Seventh International Workshop on Optimal Codes and Related Topics
September 6-12, 2013, Albena, Bulgaria *pp. 85-90*

Anonymous and secure network coding scheme[1]

E. M. Gabidulin ernst_gabidulin@yahoo.com
Moscow Institute of Physics and Technology (State University)
O. V. Trushina oksana.trushina@gmail.com
Moscow Institute of Physics and Technology (State University)

Dedicated to the memory of Professor Stefan Dodunekov

Abstract. We consider the problem of anonymous transmission against eavesdropper in network with linear network coding. Our problem is to achieve information flow untraceability in network with multiple flows. We propose the new simple anonymous scheme based on coset coding. We show that this coset coding scheme allows to hide correlation between incoming and outgoing packets by adding random message in such a way that the destination node can recover easily information messages. We also discuss a possible attack.

1 Introduction

In addition to *confidentiality*, *anonymity* is also an important concern of secure data communication. There are three anonymity models given by [1]:

- it is impossible to ascertain whether a communication exists, such situation is called *unobservability*;
- it is impossible to identify the sender/receiver of data flow, but existence of communication session is clear, it is referenced as *sender/receiver anonymity*;
- it is impossible to relate a sender and a receiver of a communication on the assumption that a sender or a receiver may be identified, it is called *relationship anonymity*.

Unobservability provides for strongest privacy protection, while relationship anonymity is the weakest model among the three. However, providing either unobservability or sender/receiver anonymity may results in network performance

[1]This research is partially supported by the Russian Foundation for Basic Research, project № 12-07-00122-a.

degradation. An approach to achieve these models of privacy often involves injecting dummy packets and probabilistically forwarding the packets, which lead to significant network bandwidth consuming and large delays correspondingly.

In this work we focus on relationship anonymity when it is impossible to trace the messages. It is sufficient for many applications.

Ensuring anonymity consists of two parts. One is to setup a path of data transmission confidentially. Only intended nodes know path setup information and every such node owns the information as few as possible for the right operating. The other is to guarantee the message forwarding to be untraceable.

We concentrate on the second part and assume the forwarding paths to be established confidentially.

To provide relationship anonymity the input and output packets of a relay node on a path must not be linked. In linear network coding incoming packets are mixed at relay nodes. It hides correlation between incoming and outgoing packets. But a dependance between this packets is linear. Thus adversary may easy relate them and reveal information of the packet forwarding path. Therefore, on the assumption that forwarding paths are established confidentially the main task is to hide correlation between incoming and outgoing packets.

To achieve this goal there are two key approaches proposed for network coding. The first is encryption of *global encoding vectors* [2]. The second is to make global encoding vectors belonging to different flows to be similar [3].

The present work offers method based on secure scheme proposed by Silva and Kschischang in [4]. This secure scheme is based on Ozarow-Wyner coset coding scheme with rank-metric codes. According to this scheme a real message to be transmitted is a syndrome of message being transmitted through the network. The Silva-Kschischang scheme guarantees source information to remain information-theoretically secure from adversary eavesdropping certain number of links.

The idea of our approach is to modify the Silva–Kschischang scheme in such a manner that the incoming and outgoing messages of any relay node are statistically independent. Therefore it is possible to achieve relationship anonymity and security simultaneously.

2 Network and adversary models

We start by describing our network model and adversary model against which our anonymity scheme is designed.

We consider a network represented by a directed multigraph. Vertices are network nodes, edges are error free links. There are multiple source nodes and multiple destination nodes. We assume routes between sources and destinations to be established confidentially by some secure routing protocol.

Let h sources be in the network. Every source node produces n packets X consisting of m symbols from a finite field $\mathbb{F}_q$. The vector space isomorphism

$\mathbb{F}_q^{1\times m} \cong \mathbb{F}_{q^m}$ is used. Then packets are regarded as elements of the extension field $\mathbb{F}_{q^m}$ and packets of a source are regarded as vectors over $\mathbb{F}_{q^m}$, $X \in \mathbb{F}_{q^m}^n$. Every source node transmits the linear combinations of the packets to its outgoing links. The coefficients of this combinations are the elements of field F_q. Any relay node transmits the linear combinations of its incoming packets to its outgoing links. Its own set of coefficients corresponds to any link e. This set of coefficients forms a *global incoming vector* $c_e \in F_q^{1\times n}$. Then over link e packet c_eX is forwarded. We assume that the relay nodes can distinguish packets owned to different sources.

For adversary to be specified it is necessary to make preliminary consideration of Ozarow-Wyner coset coding scheme.

Let $S \in F_{q^m}^k$ be a message to be forwarded confidentially. To do that the message is transformed to a message $X \in F_{q^m}^n$ using parity-check matrix of $[n, n-k]$ linear code $H \in F_{q^m}^{k\times n}$ like $S = HX$.

Suppose that there are h source nodes. Denote $\mathbf{S} = \begin{pmatrix} S_1 & S_2 & \dots & S_h \end{pmatrix}^\top$ and $\mathbf{X} = \begin{pmatrix} X_1 & X_2 & \dots & X_h \end{pmatrix}^\top$. Then $\mathbf{S} = \mathbf{HX}$, where

$$\mathbf{H} = \begin{pmatrix} H_1 & 0 & 0 & \dots & 0 \\ 0 & H_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & H_h \end{pmatrix}$$

Consider the case where the adversary is allowed to observe $\mathbf{Z} = \mathbf{WX}$. Assume that the random vector $\mathbf{X}$ has the uniform distribution. For no information to be leaked to adversary, the vectors $\mathbf{S}$ and $\mathbf{Z}$ must be *statistically* independent. This condition is satisfied if matrices $\mathbf{H}$ and $\mathbf{W}$ are *linearly* independent. In other words, it must be $\mathrm{Rk}\begin{pmatrix} \mathbf{H} \\ \mathbf{W} \end{pmatrix} = \mathrm{Rk}(\mathbf{H}) + \mathrm{Rk}(\mathbf{W})$. For example, this equality is valid, if $\mathbf{W}$ is a block diagonal matrix and matrices H_i, W_i are pairwise linearly independent. It means that adversary is allowed to eavesdrop certain piece of message of each source node. In such case it is said that scheme is *perfect secure under some number of observations*.

3 Anonymous scheme

In this section we discuss our approach in detail.

Consider one source node. Let $S \in \mathbb{F}_{q^m}^k$ be the packets, which must be forwarded anonymous. To do this Silva-Kschischang scheme is used in the following way.

Let C be an $[n, n-k]$ linear maximum-rank-distance code over $\mathbb{F}_{q^m}$ with parity-check matrix $H \in \mathbb{F}_{q^m}^{k\times n}$, $m \geq n$. Silva and Kschischang proved the coset coding scheme based on H to be perfect secure under $\mu \leq n-k$ observations.

The vector S is transformed to the new vector $X \in \mathbb{F}_{q^m}^n$ as follows. The source node chooses uniformly at random and independent from S a vector $V_1 \in \mathbb{F}_{q^m}^{(n-k)}$. The vector X is produced by

$$X = T \begin{pmatrix} S \\ V_1 \end{pmatrix},$$

where $T \in \mathbb{F}_{q^m}^{n \times n}$ is an invertible matrix such that

$$T^{-1} = \begin{pmatrix} H \\ L \end{pmatrix}$$

for some $L \in \mathbb{F}_{q^m}^{(n-k) \times n}$.

The matrix T may be represented as

$$T = \begin{pmatrix} T_1 & T_2 \end{pmatrix},$$

where $T_1 \in \mathbb{F}_{q^m}^{n \times k}$, $T_2 \in \mathbb{F}_{q^m}^{n \times (n-k)}$, $\mathrm{Rk}(T_1) = k$, $\mathrm{Rk}(T_2) = n - k$. Then

$$X = \begin{pmatrix} T_1 & T_2 \end{pmatrix} \begin{pmatrix} S \\ V_1 \end{pmatrix} = T_1 S + T_2 V_1. \tag{1}$$

Consider the first relay node. Its incoming message is X. The relay node produce the new message X' by choosing $V_2 \in \mathbb{F}_{q^m}^{n-k}$ uniformly at random and independently of X and calculating $X' = X + T_2 V_2$. In that case X' is uniformly distributed and mutual information between X and X' satisfies $I(X; X') = 0$. It means that it is impossible to relate incoming message X and outgoing message X'. Every relay node behaves similarly computing its outgoing message as sum of its incoming message and $T_2 V$, where V is a randomly chosen vector.

Since several information flows from several sources enter in the relay node then $I(X_i; X'_j) = 0$, $i, j = 1, 2, \ldots r$, r – number of entering flows. By owning some outgoing message the adversary can not guess which incoming message it corresponds to. Consequently it is impossible for adversary to trace the message. Thereby our approach provides relationship anonymity and perfect secrecy under μ observations simultaneously.

We consider the case when the matrix T_2 is the same for all information flows and each information flow has its own H, there are no two equal matrices H among all information flows. Since

$$T^{-1} T = \begin{pmatrix} I_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$$

then $HT_1 = I_k$, $\quad HT_2 = 0$. The matrix T_2 being known for everyone doesn't compromise secrecy. An estimation of how many matrices T_1 are there given T_2 is discussed below.

The dimension of the subspace spanned by columns of T_2 is $n-k$. The columns of the matrix T_1 must be linearly independent on the matrix T_2 and each other. The first column of T_1 may be chosen in $q^n - q^{n-k}$ ways. The second in $q^n - q^{n-k+1}$ ways. The third in $q^n - q^{n-k+2}$ ways and so forth. The last k-th column may be chosen in $q^n - q^{n-1}$ ways. Then the number of matrices T_1 such that for given T_2 the matrix $\begin{pmatrix} T_1 & T_2 \end{pmatrix}$ is equal to

$$(q^n - q^{n-k})(q^n - q^{n-k+1})(q^n - q^{n-k+2}) \dots (q^n - q^{n-1}) = q^{nk} \prod_{i=1}^{k} \left(1 - \frac{1}{q^i}\right).$$

Typical practical network coding uses following parameters $q = 2^8,\ n \sim 50$ [5]. Then number of matrices $T_1 \sim q^{nk}$.

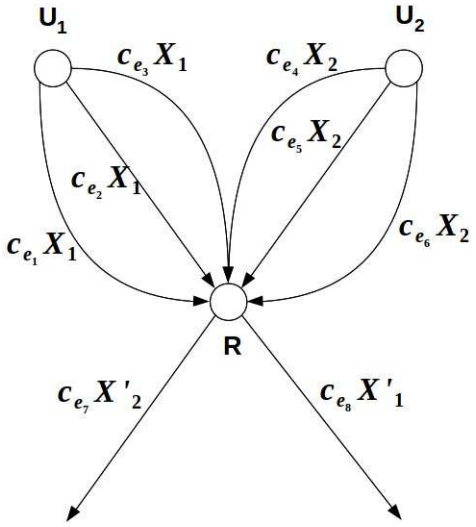

Figure 1: Example

The following example illustrates the basic idea of the scheme.

Example 1. *Let* $q = 2,\ m = n = 3,\ k = 1,\ \mu = n - k = 2$. $\mathbb{F}_{2^3}$ *is generated by a root of* $p(x) = x^3 + x + 1,\ p(\alpha) = 0$. *There are two sources* U_1 *and* U_2 *(Fig. 1). Matrix* H_1 *corresponds to the information flow from* U_1 *and* H_2 *corresponds to the information flow from* U_2. *The matrix* T_2 *is given.* $S_1,\ S_2 \in \mathbb{F}_{2^3}$ *are the messages of* U_1 *and* U_2 *correspondingly.* $S_1 = H_1 X_1$ *and* $S_2 = H_2 X_2,\ X_1, X_2 \in \mathbb{F}_{2^3}^3$. *The eavesdropper is allowed to intercept any* $\mu = 2$ *outgoing links of each node. Suppose the eavesdropper intercepts links* $e_1,\ e_3,\ e_4,\ e_5,\ e_7$ *and* e_8. *In that case*

it has messages $c_{e_1}X_1$, $c_{e_3}X_1$, $c_{e_4}X_2$, $c_{e_5}X_2$ *and* $c_{e_7}X_2'$, $c_{e_8}X_1'$. *According to Silva-Kschischang scheme by owning* $c_{e_1}X_1$ *and* $c_{e_3}X_1$ *it is impossible to find* S_1, *similarly for* S_2. *Denote* $\begin{pmatrix} c_{e_7} \\ c_{e_8} \end{pmatrix} \begin{pmatrix} X_1 + T_2V_1 \\ X_2 + T_2V_2 \end{pmatrix}$ *as* W_1, $\begin{pmatrix} c_{e_7} \\ c_{e_8} \end{pmatrix} \begin{pmatrix} X_2 + T_2V_3 \\ X_1 + T_2V_4 \end{pmatrix}$ *as* W_2 *and adversary observation as* W. *Then* $P(W = W_1) = P(W = W_2) = \frac{1}{2}$.

4 Conclusion

In this work we have addressed to the problem of achieving anonymity and security simultaneously. We proposed the scheme based on the Silva-Kschischang secure scheme which realizes an interesting idea: to break correlation between incoming and outgoing messages at a relay node by producing outgoing message as the sum of incoming and random messages. It was given an example. We showed that our scheme provides anonymity and security simultaneously without increasing the decoding complexity for the authorized destination.

The authors are thankful to Dr. Nina Pilipchuk for discussions.

References

[1] A. Pfitzmann, M. Hansen, A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management, http://dud.inf.tu-dresden.de/literatur/Anon_terminology_v0.34.pdf, 2010.

[2] Y. Fan, Y. Jiang, H. Zhu, J. Chen, X. Shen, Network coding based privacy preservation against traffic analysis in multi-hop wireless networks, in *IEEE Transactions on Wireless Communication,* IEEE, Piscataway, 2011, 834–843.

[3] J. Wang, J. Wang, C. Wu, K. Lu, N. Gu, Anonymous communication with network coding against traffic analysis attack, in *Proc. 30th IEEE International Conference on Computer Communications, 2011, China,* IEEE, 2011, 1008–1016.

[4] D. Silva, R. Kschischang, Univelsal secure network coding via rank-metric codes, *IEEE Transactions on Information Theory,* vol. 57, no.2, 1124–1135, 2011.

[5] P. A. Chou, Y. Wu, K. Jain, Practical network coding, *Proc. Allerton Conf. on Commun., Contr., Comput., Monticello, IL,2003,* 40–49, 2003.