, no previous works have been conducted to address such a secure unicast routing problem. The main contributions of the paper are summarized as follows:

- *Problem modeling*: We prove that the secure unicast routing problem is equivalent to a *constrained link-disjoint path* problem under the requirements of weakly secure.
- *Secure unicast routing*: We develop efficient polynomial algorithm that can find the secure unicast routing topology.
- *Deterministic linear coding*: Given a secure unicast routing topology, we construct a linear network coding scheme that can achieve the maximum secure transmission rate.
- *Random linear coding*. We develop the lower bound of the probability that a random linear coding scheme is weakly secure. The derived lower bound clearly shows the relationship between the size of the finite field and the security requirements.

The rest of the paper is organized as follows. In Section II, we provide the definitions of the network model, the attack model, and the problem to be addressed. We will then model the secure unicast routing problem as a graph problem in Section III. The topology construction method will be presented in Section IV, followed by a deterministic code design in Section V. The behavior of random linear code will be addressed in Section VI. And finally, we discuss related work in Section VII and conclude the paper in Section VIII.

II. A Secure Unicast Routing Problem

In this section, we will define the secure unicast routing problem that will be investigated. Specifically, we first summarize the notations and parameters to be used in the paper. We then introduce the network model and the attack model that will be addressed. Finally, we provide the definition of the problem.

A. Notations

To facilitate the discussion, we define the following parameters and notations.

- Vectors, matrixes and linear spans will be denoted by symbols with **bold** font.
- Symbol $^{\mathrm{T}}$ will be used to indicate transpose.
- Parameters i and j will be used as indexes.
- $\langle\cdot\rangle$ will be used as the linear span of a set of vectors.

B. The Network Model

In our study, we consider a directed acyclic graph $G = (V, E)$, where V is the set of vertices (or nodes) and E is the set of edges (or links). We assume that a *time-division multiplexing* (TDM) scheme will be applied, and the whole time horizon will be partitioned into fixed-size time slots.

We further assume that each link in this graph has the same *unit capacity*, which is 1 data unit per time slot. Note that the capacity of different links can be different in reality. However, we can always convert a link with a certain capacity C (where C is a nature number) data units per time slot to C links, each of which has a unit capacity. Suppose that all data packets have

[1]The formal definition of weakly secure can be found in Section II-D

the same size and will occupy a single time slot, then each link can transmit one packet per time slot.

We now consider a total of L unicast data streams that share the same source and destination nodes, which are denoted by s and d, respectively. We suppose that, for each data stream, packets arrive periodically with rate k packets per time slot.

C. The Linear Network Coding Scheme

Upon arrival, data packets from different streams are buffered at node s, waiting for encoding and transmission towards node d. We let T be the coding interval in units of time slots. Therefore, kTL packets will be encoded together in each T time slots. In this paper, we let k, L and T be integers, and we let $K = kT$.

To encode the packets, we will apply linear network coding, in which a total of $(K \times L)$ packets will be coded and delivered to node d. Let $\mathbf{M} = [m_{1,1}, m_{1,2}, \ldots, m_{1,K}, m_{2,1}, \ldots, m_{L,K}]^{\mathrm{T}}$ represent the data packets to be encoded at node s, and let $\mathbf{M}_i = [m_{i,1}, m_{i,2}, \ldots, m_{i,K}]^{\mathrm{T}}$ represent the set of data packets from stream i. For link e and time slot t, we let vector $\mathbf{f}_e(t)$ with length $(K \times L)$ be the global encoding vector, so that the coded data packet transmitted on e at t is $\mathbf{f}_e(t)\mathbf{M}$.

Let $In(v_i)$ and $Out(v_i)$ be the set of input and output links of a given node v_i, respectively. For the source node s, we assume that there are L virtual links, each of which represents one data stream. We define $|In(v_i)|$ and $|Out(v_i)|$ as the cardinality of the two sets, respectively. We can define the linear network code for unicast with multiple streams as below.

Definition 1: A $(K \times L)$-dimensional linear network code $\phi(G, \mathbb{F}_q, T, K, L)$, where $\mathbb{F}_q$ is a field of size q over which the code is defined, consists of vectors $\mathbf{f}_e(t)$ with length $(K \times L)$, $\forall t \in \{1, 2, \cdots, T\}$ and $\forall e \in E \cup In(s)$, such that:

1) For each virtual link $e \in In(s)$, $\bigcup_{t=1}^{T}\{\mathbf{f}_e(t)\}$ forms a basis of the vector space $\mathbb{F}_q^{K \times L}$.
2) For $e \in Out(v_i)$, $\mathbf{f}_e(t)$ is a linear combination of $\bigcup_{t'=1}^{t}\{\mathbf{f}_{e'}(t'), e' \in In(v_i)\}$.

Let the set of intermediate nodes be $I = V - \{s, d\}$. For each $v_i \in I$, $\mathbf{A}_i = \bigcup_{t=1}^{T}\{\mathbf{f}_e(t), e \in In(v_i)\}$ and $\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_{|I|}\}$. We also let $\mathbf{Y}_{v_i} = \langle \mathbf{A}_i \rangle$, and we note that $dim(\mathbf{Y}_{v_i}) = Rank(\mathbf{A}_i)$. We denote that $R_i = Rank(\mathbf{A}_i)$ and $R = \max_i R_i$.

Definition 2: A $(K \times L)$-dimensional linear network code $\phi(G, \mathbb{F}_q, T, K, L)$ is a *linear unicast code* $\upsilon(G, \mathbb{F}_q, T, K, L)$, if $dim(\mathbf{Y}_d) = dim(\mathbf{Y}_s) = K \times L$.

Fig. 2 shows an example of a 6-dimensional linear unicast code, with $k = 1$, $L = 3$, $T = 2$.

D. The Attack Model and Requirements of Weakly Secure

In this paper, we consider the passive attack model, in which each node $v_i \in I$ may try to acquire data information passing through it. Moreover, intermediate nodes do not cooperate with each other to decode the packets sent from the source node s. The scenarios that two or more nodes can cooperate in the attack will be out of the scope of this paper.

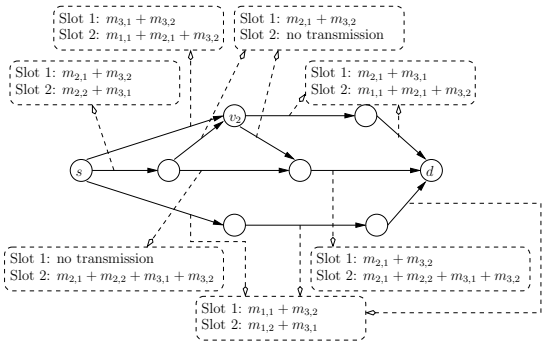


Fig. 2. A 6-Dimensional linear network code with $k = 1$, $L = 3$, $T = 2$.

With the assumptions above, we consider the *weakly secure* conditions, which was proposed in [10]. Specifically, if we let $\mathbf{Z}_i$ be the matrix formed by nonzero vectors in $\mathbf{A}_i$ as its rows, which implies that $Rank(\mathbf{Z}_i) = R_i$, we have

Definition 3: A $(K \times L)$-dimensional linear network code $\phi(G, \mathbb{F}_q, T, K, L)$ is weakly secure, if $\phi(G, \mathbb{F}_q, T, K, L)$ satisfies that for every intermediate node $v_i \in I$

$$H(\mathbf{M}_j | \mathbf{Z}_i \mathbf{M}) = H(\mathbf{M}_j), \forall j \in \{1, 2, \cdots, L\}, \qquad (1)$$

where $H(\cdot)$ and $H(\cdot|\cdot)$ denote entropy and conditional entropy, respectively.

Eq. (1) indicates that $\mathbf{M}_j$ and $\mathbf{Z}_i\mathbf{M}$ are independent for any i and j. It means that every intermediate node cannot obtain any nonzero messages which are linear combination of data packets from the same data stream.

Definition 4: A linear unicast code $\upsilon(G, \mathbb{F}_q, T, K, L)$ is *weakly secure linear unicast code* if it satisfies the requirements of weakly secure.

As an example, the linear unicast code illustrated in Fig. 2 is not weakly secure because node v_2 can decode $m_{1,1}$ by the end of the second time slot.

E. Problem Description

In this paper, we aim to address the following secure unicast routing problem.

Definition 5: Given graph G, number of data streams L and coding interval T, the optimal secure unicast routing problem is to find the maximum value of k^* (e.g. maximum transmission rate $k^* \times L$) and its correspondent secure linear unicast code $\upsilon(G, \mathbb{F}_q, T, K^*, L)$, where $K^* = k^* \times T$.

III. Problem Modeling

In this section, we prove that the secure unicast routing problem explained above can be modeled as a *constrained link-disjoint path* problem.

Lemma 1: If $\upsilon(G, \mathbb{F}_q, T, K, L)$ is a weakly secure linear unicast code, then for every intermediate node $v_i \in I$, $dim(\mathbf{Y}_{v_i}) \leq K \times (L - 1)$.

Proof: We prove the Lemma by contradiction. Because every global encoding vector has length $K \times L$, for each intermediate node, $dim(Y_{v_i}) < K \times L$, otherwise, if there exists an intermediate node v_i with $dim(Y_{v_i}) = K \times L$, it

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings
This paper was presented as part of the main Technical Program at IEEE INFOCOM 2010.

can decode the message as the destination node d does, which contradict with the condition that $\upsilon(G, \mathbb{F}_q, T, K, L)$ is weakly secure.

We suppose that there exists an intermediate node v_i where $K \times L > dim(\mathbf{Y}_{v_i}) \geq K \times (L-1) + 1$. According to the definition, we can find R_i independent vectors in $\mathbf{A}_i$. We then let $\tilde{\mathbf{Z}}_i$ be the matrix formed by these independent vectors and each vector is placed in $\tilde{\mathbf{Z}}_i$ as a row. Now let $\tilde{\mathbf{Z}}_i(j,u)$ be the element at the j-th row and the u-th column. We can define $\boldsymbol{\alpha}_j = \{\tilde{\mathbf{Z}}_i(j,1), \cdots, \tilde{\mathbf{Z}}_i(j,K)\}$ and $\boldsymbol{\beta}_j = \{\tilde{\mathbf{Z}}_i(j,K+1), \cdots, \tilde{\mathbf{Z}}_i(j,K+K), \cdots, \tilde{\mathbf{Z}}_i(j,K\times L)\}$. Therefore, $\tilde{\mathbf{Z}}_i$ can be represented by

$$\tilde{\mathbf{Z}}_i = \begin{bmatrix} \boldsymbol{\alpha}_1 & \boldsymbol{\beta}_1 \\ \boldsymbol{\alpha}_2 & \boldsymbol{\beta}_2 \\ \vdots & \vdots \\ \boldsymbol{\alpha}_{R_i} & \boldsymbol{\beta}_{R_i} \end{bmatrix} \tag{2}$$

Since the length of each vector in set $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{R_i}\}$ is $K \times (L-1)$, and $R_i = dim(\mathbf{Y}_{v_i}) > K \times (L-1)$, we can derive that vectors in $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{R_i}\}$ are linearly dependent. Therefore, there exists a non-zero vector $\boldsymbol{\Lambda}_i = [\lambda_1, \cdots, \lambda_{R_i}]$ such that $\sum_{j=1}^{R_i} \lambda_j \boldsymbol{\beta}_j = [0, \cdots, 0]$.

Consequently, we have $\boldsymbol{\Lambda}_i \tilde{\mathbf{Z}}_i = [\gamma_1, \cdots, \gamma_K, 0, \cdots, 0]$. Since $\tilde{\mathbf{Z}}_i$ consists of independent row vectors, there exists at least one $\gamma_j \neq 0, j \in \{1, \cdots, K\}$. And we can derive

$$\boldsymbol{\Gamma}_i \tilde{\mathbf{Z}}_i \mathbf{M} = [\gamma_1, \cdots, \gamma_K, 0, \cdots, 0]\mathbf{M}, \tag{3}$$

which shows that we can derive a linear combination of packets from stream 1 (i.e. $\mathbf{M}_1$). If this is true, then the linear network code is not weakly secure, which contradicts with the condition that $\upsilon(G, \mathbb{F}_q, T, K, L)$ is weakly secure.

Therefore, we have proved that, for all intermediate nodes v_i, $dim(\mathbf{Y}_{v_i}) \leq K \times (L-1)$. ■

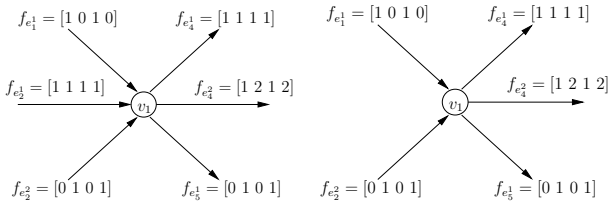
Lemma 1 shows that, if $K \times L$ data packets are coded under a weakly secure linear unicast code $\upsilon(G, \mathbb{F}_q, T, K, L)$, then the maximum number of *independent* vectors received at every intermediate node is no more than $K \times (L-1)$.

In order to find a transmission topology $G' \subset G$, in which a weakly secure linear unicast code exists and the maximum transmission rate from s to d can be achieved, we need to find the necessary and sufficient conditions that a weakly secure linear unicast code exists.

To find these conditions, we first transform graph G into a multigraph $G_m(V_m, E_m)$, where $V_m = V$ and for each link $e_{i,j} \in E$, there are T links in G_m between v_i and v_j, namely, $e^1_{i,j}, e^2_{i_j}, \cdots, e^{\mathrm{T}}_{i,j} \in E_m$.

Lemma 2: If there exists a weakly secure linear unicast code $\upsilon(G, \mathbb{F}_q, T, K, L)$ on G, then there exists a weakly secure linear unicast code $\upsilon_m(G_m, \mathbb{F}_q, 1, K, L)$ on G_m.

Proof: According to the transformation of G_m, data packets transmitted in T time slots on G can be transmitted in one time slot on G_m. Let the global encoding vector of link $e^t_{i,j} \in E_m$ be $\mathbf{f}_{e^t_{i,j}}$, which is same as $\mathbf{f}_{e_{i,j}}(t)$. Obviously, the global encoding vectors for the outgoing links of every intermediate node in G_m are the linear combinations of the incoming global encoding vectors. Therefore, such code assignment is a linear network code. We denote such linear network code as $\upsilon_m(G_m, \mathbb{F}_q, 1, K, L)$.



(a) Before pruning in G_m. (b) After pruning in G'_m.

Fig. 3. An example of link pruning.

Since the data packets obtained at every intermediate node of G_m in one time slot is the same as the data packets that its correspondent node in G receives over T time slots, υ_m is a $(K \times L)$-dimensional linear unicast code satisfying the requirements of weakly secure. ■

According to Lemma 1 and Lemma 2, if there exists a secure linear unicast code $\upsilon_m(G_m, \mathbb{F}_q, 1, K, L)$ on G_m, then $dim(\mathbf{Y}_{v_i}) \leq K \times (L-1), \forall v_i \in I$. This only proves that the maximum number of linear independent vectors received at every intermediate node is no more than $K \times (L-1)$, while the maximum number of vectors received by some intermediate nodes may be more than $K \times (L-1)$.

To deal with this issue, we now show that given a weakly secure linear unicast code $\upsilon_m(G_m, \mathbb{F}_q, 1, K, L)$ on G_m, we can find a weakly secure linear unicast code $\upsilon'_m(G_m, \mathbb{F}_q, 1, K, L)$ so that the maximum number of vectors received by every intermediate node is no more than $K \times (L-1)$.

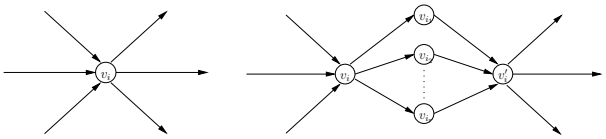
Lemma 3: If there exists a weakly secure linear unicast code $\upsilon_m(G_m, \mathbb{F}_q, 1, K, L)$ on G_m, then there exist $K \times L$ link-disjoint paths in G_m between s and d, and for every intermediate node, the number of different paths passing through it is no more than $K \times (L-1)$.

Proof: For every intermediate node $v_i \in G_m$, we remove its incoming links if the corresponding global encoding vectors do not belong to the maximum independent group of $\mathbf{A}_i$. The link pruning strategy is shown in Fig. 3. After removing such links, we denote the graph as $G'_m = (V_m, E'_m)$ and the correspondent linear network code as $\upsilon'_m(G_m, \mathbb{F}_q, 1, K, L)$.

Because the span space of the maximum independent group of $\mathbf{A}_i$ is the same as the span space of $\mathbf{A}_i$, for each link $e' \in Out(v_i)$ in G'_m, the corresponding global encoding vector $\mathbf{f}_{e'}$ is a linear combination of the global encoding vectors of incoming links and $dim(\mathbf{Y}_d) = K \times L$. Therefore, υ'_m is also a $(K \times L)$-dimensional linear unicast code on G'_m, which can achieve transmission rate $K \times L$ in one time slot.

For every intermediate node $v_i \in G'_m$, the set of global encoding vectors of the incoming links is a subset of $\mathbf{A}_i$ in G_m. Therefore, if υ_m is a $(K \times L)$-dimensional linear unicast code satisfying the requirements of weakly secure on G_m, υ'_m is also a $(K \times L)$-dimensional linear unicast code satisfying the requirements of weakly secure on G'_m.

If the linear network coding scheme υ'_m on G'_m can achieve transmission rate $K \times L$, there must exist a network flow F_m



(a) Before transformation. (b) After transformation.

Fig. 4. An example of link transformation.

from s to d with capacity $K \times L$ in graph G'_m. Because all the links in G'_m have unit capacity, there exist $K \times L$ link-disjoint paths between s and d. According to Lemma 1 and the construction of G'_m, for every intermediate node $v_i \in G'_m$, we have $|In(v_i)| = Rank(\mathbf{A_i}) \leq K \times (L-1)$. Thus, for every intermediate node, the number of different link-disjoint paths passing through it is no more than $K \times (L-1)$. ■

Lemma 4: If there exist $K \times L$ link-disjoint paths in G_m between s and d, and for every intermediate node, the number of different paths passing through it is no more than $K \times (L-1)$, then there exist $k \times L$ link-disjoint paths in G between s and d, and for every intermediate node, the number of different paths passing through it is no more than $k \times (L-1)$.

Proof: If there exist $K \times L$ link-disjoint paths in G_m between s and d, and for every intermediate node, the number of different paths passing through it is no more than $K \times (L-1)$, we set the flow on each link be 1 if and only if the link belongs to one path. Such flow assignment obviously composes a network flow between s and d with capacity $K \times L$ in G_m.

If the number of the links which belong to different paths between v_i and v_j in G_m is f, then we set the flow on link $e_{i,j}$ in G as f/T. Obviously, under such a flow assignment, the flow conservation constraints hold obviously and the flow capacity between s and d in G has capacity $\frac{K \times L}{T} = k \times L$ and the total positive flow entering every intermediate node in G is no more than $r = \frac{K \times (L-1)}{T} = k \times (L-1)$.

After finding such flow in G, we can transform the network $G = (V, E)$ as follows: transform a node v_i into two nodes v_i and v'_i, and add r intermediate nodes $\{v_{i1}, v_{i2}, \cdots, v_{ir}\}$ between them, each of which has two links connecting with v_i and v'_i, respectively. All the new links have unit capacity. An example for the transformation is shown in Fig. 4. The node transformation is done on each node in G except nodes s and d. We denote the graph constructed based on G as $\widetilde{G}$.

Because there exists a network flow between s and d in G with capacity $k \times L$ and the total positive flow entering every intermediate node in G is no more than r, there also exists a network flow with capacity $k \times L$ in $\widetilde{G}$. Due to unit link capacity, there exist $k \times (L-1)$ link-disjoint paths in $\widetilde{G}$. The set of $k \times (L-1)$ link-disjoint paths is denoted as $\widetilde{P}$.

Since a path coming into node v_i must go out from v'_i and there are at most r link-disjoint paths between them, the number of different paths passing through v_i is at most r. Suppose that the path $\widetilde{p}_n \in \widetilde{P}$ in $\widetilde{G}$ is denoted as $\{s, \cdots, v_i, v_{i,l}, v'_i, v_j, v_{j,m}, v'_j, \cdots, d\}$, the corresponding path p_n in G is $\{s, \cdots, v_i, v_j, \cdots, d\}$. Here we note that each three nodes $v_i, v_{i,l}, v'_i$ in $\widetilde{G}$ is replaced by one node v_i in G. If there is a link from v'_i to v_j, there exists a link from v_i to v_j. Therefore, for each path in $\widetilde{G}$, there is a corresponding path from s to d in G. The set of $k \times (L-1)$ link-disjoint paths in G is denoted as P.

We now let the set of links in $\widetilde{p}_n$ be $\widetilde{E}_n$ and the set of links in p_n be E_n. For any $n_1 \neq n_2$, $\widetilde{E}_{n_1} \cap \widetilde{E}_{n_2} = \emptyset$. Because $E_n \subset \widetilde{E}_n, \forall n$, then for any $n_1 \neq n_2$, $E_{n_1} \cap E_{n_2} = \emptyset$. Therefore, the set of paths p_n, $n \in \{1, \cdots, k \times L\}$ are link-disjoint.

We then let the set of nodes in $\widetilde{p}_n$ be $\widetilde{N}_n$ and the set of nodes in p_n be N_n. For every intermediate node v_i, there are at most r paths in $\widetilde{P}$ passing through it in $\widetilde{G}$. Because $N_n \subset \widetilde{N}_n, \forall n$, for every node v_i, there are at most r paths in P passing through it in G.

Therefore, there exist $k \times L$ link-disjoint paths from s to d in G, and for every intermediate node, the number of different paths passing through it is no more than $k \times (L-1)$. ■

Theorem 1: If there exists a $(K \times L)$-dimensional linear unicast code $\upsilon(G, \mathbb{F}_q, T, K, L)$ on G which is weakly secure, then there exist $k \times L$ link-disjoint paths from s to d in G, and for every intermediate node, the number of different paths passing through it is no more than $k \times (L-1)$.

Proof: According to Lemma 2, 3 and 4, this theorem holds obviously. ■

Theorem 2: If there exist $k \times L$ link-disjoint paths from s to d in G, and for every intermediate node, the number of different paths passing through it is no more than $k \times (L-1)$, then there exists a $(K \times L)$-dimensional linear unicast code $\upsilon(G, \mathbb{F}_q, T, K, L)$ on G which is weakly secure.

Proof: If there exist $k \times L$ link-disjoint paths from s to d in G, and for every intermediate node, the number of different paths passing through it is no more than $k \times (L-1)$, then, by routing packets through these link-disjoint paths, the unicast session can transmit $k \times L$ messages in each time slot and can transmit a total of $K \times L$ messages over T time slots.

We denote the set of $k \times L$ disjoint paths in G as $P = \{p_1, \cdots, p_{k \times L}\}$. For each intermediate node v_i, we denote the set of link-disjoint paths passing through it in P as P_i.

When $\mathbf{M} = [m_{1,1}, m_{1,2}, \ldots, m_{1,K}, m_{2,1}, \ldots, m_{L,K}]^{\mathrm{T}}$ and the size of finite field satisfying $q^{K \times L} > |\mathbf{A}| q^{R+K-1} + q^{K \times (L-1)}$, where $|\mathbf{A}| = |\{P_1, \cdots, P_{|I|}\}|$, there exists a transformation matrix $\mathbf{B}$. When we transmit messages $\mathbf{BM}$ instead of $\mathbf{M}$ through $k \times L$ link-disjoint paths and the intermediate nodes only perform the store and forward role, the transmission is weakly secure. This means that the global encoding vectors corresponding to the links in the same path are same, and, for each path, the corresponding global vector is one of the rows in $\mathbf{B}$. The construction of matrix of $\mathbf{B}$ is shown in section V.

Therefore, there exists a $(K \times L)$-dimensional linear unicast code $\upsilon(G, \mathbb{F}_q, T, K, L)$ that is weakly secure. ■

Theorem 3: There exists $(K \times L)$-dimensional linear unicast code $\upsilon(G, \mathbb{F}_q, T, K, L)$ on G which is weakly secure, if and only if there exist $k \times L$ link-disjoint paths from s to d in G, and for every intermediate node, the number of different paths passing through it is no more than $k \times (L-1)$.

Proof: According to Theorem 1 and 2, this theorem holds obviously. ■

For every $k \times L$ link-disjoint paths from s to d in G, and for every intermediate node, the number of different paths passing through it is no more than $k \times (L-1)$, we refer it as *weakly secure $k \times L$ link-disjoint paths*. From Theorem 3, we have proved that the optimal secure unicast routing problem defined in Section II-E is equivalent to finding the weakly secure $k \times L$ link-disjoint paths with maximum value of k.

IV. A Secure Unicast Routing Algorithm

In the previous section, we have proved that the optimal secure unicast routing problem is equivalent to a constrained link-disjoint path problem. In this section, we develop an efficient algorithm to find a secure unicast routing topology in a polynomial amount of time.

In the proof of Lemma 4, we have seen that, if there exists a flow between s and d with capacity $k \times L$ in G, and if the total positive flow entering every intermediate node is no more than $k \times (L-1)$, there exist $k \times L$ link-disjoint paths from s to d in G, and for every intermediate node, the number of different paths passing through it is no more than $k \times (L-1)$. On the opposite direction, the conclusion holds obviously.

Therefore, we can design the algorithm as the following. First, we address the constraint that the maximum positive flow entering each intermediate node is $r = k \times (L-1)$. This node-capacity constraint can be easily converted to a link-capacity constraint by simply "splitting" a node into two and introducing an extra link with the link capacity constraint r between them [12].

After transforming the node capacity constraints to link capacity constraints, the FORD-FULKERSON method can be used to verify whether there exists a network flow with capacity $k \times L$. If such a flow is found, then we know there exist $k \times L$ link-disjoint paths from s to d in G, and for every intermediate node, the number of different paths passing through it is no more than $k \times (L-1)$. Therefore, a $kT \times L$ weakly secure linear unicast code can be constructed on it.

Theorem 4: If there exists a network flow between s and d with capacity $k \times L$ in G and the total positive flow entering every intermediate node is no more than $k \times (L-1)$, there exists a network flow between s and d with capacity $(k-1) \times L$ in G and the total positive flow entering every intermediate node is no more than $(k-1) \times (L-1)$.

Proof: Suppose that there is a network flow F_k between s and d with capacity $k \times L$ in G where the total positive flow entering every intermediate node is no more than $k \times (L-1)$ and the flow on each link $e_{i,j}$ is $f^k_{i,j}$, we can set the flow $f^{k-1}_{i,j}$ on every link $e_{i,j}$ be $\frac{k-1}{k} f^k_{i,j}$. Because $f^k_{i,j} \leq 1$, we have $f^{k-1}_{i,j} \leq 1$, which means such flow assignment does not exceed the link capacity. Such a flow assignment composes a network flow F_{k-1} in G. If the network flow F_k satisfies the flow conservation constraints, F_{k-1} also satisfies the flow conservation constraint. Moreover, the flow capacity between s and d is $\frac{k-1}{k} \times k \times L = (k-1) \times L$; and the total positive flow entering every intermediate node is no more than $\frac{k-1}{k} \times k \times (L-1) = (k-1) \times (L-1)$. ■

Algorithm 1 A secure unicast routing algorithm (SURA)

```
Run FORD-FULKERSON method on graph G and obtain the maximum capacity W;
if W < L then
    return (false, Null);
end if
l_k = 0, h_k = ⌊W/L⌋ − 1, F(k) = Null, ∀1 ≤ k ≤ ⌊W/L⌋;
while l_k < h_r do
    m_k = l_k + ⌊(h_k − l_k)/2⌋
    if P(m_k) = true then
        h_k = m_k;
    else
        l_k = m_k + 1;
    end if
end while
if P(l_k) = false then
    return (false, Null);
end if
return (⌊W/L⌋ − l_k, F(⌊W/L⌋ − l_k));
```

Algorithm 2 Calculating $P(k)$

```
for each node in set I do
    Set node capacity constraint (⌊W/L⌋ − k) × (L − 1);
end for
Transform the graph G with node capacity constraints to graph G' with only link capacity constraints;
Run FORD-FULKERSON method on graph G' to find whether network flow F' with capacity (⌊W/L⌋ − k) × L exists;
if there exists such flow F' then
    Convert flow F' in G' to corresponding flow F in G;
    F(⌊W/L⌋ − k) = F;
    return true
else
    return false
end if
```

Theorem 4 shows that, if a network flow with capacity $k \times L$ can be found under node capacity constraint $k \times (L-1)$, then, for any $k'(1 \leq k' \leq k)$, network flow with capacity $k' \times L$ can be also found under node capacity $k' \times (L-1)$.

We can then utilize this result and find the maximum k, in which we exploit the binary search method. The algorithm shown in Algorithm 1-2 for finding the secure unicast routing is referred to as *secure unicast routing algorithm* (SURA).

If the algorithm returns $(false, Null)$, then we cannot construct weakly secure linear unicast code for any $k \geq 1$. Otherwise, if the algorithm returns $(k, F(k))$, then a network flow $F(k)$ with capacity $k \times L$ can be found. In this case, we can find weakly secure $k \times L$ link-disjoint paths which compose a unicast routing topology on which a weakly secure linear unicast code can be constructed.

Note that in Algorithm 1, the FORD-FULKERSON method is at most run $\log(\lfloor \frac{W}{L} \rfloor) + 1$ times. So, The time complexity of it is $\log(\lfloor \frac{W}{L} \rfloor) + 1) O(|E|W) = O(|E|W \log \lfloor \frac{W}{L} \rfloor)$.

V. Deterministic Network Coding Scheme

In this section, we will construct a weakly secure linear network code in a deterministic manner. Here we first note

that, given a secure unicast routing topology, the coding operations only need to be done at the source node and destination node to achieve the maximum secure transmission rate for weakly secure.

To construct the code, we introduce a matrix $\mathbf{B}$. We use $\{\mathbf{B}\}_l$ to denote the l-th row of $\mathbf{B}$; use $\{\mathbf{B}\}_l^{l+i}$ to denote the set of rows with indexes from l to $l+i$ in $\mathbf{B}$; and use $\{\mathbf{B}\}$ to represent the set of rows of $\mathbf{B}$.

Theorem 5: If there exists a $(K \times L)$-dimensional linear unicast code $\upsilon(G, \mathbb{F}_q, T, K, L)$ on G, and for each intermediate node v_i, $R_i \leq K \times (L-1)$, then there exists a transformation matrix $\mathbf{B}$, with dimension $(K \times L) \times (K \times L)$ over $\mathbb{F}_q$ ($q^{K\times L} > |\mathbf{A}|q^{R+K-1} + q^{K\times(L-1)}$), which can make υ weakly secure.

Proof: First, given a transformation matrix $\mathbf{B}$, instead of $\mathbf{M}$, we can transmit $\mathbf{BM}$ at source s . In this manner, the messages acquired at intermediate node v_i are $\mathbf{Z}_i\mathbf{BM}$. Since the destination node d shall still be able to decode $K \times L$ data after the transformation, the matrix $\mathbf{B}$ is full rank.

Next, we give a constructive proof. Assuming that there exists such transformation, then, intermediate node v_i cannot obtain any message which forms a linear combination of messages only in one data stream $\mathbf{M}_j, j \in \{1, 2, \cdots, L\}$, by taking linear combinations of the acquired messages $\mathbf{Z}_i\mathbf{BM}$, $\forall v_i \in I$; otherwise, $H(\mathbf{M}_j|\mathbf{Z}_i\mathbf{BM}) \neq H(\mathbf{M}_j)$. Therefore, the unicast session will be secure if a full rank matrix $\mathbf{B}$ is found satisfying the condition that:

$$\mathbf{b}\mathbf{Z}_i\mathbf{BM} \neq \mathbf{\Gamma}_j\mathbf{M}_j. \tag{4}$$

for $\forall \mathbf{\Gamma}_j = [\gamma_{(j-1)K+1}, \gamma_{(j-1)K+2}, \cdots, \gamma_{jK}] \neq \mathbf{0}$, $\forall 1 \leq j \leq L, \forall v_i \in I, \forall \mathbf{b}$, where $\mathbf{b}$ is a non-zero vector with length $|In(v_i)|$ on $\mathbb{F}_q$.

Ineq. (4) means that the row space of the K rows from $(j-1)K+1$ to jK, $\forall 1 \leq j \leq L$ of $\mathbf{B}^{-1}$ has no nonzero vectors the same with the row space of the rows in $\mathbf{Z}_i$, $\forall v_i \in I$.

For the $(j-1)K+1$ to jK row of $\mathbf{B}^{-1}$, we select the $((j-1)K+l+1)$-th row of $\mathbf{B}^{-1}$ as a vector not in span $\{(\bigcup_{v_i\in I}\{\mathbf{Z}_i\}) \cup \{\mathbf{B}^{-1}\}_{(j-1)K+1}^{(j-1)K+l}\}$ and span $\{\mathbf{B}^{-1}\}_1^{(j-1)K}$.

Next, we shows that such construction of matrix $\mathbf{B}^{-1}$ satisfies Ineq. (4). We suppose that there exists a nonzero vector $\mathbf{\Delta} = [\delta_1, \cdots, \delta_{|\{\mathbf{Z}_i\}|}]$ and nonzero vector $\mathbf{\Lambda} = [\lambda_{(j-1)K+1}, \cdots, \lambda_{jK}]$, where $\sum_{k=1}^{|\{\mathbf{Z}_i\}|} \delta_k\{\mathbf{Z}_i\}_k = \sum_{l=(j-1)K+1}^{jK} \lambda_l\{\mathbf{B}^{-1}\}_l$. We assume that l' ($(j-1)K+1 \leq l' \leq jK$) is the largest l satisfying $\lambda_{l'} \neq 0$ and

$$\{\mathbf{B}^{-1}\}_{l'} = \lambda_{l'}^{-1}\Big(\sum_{u=1}^{|\{\mathbf{Z}_i\}|} \delta_u\{\mathbf{Z}_i\}_u - \sum_{l=(j-1)K+1}^{l'-1} \lambda_l\{\mathbf{B}^{-1}\}_l\Big). \tag{5}$$

In other words, $\{\mathbf{B}^{-1}\}_{l'}$ is in span $\{(\bigcup_{v_i\in I}\{\mathbf{Z}_i\}) \cup \{\mathbf{B}^{-1}\}_{(j-1)K+1}^{l'-1}\}$ which is a contradiction. Therefore, our construction can find the matrix $\mathbf{B}$ satisfying Ineq. (4).

Suppose that the first $(j-1)K+l, (1 \leq j \leq L)$ row of $\mathbf{B}^{-1}$ can be found, then it is possible to find $\{\mathbf{B}^{-1}\}_{(j-1)K+l+1}$ if

$$q^{K\times L} > \sum_{\mathbf{A}_i\in\mathbf{A}} q^{R_i+l} + q^{(j-1)K}, \forall l = 0, 1, \cdots, K-1.$$

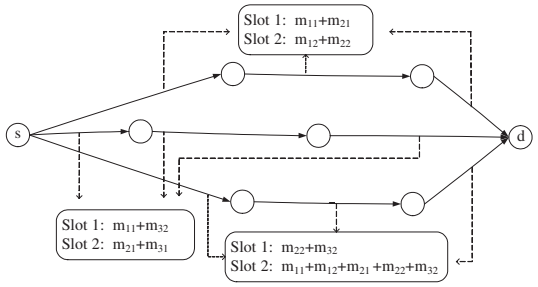


Fig. 5. A 6-dimensional weakly secure linear unicast code on $\mathbb{F}_2$ with $k = 1$, $L = 3$, and $T = 2$.

When $R = \max_i R_i$, the above condition is satisfied if:

$$q^{K\times L} > |\mathbf{A}|q^{R+l} + q^{(j-1)K}, \forall l = 0, 1, \cdots, K-1$$

If $R \leq K \times (L-1)$, then the matrix $\mathbf{B}^{-1}$ can be found when the condition is satisfied:

$$q^{K\times L} > |\mathbf{A}|q^{R+K-1} + q^{(L-1)K} \tag{6}$$

Note that, if $k = T = 1$, the Ineq. (6) becomes $q^L > |\mathbf{A}|q^R + q^{L-1}$, which is the conclusion of Theorem 1 [10]. ■

When given a unicast routing topology $G_u = (V_u, E_u)$ induced by $k \times L$ link-disjoint paths from s to d in G, and for every intermediate node, the number of different paths passing through it is no more than $k\times(L-1)$, we will show that there exists a linear network code which can transmit $K \times L$ data within T time slots. $I_u = V_u - \{s, d\}$.

In G_u, routing by these link-disjoint paths, the unicast session can transmit $k \times L$ messages in each time slot and totally transmit $K \times L$ messages over T time slots.

We first assume that the unicast session just transmits $k \times L$ messages without mixing them together in each time slot and the intermediate nodes only perform the store and forward role. Because the intermediate nodes only perform the store and forward role and $\bigcup_{t_l=1}^{T} \mathbf{f}_e(t_l), e \in In(s)$, form a basis of the vector space $\mathbb{F}_q^{K\times L}$, Such transmission also can be seen as a linear network code with time period T. Under such linear network code, $\mathbf{Z}_i$ is composed by the global encoding vectors received by v_i in the time period as its rows. For every intermediate node, the maximum number of messages it can receive is $K \times (L-1)$ within time period T. In the whole transmission precess, $dim(\mathbf{Y}_{v_i}) \leq K \times (L-1), \forall v_i \in I$.

Apparently, such a simple network coding scheme may be not secure. Nevertheless, we can construct a weakly secure code as follows. Suppose that the number of different paths passing through v_i is r_i, then set $r = \max_{v_i\in I} r_i$. According Theorem 5, when the size of finite field satisfies $q^{K\times L} > |\mathbf{A}|q^{Tr+K-1} + q^{(L-1)K}$ and $r \leq k \times (L-1)$, there exists a transformation matrix $\mathbf{B}$, which has $H(\mathbf{M}_j|\mathbf{Z}_i\mathbf{BM}) = H(\mathbf{M}_j), \forall j = \{1, 2, \cdots, L\}$. Therefore, there exists a $(K \times L)$-dimensional weakly secure linear unicast code $\upsilon(G, \mathbb{F}_q, T, K, L)$.

We denote the set of $k \times L$ disjoint paths in G as $P = \{p_1, \cdots, p_{k\times L}\}$. For each intermediate node v_i, we denote

the set of paths which pass through it in P as P_i. For the construction of weakly secure linear unicast code, the vector transmitted on the same path are the same. We can see that $|\mathbf{A}| = |\{P_1, \cdots, P_{|I|}\}| \leq |I|$.

The unicast topology found in section IV is only dependent on L and k. Therefore, when we find a unicast topology for given L and k, for any $T \geq 1$, the weakly secure linear unicast code can be designed on it. Next we will show that when given L and k the size of field sufficient to construct secure network coding scheme decreases with the increase of T.

Theorem 6: The size of a field which is sufficient to construct secure network code, decreases with the increase of T.

Proof: Suppose $1 \leq T_1 < T_2$ and q satisfies that : $q^{T_1 Lk} > |\mathbf{A}|q^{T_1 r + T_1 k - 1} + q^{T_1(L-1)k}$, then

$$q^{T_1(L-1)k+1} > |\mathbf{A}|q^{T_1 r} + q^{T_1(L-2)k+1} \tag{7}$$

Set function $g(T) = q^{T(L-1)k+1} - |\mathbf{A}|q^{Tr} - q^{T(L-2)k+1}$ and $\Delta = q^{T_1 r} g(T_2) - q^{T_2 r} g(T_1)$. Then:

$$\begin{aligned}\frac{\Delta}{q^{T_1 r + T_2 r + 1}} = \quad & q^{T_2(L-1)k - T_2 r} - q^{T_2(L-2)k - T_2 r} \\ & -(q^{T_1(L-1)k - T_1 r} - q^{T_1(L-2)k - T_1 r})\end{aligned}$$

Function $P(T) = q^{T(L-1)k - Tr} - q^{T(L-2)k - Tr}$

$$\begin{aligned}\frac{d(P(T))}{dT} = \quad & (((L-1)k - r)q^{T(L-1)k - Tr} \\ & -((L-2)k - r)q^{T(L-2)k - Tr})\ln q \\ = \quad & (((L-1)k - r)q^{T(L-2)k - Tr}(q^{Tk} - 1) \\ & + kq^{T(L-2)k - Tr})\ln q \\ > \quad & 0 \quad (\text{where } r \leq k \times (L-1), q > 1, k \geq 1)\end{aligned}$$

It means that $P(T)$ is a monotonic increasing function when $T \geq 1$. Then:

$$\begin{aligned}& \frac{\Delta}{q^{T_1 r + T_2 r + 1}} = P(T_2) - P(T_1) > 0, \quad (\text{when } T_2 > T_1) \\ & \Rightarrow \Delta = q^{T_1 r} g(T_2) - q^{T_2 r} g(T_1) > 0 \\ & \Rightarrow q^{T_1 r} g(T_2) > q^{T_2 r} g(T_1) > 0 \\ & \Rightarrow g(T_2) > 0 \\ & \Rightarrow q^{T_2 Lk} > |\mathbf{A}|q^{T_2 r + T_2 k - 1} + q^{T_2(L-1)k}\end{aligned}$$

We have conclusion that given L and k, if q satisfies the sufficient conditions for T_1 then for any $T_2 > T_1$, q also satisfies the sufficient conditions. It means that the size of field sufficient to construct secure network code decreases with the increase of T. ■

In summary, Theorem 5 shows a constructive upper bound of q on which secure network code can be constructed; and Theorem 6 indicates that the value of q can be reduced if the unicast streams can tolerate more coding delay.

VI. Random Linear Coding Scheme

In practice, *random linear coding* can also be utilized to realize linear network coding [13], [14]. With random linear coding, a node can forward random linear combinations of the messages, which it received previously, to outgoing links. In [13], [14], the authors have proved that such a simple approach can obtain valid linear codes for multicast with probability $(1 - d/q)^\eta$, where η is the number of links with associated randomized coefficients, q is the size of a finite field, and d is the number of destination nodes.

In this section, we investigate the behavior of the random linear coding, when it is applied to the weakly secure unicast routing problem we discuss in this paper. The usage of such linear code is similar to the one we discussed in Section V. The coding operations are only done at the source node and destination node. The major difference is that, instead of computing the transformation matrix $\mathbf{B}$ at the source node, the elements in $\mathbf{B}$ are randomly chosen from a finite field $\mathbb{F}_q$.

Lemma 5: Given a $(K \times L) \times (K \times L)$ matrix $\mathbf{B}$ whose elements are randomly selected in the finite field $\mathbb{F}_q$ and $\tilde{\mathbf{A}}_i$ represents the set of linearly independent vectors obtained by an intermediate node v_i, the probability that the node can not obtain the linear combination of data packets from the same data stream is no less than $\prod_{j=1}^{|\tilde{\mathbf{A}}_i|}(1 - \frac{L}{q^{K(L-1)+1-j}})$.

Proof: An intermediate node can obtain the linear combination of data packets from the same data stream if and only if there exists a non-zero vector $\mathbf{\Gamma}_j$ $(1 \leq j \leq L)$ in the row space of $\tilde{\mathbf{Z}}_i$, such that $\mathbf{\Gamma}_j = [0, \cdots, 0, \gamma_{(j-1)K+1}, \gamma_{(j-1)K+2}, \cdots, \gamma_{jK}, 0, \cdots, 0]$.

Let σ be the number of matrix $\mathbf{Z}$ (with dimension $(|\tilde{\mathbf{A}}_i)| \times (K \times L)$) whose row space does not include $\mathbf{\Gamma}_j$, $\forall 1 \leq j \leq L$.

$$\sigma \geq (q^{KL} - q^K L)(q^{KL} - q^{K+1} L) \cdots (q^{KL} - q^{K+|\tilde{\mathbf{A}}_i|} L). \tag{8}$$

In Ineq. (8), each term is a lower bound for the number of values of $\{\mathbf{Z}\}_l (1 \leq l \leq |\tilde{\mathbf{A}}_i|)$ given $\{\mathbf{Z}\}_1, \cdots, \{\mathbf{Z}\}_{l-1}$ such that the span of $\bigcup_{u=1}^{l} \{\mathbf{Z}\}_u$ does not include $\mathbf{\Gamma}_j, \forall 1 \leq j \leq L$. The number of different matrix with dimension $|\tilde{\mathbf{A}}_i| \times (K \times L)$ in $\mathbb{F}_q$ is $q^{|\tilde{\mathbf{A}}_i| KL}$.

Let p_i denote the probability that the row space of Z_i does not include $\mathbf{\Gamma}_j$, $\forall 1 \leq j \leq L$, then

$$p_i \geq \frac{\sigma}{q^{|\tilde{\mathbf{A}}_i| KL}} = \prod_{j=1}^{|\tilde{\mathbf{A}}_i|}(1 - \frac{L}{q^{K(L-1)+1-j}}) \tag{9}$$

■

Now we can get the lower bound of the probability that a random linear coding is weakly secure, denoted as p_s. Then

$$\begin{aligned}1 - p_s \quad & \leq \quad \sum_i (1 - p_i) \\ & \leq \quad \sum_i \left(1 - \prod_{j=1}^{|\tilde{\mathbf{A}}_i|} \left(1 - \frac{L}{q^{K(L-1)+1-j}}\right)\right) \\ & \leq \quad \sum_i \left(1 - \left(1 - \frac{L}{q^{Tk(L-1)+1-|\tilde{\mathbf{A}}_i|}}\right)^{|\tilde{\mathbf{A}}_i|}\right) \\ & \leq \quad \sum_i \frac{|\tilde{\mathbf{A}}_i| L}{q^{K(L-1)+1-|\tilde{\mathbf{A}}_i|}} \leq \frac{|\mathbf{A}| LR}{q^{K(L-1)+1-R}}\end{aligned} \tag{10}$$

Therefore, given a unicast routing topology, the probability that a random linear coding scheme is weakly secure is greater than $1 - \frac{|\mathbf{A}| LR}{q^{K(L-1)+1-R}}$. And the lower bound of the probability

that a random linear coding scheme is weakly secure increases with the increase of T.

VII. Related work

For secure linear network coding, there are mainly two secure models in previous work, namely, *information theoretical secure* and *weakly secure*. Compared to the weakly secure model, the main feature of information theoretical secure is that the attacker cannot obtain any linear combination of the original messages. To fulfill such requirements, random numbers must be included in the coding process. To achieve information theoretical security, in [8], Cai and Yeung gave a sufficient condition for finding an admissible code to protect the data from being decoded if a set of channels can be accessed by wiretapper. The same secure requirement is considered by Fedman *et al.* in [9], in which they showed that the problem of finding secure network code is the same as finding a block code with certain distance properties. Although we have not investigated information theoretical secure in this work, we believe that our approach can be extended to address issues in this category.

From the perspective of traffic pattern, most existing studies addressed multicast [8]–[11]. As a special case of the secure multicast, the secure unicast routing is studied in [15], in which the authors considered a single unicast flow over cyclic network with the information theoretical secure model. To the best of the authors' knowledge, there are no previous work on secure linear network coding for unicast with multiple streams, which has been addressed in this work.

The size of finite field is an important parameter for linear network coding because a smaller size of finite filed can lead to lower implementation complexity and a higher network link usage. [16] gives a multicast code construction algorithm and the lower bound on field size. For the information theoretical secure model, the authors in [9] studied the tradeoff between the required size of finite field and the multicast rate. For the weak secure model, the authors in [10] presented the sufficient conditions on size of finite filed to transform a linear network code to make the system weakly secure. [11] shows the relationship between the size of finite field and the probability that a random linear code is weakly secure. In our paper, we proved a constructive sufficient condition on the size of finite field to construct a deterministic weakly secure unicast code, and we also developed the lower bound of the probability that a random linear coding scheme is weakly secure.

Finally, the coding schemes designed in [8]–[10] are deterministic. Random linear network coding were discussed in [11], [13], [14]. For instance, [11] analyzed the probability that an intermediate node can decode the message under a random coding scheme. In our work, we studied both deterministic and random coding schemes.

VIII. Conclusions

In this paper, we have investigated the linear network coding design issue for weakly secure unicast with multiple streams between the same source and destination nodes. We firstly proved that the secure unicast routing problem is equivalent to a constrained link-disjoint path problem. Based on such understanding, we developed efficient algorithm to find the optimal unicast topology, whose time complexity is polynomial. A deterministic linear network code is then designed that can be created at the source node. We also discussed the usage of random linear code for weakly secure unicast. We proved a lower bound for the probability that the random linear code is weakly secure.

Acknowledgment

The work is supported in part by the grant from City University of Hong Kong under grant 7002468 and NSF under grant CNS-0424546 and CNS-0922996.

References

[1] R. Ahlswede, N. Cai, S. Li, and R. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[2] S. Li, R. Yeung, and N. Cai, "Linear network coding," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 371–381, 2003.

[3] T. Ho, B. Leong, R. Koetter, M. Medard, M. Effros, and D. Karger, "Byzantine modification detection in multicast networks using randomized network coding," in *Proceedings of IEEE International Symposium on Information Theory (ISIT) 2004*, 2004, p. 144.

[4] M. Krohn, M. Freedman, and D. Mazieres, "On-the-fly verification of rateless erasure codes for efficient content distribution," in *Proceedings of IEEE Symposium on Security and Privacy 2004*, 2004, pp. 226–240.

[5] C. Gkantsidis and P. R. Rodriguez, "Cooperative security for network coding file distribution," in *Proceedings of INFOCOM 2006, the 25th IEEE International Conference on Computer Communications*, 2006, pp. 1–13.

[6] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, and M. Medard, "Resilient network coding in the presence of byzantine adversaries," in *Proceedings of IEEE INFOCOM 2007, the 26th Conference on Computer Communications*, 2007, pp. 616–624.

[7] Z. Yu, Y. Wei, B. Ramkumar, and Y. Guan, "An efficient Signature-Based scheme for securing network coding against pollution attacks," in *Proceedings of IEEE INFOCOM 2008, the 27th Conference on Computer Communications*, 2008, pp. 1409–1417.

[8] N. Cai and R. Yeung, "Secure network coding," in *Proceedings of IEEE International Symposium on Information Theory (ISIT) 2002*, 2002, p. 323.

[9] J. Feldman, T. Malkin, C. Stein, and R. A. Servedio, "On the capacity of secure network coding," in *Proceedings of the 42nd Annual Allerton Conference on Communication, Control, and Computing*, 2004.

[10] K. Bhattad and K. R. Narayanan, "Weakly secure network coding," in *Proceedings of the First Workshop on Network Coding, Theory, and Applications(NetCod)*, Riva del Garda, Italy, 2005.

[11] L. Lima, M. Medard, and J. ao Barros, "Random linear network coding: A free cipher?" in *Proceedings of IEEE International Symposium on Information Theory (ISIT) 2007*, 2007, pp. 546–550.

[12] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Feb. 1993.

[13] T. Ho, R. Koetter, M. Medard, D. Karger, and M. Effros, "The benefits of coding over routing in a randomized setting," in *Proceedings of IEEE International Symposium on Information Theory (ISIT) 2003*, 2003, p. 442.

[14] T. Ho, M. Medard, J. Shi, M. Effros, and D. R. Karger, "On randomized network coding," in *Proceedings of the 41st Annual Allerton Conference on Communication Control and Computing*, vol. 41. The University; 1998, 2003, pp. 11–20.

[15] K. Jain, "Security based on network topology against the wiretapping attack," *IEEE Wireless Communications*, vol. 11, no. 1, pp. 68–71, 2004.

[16] S. Jaggi, P. Chou, and K. Jain, "Low complexity algebraic multicast network codes," in *Proceedings of IEEE International Symposium on Information Theory (ISIT) 2003*, 2003, p. 368.