Effectiveness Analysis of ALNCode Scheme for Anonymous Communication with Network Coding*

YIHUI ZHOU, XUENING LIAO AND ZHENQIANG WU
School of Computer Science
Shaanxi Normal University
Xi’an 710062, P.R. China
E-mail: zhouyihui@snnu.edu.cn

Network coding provides a new perspective for anonymous communication. Wang et al. propose an anonymous communication scheme named ALNCode, in which the linear relation of the global encoding vectors can be hidden without the participation of encryption/decryption operations. But the effectiveness of the scheme needs to be further analyzed. In this paper, we present an exact formula of the effectiveness of ALNCode scheme, which goes a step further than the lower bound given by Wang et al. Based on the exact formula, the influential factors of effectiveness of the scheme are taken into account in details. Moreover, a comparison between the exact effectiveness and the lower bound shows that the lower bound is not true. The reason of the error is analyzed finally.

Keywords: anonymity, linear space, intersection probability, effectiveness analysis, network coding

1. INTRODUCTION

In addition to security, more and more attention has been paid to privacy protection in network communication. In online social networks such as Facebook and Twitter, friends exchanging private messages or contents may not want the others to know whom they are communicating with; In a peer-to-peer file sharing system, anonymous P2P communication is desired to share copyrighted files without revealing one’s identity and risking litigation, as well as to prevent tracking or data mining activities from spammers. According to the current network protocols, anonymous communication system aims to protect personal privacy and guarantee the unlinkability of the sender and the receiver.

Traffic analysis attack, a major passive eavesdropping attack threatening anonymous communication system, has the ability of flow analysis by analyzing the content, size, time correlation of the data packets to obtain information of the sender and the receiver. The traditional methods against traffic analysis attack mainly include encryption, data filling and mix. D. Chaum used the encryption method to protect the content of data packets [1]. M. G. Reed et al. used data filling to protect the packets from size correlation attack [2]. The time correlation among the messages is hidden by mixing the order of messages received at intermediate nodes [3]. However, the time cost of encryption and decryption is high, and the distribution and management of keys is complicated. Also, the time complexity of data filling and mix is comparatively high.

Network coding [4], especially linear network coding (LNC) [5], provides a new perspective to realize anonymous communication and hence enriches traditional mixing...
method. Linear network coding has the feature of intrinsic mixing which can protect the messages without complicated encryption/decryption operations. The encoded packets have the same size as the original ones, so it is easy to resist the packet size correlation attack. The packets will be buffered at intermediate nodes, then coded, and finally sent, which prevents the time correlation attack on the packets.

Considering the fact that network coding breaks the current privacy-preserving protocols (e.g., Onion Routing), Zhang et al. apply the idea of cooperative network and design an anonymity scheme named ANOC [6], which is built upon the classic Onion Routing protocol. Wan et al. propose a privacy preserving scheme based on network coding [7], and the scheme can achieve stronger privacy protection than the mix system while achieving better network performance. However, each encoded packet is associated with a Global Encoding Vector (GEV) in the schemes of [6, 7], and if enough encoded packets are received, the attacker could obtain the path information of a data flow by analyzing the relationship between the upstream and downstream GEVs. Moreover, the routing information in the communication process will reveal information of source and destination node. Therefore, how to conceal the linear relations of GEVs and protect the routing information becomes a main problem of anonymous communication with linear network coding.

In order to protect the routing information, secure routing protocols are designed with methods such as information slicing [8-10], multi-path routing [8-11], opportunistic routing [12] and encryption [13].

To conceal the correlation of GEVs, Fan et al. propose a network coding based privacy-preserving scheme against traffic analysis in multi-hop wireless networks [14, 15]. With homomorphic encryption on GEVs, the scheme could offer two significant privacy-preserving features, packet flow untraceability and message content confidentiality. As to the traditional cryptographic mechanisms, the time complexity of homomorphic encryption is very high for a large network. Based on linear space theory, Wang et al., in 2011, first propose an anonymous linear network coding scheme against traffic analysis attack—ALNCode [16] without employing the existing privacy-preserving techniques, such as Onion Routing in [6] and homomorphic encryption in [14, 15]. It is a crucial work since the scheme provides a novel method to realize anonymous communication with lower time complexity and lower network cost. WANG et al. analyze the effectiveness of ALNCode scheme and give a lower bound of it. However, the exact effectiveness needs to be further studied, which can precisely describe in what degree the scheme is effective. Based on the idea of generation and the theory of linear spaces, effectiveness analysis of ALNCode scheme for anonymous communication with network coding is discussed in this paper. The main contributions of this paper are as follows:

(1) The effectiveness of the ALNCode scheme is further analyzed, and an exact formula of the effectiveness is given.
(2) Based on the exact formula of effectiveness, we consider the influential factors of effectiveness which consists of the number of GEVs received at an intermediate node from some flow “f”, the number of GEVs from other flows “f’”, the number of packets per generation of a flow “h” and the size “q” of a finite field \(\mathbb{F}_q\).
(3) A comparison between the exact effectiveness and the lower bound shows that the lower bound is not true, which is justified by a counter example. After analyzing the
proof in [16], we give the reason of the error.

The remainder of the paper is organized as follows. Section 2 introduces the network model, the attack model, the goals of anonymous communication and the key idea of the ALNCode scheme. In section 3, the effectiveness of ALNCode scheme is discussed, and the exact formula of the effectiveness is given. In section 4, an error about the lower bound is pointed out and the reason of the error is analyzed. The influential parameters of the effectiveness are studied. Section 5 concludes this paper.

2. SYSTEM MODEL

2.1 Network Model

We consider a communication network with multiple unicast flows between multiple pairs of source nodes denoted by \( S \) and destination nodes denoted by \( D \) [16]. Each flow has a unique flow number \( i \) and may go through multiple paths; the paths of different flows may intersect at common intermediate nodes. LNC is applied in the transmission of each unicast flow: a source node partitions the data flow into messages of the fixed size \( H \), and every \( h \) consecutive messages in the flow form a generation with a generation number \( j \). The network model is shown in Fig. 1, in which \( S = \{s_1, s_2, s_3\} \), \( D = \{d_1, d_2, d_3\} \) and \( k \) represents an intermediate node.

![Network model](image)

In a practical communication network, each coded message to be derived is tagged with its routing information, flow number, generation number, and the GEV, which together is referred to as a data packet. We assume all data packets in the network have an equal size. We also assume that a secure routing protocol is in place [13] (It is also similar to the assumptions made in [14, 15]). With such a secure routing protocol, the route of each flow from the source to the destination is decided locally at each intermediate node, who knows only the previous-hop and next-hop nodes along the path. The routing information, flow and generation numbers attached to each coded message are protected by encryption with public/private keys generated locally by each intermediate node and exchanged only among neighbors. On the other hand, GEVs and message contents are not encrypted.

2.2 Attack Model and Anonymous Communication Goals

We consider passive wiretapping attackers from outside of the network with traffic
analysis abilities [14, 15]. For such an outside attacker, we assume it can observe all the packets along all the links in the network and analyze them, attempting to identify sources, destinations, and paths of the flows.

The goals of anonymous communication include: (1) Flow path anonymity. The attacker cannot deduce the flow paths of each flow, i.e., if an attacker observes an upstream packet and a downstream packet at a node, it cannot distinguish whether they are in the same flow or not; (2) Source and destination anonymity. The attacker cannot determine which node each flow originates from or terminates at, i.e., it is not able to tell which nodes in the network (sources) are communicating with which other nodes (destinations).

2.3 Basic Idea of ALNCode

The key idea of ALNCode is to produce new coded messages with obfuscated GEVs at intermediate nodes, which are linearly correlated not only with received GEVs from the same flow, but also those from other flows. Fig. 2 gives an illustration: let $V_{i,k}$ denote the set of GEVs of coded messages received at an intermediate node $k$ from generation $j$ of flow $i$ in the past $T$ time slots, and $L(V_{i,k})$ be the linear span of those GEVs; let $U_{i,k}$ be the set of GEVs received at $k$ from flows other than $i$ before and within the $T$ time slots, and $L(U_{i,k})$ be its linear span.

To produce a coded message for generation $j$ of flow $i$, we seek to generate a GEV $v_{ij}$ that is linear combination of $a$ and $b$, where vector $a$ is generated from $L(V_{i,k})$ and vector $b$ is generated from $L(V_{i,k}) \cap L(U_{i,k})$. In this way, $v_{ij}$ could have linear correlation with both GEVs in $V_{i,k}$ and $U_{i,k}$. If an attacker attempts to trace back the source of the coded packet with GEV $v_{ij}$, both the GEVs in $V_{i,k}$ and $U_{i,k}$ could be correlated with $v_{ij}$, and the attacker would fail to identify which flow the packet actually belongs to.

![Fig. 2. Basic idea of ALNCode.](image)

3. EFFECTIVENESS ANALYSIS OF ALNCODE SCHEME

3.1 Related Propositions

In this part, some conclusions in Linear algebra and Combinatorics needed are presented.

**Proposition 3.1.1** [17] Let $r_1 \leq r_2$ and $A=(a_{ij})$ be an $r_1 \times h$ matrix with rank $r_1$, $B=(b_{ij})$ an $r_2 \times h$ matrix with rank $r_2$. Then the vector subspace $A$ represents is contained in the vector...
Proposition 3.1.2 [18] (i) Let \( r \leq h \) and \( N_q(r, h) \) be the number of \( r \)-dimensional vector subspaces in \( \mathbb{F}_q^h \). Then

\[
N_q(r, h) = \prod_{i=1}^{h} \frac{(q^r - 1)}{(q^i - 1)}. \tag{1}
\]

(ii) Let \( r \leq h \) and \( n_q(r, h) \) be the number of \( r \times h \) matrices of rank \( r \) over \( \mathbb{F}_q \). Then

\[
n_q(r, h) = q^{c(r-1)} \prod_{i=d+1}^{h} (q^i - 1). \tag{2}
\]

(iii) The number of \( h \times h \) matrices of rank \( h \) over \( \mathbb{F}_q \) is \( N_q(h, h) \).

(iv) Let \( 0 \leq k \leq h \) and \( N_q(k, r, h) \) be the number of \( k \)-dimensional vector subspaces contained in a given \( r \)-dimensional vector subspaces of \( \mathbb{F}_q^h \). Then \( N_q(k, r, h) = N_q(k, r) \).

Proposition 3.1.3 (i) The number of bases of \( \mathbb{F}_q^h \) is \( n_q(h, h)/h! \). The number of bases of an \( r \)-dimensional subspace of \( \mathbb{F}_q^h \) is \( n_q(r, r)/r! \).

(ii) The number of the complementary subspaces of an \( r \)-dimensional vector subspace \( P \) of \( \mathbb{F}_q^h \) is \( q^{r(h-r)} \) [17].

Notation 3.1.4 (i) we use \( C_m^n \) to denote the number of ways to select \( n \) objects from a collection of \( m \) different objects, with the order of selection irrelevant and repetition not allowed. It is known to all that \( C_m^n = m^n/(n!(m-n)!)) \).

(ii) we use \( F_m^n \) to denote the number of ways to select \( n \) objects from a collection of \( m \) different objects, with repetition allowed. Then it can be proved that \( F_m^n = C_m^{n+1} \).

Remark 3.1.5 Distinguish the notation \( \mathbb{F}_q^h \) with \( F_m^n \). On the definitions and conclusions in Linear algebra and Combinatorics, please refer to [19, 20], respectively.

3.2 Exact Formula of Effectiveness

We assume that the GEVs received at node \( k \) are randomly selected from the vector space \( \mathbb{F}_q^h \). For the network model, \( h \) is the number of messages for each generation of a flow. In this case, \( P_1 \) is the span space of \( f_1 \) vectors randomly selected from \( \mathbb{F}_q^h \) and \( P_2 \) is the span space of \( f_2 \) vectors randomly selected from \( \mathbb{F}_q^h \). Here \( f_1 \) and \( f_2 \) are the numbers of GEVs received at \( k \) from generation \( j \) of flow \( i \) and from other flows, respectively. In the following, we calculate the probability of \( \dim(P_1 \cap P_2) \neq 0 \), which is crucial to the practical effectiveness of ALNCode.

Theorem 3.2.1 [16] At intermediate node \( k \), an obfuscated GEV \( v_{ij} \) exists for generation \( j \) of flow \( i \), if and only if \( \dim(P_1 \cap P_2) \neq 0 \), that is, \( P_1 \cap P_2 \neq \{0\} \).

Definition 3.2.2 Given a multi-cast network, the effectiveness of the ALNCode scheme
for an intermediate node $k$ is defined as the probability of $\dim(P_1 \cap P_2) \neq 0$, denoted by $p(\dim(P_1 \cap P_2) \neq 0)$. We call this probability the intersection probability.

We use $g(q, r, f)$ to denote the number of ways to select $f$ vectors from $\mathbb{F}_q^r$, with repetition allowed, where the span space of the selected $f$ vectors is the whole space $\mathbb{F}_q^r$ (called the spanning condition). According to the Isomorphism Theorem, the number of ways to select $f$ vectors from $\mathbb{F}_q^r$ satisfying the spanning condition is the same as the number of ways to select $f$ vectors in any $r$-dimensional vector space over $\mathbb{F}_q$ with the spanning condition satisfied.

**Proposition 3.2.3** Let $r, f$ be natural numbers. Then the sequence $\{g(q, r, f)\}_{r \geq 0}$ satisfies the recurrence relation

$$g(q, r, f) = F_q^f - \sum_{i=0}^{r} N_i(r, f)g(q, i, f)$$

with the initial condition $g(q, 0, f) = 1$. Also, if $f < r$, then $g(q, r, f) = 0$.

**Proof:** There are $q^r$ vectors in $\mathbb{F}_q^r$, so the total number of ways to select $f$ vectors from $\mathbb{F}_q^r$, with repetition allowed, is $F_q^f$. Since the spanning condition is required, we conversely find the ways that do not satisfy the condition. The fact that the span space of the selected $f$ vectors is the whole space $\mathbb{F}_q^r$ means that the dimension of the span space is $r$. So, we find the number of ways in which the dimension of the span spaces is less than $r$, that is $0, 1, 2, \ldots, r-1$.

Suppose that the dimension of the span space of selected $f$ vectors is $i$, that is, the rank of the vector group of selected $f$ vectors is $i$. From Proposition 3.1.2 (i), there are $N_i(i, r)$-dimensional subspaces in $\mathbb{F}_q^r$, and hence there are $N_i(i, r)$ ways to select subspaces. Besides, in each subspace, there are $g(q, i, f)$ ways to select $f$ vectors of which the span space is the subspace. Hence, the number of ways to select $f$ vectors with rank $i$ is $N_i(i, r)g(q, i, f)$. According to the above analysis, $g(q, r, f) = F_q^f - \sum_{i=0}^{r} N_i(r, f)g(q, i, f)$.

The initial condition: $g(q, 0, f)$ means the number of ways to select $f$ vectors from a 0-dimensional space, with repetition allowed, where the span space of the selected $f$ vectors is the whole space. Since a 0-dimensional space has only one vector, there exists only one way to select the $f$ vectors which is to choose 0 $f$ times. Thus, $g(q, 0, f) = 1$.

Since the dimension of the span space of $f$ vectors is less than or equal to $f$, $g(q, r, f) = 0$ if $f < r$. \qed

**Lemma 3.2.4** [17] Let $P$ be an $r$-dimensional subspace of $\mathbb{F}_q^h$ and $a_1, a_2, \ldots, a_r$ be a basis of $P$. Expand $a_1, a_2, \ldots, a_r$ to a basis of $\mathbb{F}_q^h$, then the matrix representation of each of the complementary subspaces of $P$ in $\mathbb{F}_q^h$ has the form $(A_{i_1, i_2, \ldots, i_r})$, in which $A_{i_1, i_2, \ldots, i_r}$ is uniquely decided by the complementary subspace of $P$.

**Proposition 3.2.5** Let $r, s$ be integers with $r + s \leq h$, $P$ an $r$-dimensional subspace of $\mathbb{F}_q^h$ and $P_1$ an $s$-dimensional subspace of $\mathbb{F}_q^h$ with $P \cap P_1 = \{0\}$. Then the number of complementary subspaces of $P$ in $\mathbb{F}_q^h$ containing $P_1$ is $q^{(h-r-s)}$.  


Proof: Suppose \( a_1, a_2, \ldots, a_l \) is a basis of \( P \) and \( a_{l+1}, \ldots, a_n \) is a basis of \( P_1 \). Since \( P \cap P_1 = \{0\} \), \( a_{l+1}, \ldots, a_n \) is a basis of \( P \). Expand to \( a_1, a_2, \ldots, a_l, a_{l+1}, \ldots, a_n \), \( a_{n+1}, \ldots, a_m \) which is a basis of \( \mathbb{F}_q^h \). Then the matrix representation of \( P_1 \) under this basis is 

\[
\begin{bmatrix}
B & \mathbf{0} \\
\mathbf{0} & I_{s(h-r-s)}
\end{bmatrix}
\]

Since \( P_1 \subseteq P_2 \), by Proposition 3.1.1, there exists a matrix \( Q = ((Q_1)_{h,h} (Q_2)_{s(h-r-s)} \) such that \( (0_{r \times h}, 0_{s(h-r-s)}) = (Q_1, Q_2) \)

\[
\begin{bmatrix}
B & I_s & \mathbf{0} \\
C & \mathbf{0} & I_{s(h-r-s)}
\end{bmatrix}
\]

is also \( \mathbf{0}_{s(h-r-s)} \). Therefore, the matrix representation is uniquely decided by \( P_2 \), there are \( q^{r(h-r)} \) forms for \( C \), and the number of commentary subspaces of \( P \) containing \( P_1 \) is also \( q^{r(h-r)} \).

\[\square\]

Proposition 3.2.6 Let \( r, s \) be natural numbers with \( r+s \leq h \) and \( P \) an \( r \)-dimensional subspace of \( \mathbb{F}_q^h \). Then the number of \( s \)-dimensional subspaces \( P_1 \) satisfying \( P \cap P_1 = \{0\} \) is

\[
q^{r} \prod_{j=h-r+1}^{h} (q^{r} - 1) \prod_{i=1}^{s} (q^{r} - 1).
\]

Proof: There are \( q^{r(h-r)} \) complementary subspaces of \( P \) in \( \mathbb{F}_q^h \) from Proposition 3.1.3 (ii). In each complementary subspace, there are \( N_s(h, \text{ h-r}) \) \( s \)-dimensional subspaces since the dimension of a complementary subspace of \( P \) is \( h-r \). Any \( s \)-dimensional subspaces \( P_1 \) satisfying \( P \cap P_1 = \{0\} \) is contained in some complementary spaces of \( P \). From Proposition 3.2.5, the number of complementary subspaces of \( P \) containing \( P_1 \) is \( q^{r(h-r)} \). Therefore, the number of \( s \)-dimensional subspaces \( P_1 \) satisfying \( P \cap P_1 = \{0\} \) is

\[
q^{r(h-r)} N_s(h, h-r)/q^{r(h-r)} = q^{r} \prod_{j=h-r+1}^{h} (q^{r} - 1) \prod_{i=1}^{s} (q^{r} - 1).
\]

\[\square\]

Theorem 3.2.7 Let \( P_1 \) be a span space of \( f_1 \) vectors randomly selected from \( \mathbb{F}_q^h \) and \( P_2 \) a span space of \( f_2 \) vectors randomly selected from \( \mathbb{F}_q^h \). The probability of \( \dim(P_1 \cap P_2) \neq 0 \) is

\[
\sum_{0 \leq r \leq h, 0 \leq s \leq h-r} N_s(r, h) \times g(q, r, f_1) \times q^{r(h-r+s)} \times g(q, s, f_2)
\]

\[
1 - \frac{\prod_{0 \leq r \leq h, 0 \leq s \leq h-r} (q^{r} - 1) \times \prod_{i=1}^{s} (q^{r} - 1)}{F_q^h \times F_q^h}.
\]

Proof: The number of ways to randomly select \( f_1 \) and \( f_2 \) vectors from \( \mathbb{F}_q^h \) is \( F_q^h \times F_q^h \). Since \( \dim(P_1 \cap P_2) \neq 0 \) if and only if \( P_1 \cap P_2 \neq \{0\} \), we conversely calculate the number of ways which satisfy \( P_1 \cap P_2 = \{0\} \). Suppose the dimensions of \( P_1 \) and \( P_2 \) are \( r \) and \( s \), respec-
tively. Then, the rank of the vector group of the \( f_1 \) and \( f_2 \) vectors is \( r(0 \leq r \leq \min\{h, f_1\}) \) and \( s(0 \leq s < \min\{h, f_2\}) \), respectively. If \( P_1 \cap P_2 = \{0\} \), then \( r + s = \dim P_1 + \dim P_2 = \dim (P_1 \cap P_2) + \dim (P_1 \oplus P_2) = \dim (P_1 \oplus P_2) \leq h \).

The number of \( r \)-dimensional subspaces of \( \mathbb{F}_q^h \) is \( N_q(r, h) \). The ways of \( f_1 \) vectors selected in the chosen \( P_1 \) is \( g(q, r, f_1) \). Hence, the total number of ways to choose \( f_1 \) vectors is \( N_q(r, h) 	imes g(q, r, f_1) \). Once \( P_1 \) is chosen, from Proposition 3.2.6, there are \( q^n \prod_{i=k+1}^{h-r} (q'^i - 1) \) \( s \)-dimensional subspaces \( P_2 \) satisfying \( P_1 \cap P_2 = \{0\} \), which means that the number of ways to choose \( P_2 \) is \( q^n \prod_{i=k+1}^{h-r} (q'^i - 1) \). The ways to choose \( f_2 \) vectors in the chosen \( P_2 \) is \( g(q, s, f_2) \). Hence, the total number of ways to choose \( f_2 \) vectors is \( q^n \prod_{i=k+1}^{h-r} (q'^i - 1) \) \( g(q, s, f_2) \).

Thus, the number of ways to choose \( f_1 \) and \( f_2 \) vectors satisfying \( P_1 \cap P_2 = \{0\} \) is

\[
\sum_{0 \leq r \leq \min\{h, f_1\}} \sum_{0 \leq s < \min\{h, f_2\}} N_q(r, h) \times g(q, r, f_1) \times q^n \prod_{i=k+1}^{h-r} (q'^i - 1) \prod_{j=1}^{\min\{h-r, f_2\}} (q'^j - 1) \times g(q, s, f_2)
\]

Therefore, the probability of \( \dim (P_1 \cap P_2) \neq 0 \) is

\[
1 - \frac{F_{\phi}^v \times F_{\phi}^v}{C_{\phi}^v \times C_{\phi}^v}
\]

According to Definition 3.2.2 and Theorem 3.2.7, the effectiveness of the ALNCode scheme for an intermediate node is calculated by Eq. (5).

Generally, for the purpose of decodability at the destination node, the GEVs in a generation of a flow must be linearly independent. We study the effectiveness of the ALNCode scheme in this case. Since GEVs from the flows other than flow \( i \) may be linearly independent, we still assume that the \( f_2 \) vectors are randomly selected from \( \mathbb{F}_q^h \).

**Corollary 3.2.8** Suppose \( f_1 \leq h \). Let \( P_1 \) be a span space of \( f_1 \) linearly independent vectors randomly selected from \( \mathbb{F}_q^h \) and \( P_2 \) a span space of \( f_2 \) vectors randomly selected from \( \mathbb{F}_q^h \).

The probability of \( \dim (P_1 \cap P_2) \neq 0 \) is

\[
\sum_{0 \leq r \leq \min\{h, f_1\}} N_q(f_1, h) \times \frac{n_q(f_1, f_1)}{(f_1)!} \times q^r \prod_{i=k+1}^{h-r} (q'^i - 1) \prod_{j=1}^{\min\{h-r, f_2\}} (q'^j - 1) \times g(q, s, f_2)
\]

\[
1 - \frac{C_{\phi}^v \times F_{\phi}^v}{C_{\phi}^v \times F_{\phi}^v}
\]

**Proof:** The subspace spanned by \( f_1 \) linearly independent vectors is \( f_1 \)-dimensional, hence
there are $N_q(f_1, h)$ ways to choose subspaces. Once a subspace is chosen, to choose $f_1$ linearly independent vectors in it means to choose a basis. From Proposition 3.1.3 (i), the number of basis of an $f_1$-dimensional subspace is $n_q(f_1, f_1)/(f_1)!$, and so is the number of ways to choose $f_1$ linearly independent vectors. The rest of the proof is the same as that in Theorem 3.2.7, and we neglect it.

According to Definition 3.2.2 and Theorem 3.2.8, for a given multiple unicast network with the GEVs linearly independent in the same generation of some flow, the effectiveness of the ALNCode scheme for an intermediate node is calculated by Eq. (6).

Fig. 3 shows the effectiveness of two cases, “randomly chosen” and “linearly independent” $f_1$ vectors. It can be seen that the effectiveness of “linearly independent” is higher than that of “randomly chosen”. Therefore, the GEVs for a flow should be chosen linearly independent not only for decodability but also for high effectiveness.

4. SIMULATION AND NUMERICAL RESULTS

4.1 Lower Bound and Exact Formula of Effectiveness

In [16], Theorem 3 gives a lower bound of the probability $\dim(P_1 \cap P_2) \neq 0$ which is

$$p(\dim(P_1 \cap P_2) \neq 0) \geq \sum_{r=0}^{\max\{h, f_1\}} [(N_q(r, f_1)q^{(r-1)/2}) \prod_{j=h-1}^{f_1} (q^j - 1)]$$

$$\times (1 - \sum_{g=0}^{\max\{h, f_2\}} (N_q(g, f_1)q^{(g-1)/2}) \prod_{j=h-1}^{f_1} (q^j - 1))).$$

We compare the exact formula of effectiveness Eq. (5) and the lower bound Eq. (7) in Fig. 4, where $q=2$, $h=5$ and $f_1=6$. Fig. 4 shows that the lower bound given by [16] is higher than the exact effectiveness. This is a contradiction. A counter example is given below.

Take $h=2$, $q=2$, $f_1=1$, $f_2=2$. According to Eq. (7), the lower bound is $27/64$. The exact effectiveness is $3/8$ by Eq. (5). In fact, $\mathcal{F}_2^2=\{(0,0),(0,1),(1,0),(1,1)\}$. For simplicity,
we use 0, 1, 2, 3 to denote (0,0), (0,1), (1,0), (1,1), respectively. The ways to randomly choose 1 vector and 2 vectors in $\mathbb{F}_2^2$ are {0}{2•0}, {0}{0,1}, {0}{0,2}, {0}{0,3}, {0}{2•1}, {0}{1,2}, {0}{1,3}, {0}{2•2}, {0}{2,3}, {0}{2•3}, {1}{2•0}, {1}{0,1}, {1}{0,2}, {1}{0,3}, {1}{2•1}, {1}{1,2}, {1}{1,3}, {1}{2•2}, {1}{2,3}, {1}{2•3}, {2}{2•0}, {2}{0,1}, {2}{0,2}, {2}{0,3}, {2}{2•1}, {2}{1,2}, {2}{1,3}, {2}{2•2}, {2}{2,3}, {2}{2•3}, {3}{2•0}, {3}{0,1}, {3}{0,2}, {3}{0,3}, {3}{2•1}, {3}{1,2}, {3}{1,3}, {3}{2•2}, {3}{2•3}, {3}{2•4}. Hence, the number of ways to randomly choose 1 vector and 2 vectors is 40. The ways satisfying $P_1 \cap P_2 = 0$ are {0}{2•0}, {0}{0,1}, {0}{0,2}, {0}{0,3}, {0}{2•1}, {0}{1,2}, {0}{1,3}, {0}{2•2}, {0}{2,3}, {0}{2•3}, {1}{2•0}, {1}{0,2}, {1}{0,3}, {1}{2•2}, {1}{2•3}, {1}{2•4}, {2}{2•0}, {2}{0,1}, {2}{0,3}, {2}{2•1}, {2}{2•3}, {3}{2•0}, {3}{0,1}, {3}{0,2}, {3}{0,3}, {3}{2•1}, {3}{2•2}. Hence, the number of ways which satisfy $P_1 \cap P_2 = 0$ is 25. The probability of $P_1 \cap P_2 \neq 0$ is $1 - 25/40 = 3/8$. Therefore, the lower bound given in [16] is wrong. Next, we analyze the reason of the error in [16].

Denote the set of the $f_1$ and $f_2$ randomly selected vectors by $\mathcal{I}_1$ and $\mathcal{I}_2$, and the span spaces $P_1 = L(\mathcal{I}_1)$ and $P_2 = L(\mathcal{I}_2)$, respectively. Let $\mathbf{I}_1$ and $\mathbf{I}_2$ be the matrix whose rows are the $f_1$ and $f_2$ randomly selected vectors in an arbitrary order, respectively. In [16], the authors used the Total Probability Formula to calculate the intersection probability,

$$p(\dim(P_1 \cap P_2) \neq 0) = 1 - p(\dim(P_1 \cap P_2) = 0) = 1 - \sum_{r=0}^{\min\{f,h\}} p(\dim P_1 = r) p(\dim P_2 = 0 | \dim P_1 = r) \cdot$$

In a multiple set of vectors, the order of each vector is irrelevant, while if the order of rows changes, the matrix may be totally different. So $p(\text{rank} \mathbf{I}_1 = r)$ is not equal to $p(\text{rank} \mathbf{I}_1 = r)$ in general (see Propositions 4.1.1 and 4.1.2 below). Here, we should use $p(\text{rank} \mathbf{I}_1 = r)$, instead of $p(\text{rank} \mathbf{I}_1 = r)$. Therefore,

$$p(\dim(P_1 \cap P_2) \neq 0) = 1 - \sum_{r=0}^{\min\{f,h\}} p(\text{rank} \mathbf{I}_1 = r) p(\dim(P_1 \cap P_2) = 0 | \text{rank} \mathbf{I}_1 = r).$$

**Proposition 4.1.1** [16] Let $0 \leq r \leq \min\{f,h\}$. For any $f \times h$ dimensional matrix $\Lambda$ whose elements are randomly selected from finite field $\mathbb{F}_q$, the probability that $\text{rank} \Lambda = r$ is

$$p(\text{rank} \Lambda = r) = N_f(r,f) \times \prod_{i=0}^{h-1} (q^f-1) \times q^{f(h-r)}. \quad (8)$$

**Proposition 4.1.2** Let $0 \leq r \leq \min\{f,h\}$. For any multiple set $\Gamma$ whose elements are $f$ vectors randomly selected from $\mathbb{F}_q^h$. The probability that $\text{rank} \Gamma = r$ is

$$p(\text{rank} \Gamma = r) = \frac{N_f(r,h) \times g(q,r,f)}{F^f_h}. \quad (9)$$

Fig. 5 shows the different between $p(\text{rank} \Lambda = r)$ and $p(\text{rank} \Gamma = r)$, where $q=2$, $r=3$, $h=5$ and $f$ varies from 3 to 8.
4.2 Influential Parameters

During the process of anonymous communication with network coding, the influential parameters on effectiveness of ALNCode scheme mainly include: the number of GEVs from an upstream information flow “f₁”; the number of GEVs from other flows “f₂”; the number of encoded messages per generation “h”; the size “q” of finite field \( \mathbb{F}_q \).

To provide a better idea of the effectiveness with its influential parameters, we show the effectiveness at different values of \( f₁, f₂, h \) and \( q \). See Fig. 6.

Figs. 6 (a) and (b) show that the effectiveness increases with the increase of \( f₁ \) and \( f₂ \), respectively. Moreover, from (a), we know that when \( f₁ > 5 \), the effectiveness is nearly 1; from (b), when \( f₂ > 15 \), the effectiveness is nearly 1. The reason is straightforward: when \( h \) and \( q \) are fixed, the more GEVs a node receives from the current flow and from other flows, the higher the intersection probability is.

Fig. 6 (c) shows that the effectiveness increases first and then decreases with the in-
crease of $h$, while Fig. 6 (d) demonstrates different trends with the increase of $q$ in different cases. In particular, for any $f_1, f_2 > 0$, it shows that when $f_1 + f_2 \leq h$, the effectiveness decreases with the increase of $q$; and when $f_1 + f_2 > h$, the effectiveness increases with the increase of $q$.

The above results can guide the practical selection of the influential parameters in engineering.

(i) A reasonable routing protocol should be designed to get larger $f_1$ and $f_2$ in order to achieve higher effectiveness.

(ii) In general, an intermediate node may buffer sufficient number of messages. When divided into generations, a reasonably small $h$ should be chosen to guarantee a good effectiveness, as well as low decoding complexity.

(iii) The finite field size $q$ can then be set according: if the number of GEVs received at each node satisfies $f_1 + f_2 > h$, a relatively large $q$ can be used, but not too large considering the communication overhead and decoding complexity; if $f_1 + f_2 \leq h$, we can simply select $q=2$ to get the best effectiveness.

5. CONCLUSION

Considering the importance of anonymous communication scheme based on linear network coding, we study ALNCode—a network coding with non-encryption mechanism. First of all, the network model, the attack model, the goals of anonymous communication and the key idea of ALNCode are introduced in the section of system model; Secondly, the effectiveness of ALNCode scheme is discussed. The exact formula of effectiveness is given, which is much better than the lower bound. Finally, an error about the lower bound is pointed out and the reason of the error is analyzed. The influential parameters of the effectiveness is studied. This work gives a quantitative index for successive analysis on the effectiveness of anonymous communication scheme, and also provides a theoretical guidance for applications of anonymous communication system in engineering.

REFERENCES


Yihui Zhou (周翼辉) received her MS and Ph.D. degrees in College of Mathematics and Information Science at Shaanxi Normal University, China in 2006 and 2009, respectively. Her research focuses on network science and topology on lattice.

Xuening Liao (廖雪宁) is a Ph.D. Candidate in School of Computer Science at Shaanxi Normal University, China. Her research focuses on anonymous communication and network coding.

Zhenqiang Wu (吴振强) is a Professor in the School of Computer Science, Shaanxi Normal University, Xi’an, China. He obtained his Ph.D. degree in Computer Science and Application from School of Computer Science and Technology, Xidian University, China in 2007. His main research interests include computer networks, network security, network coding and its applications.