Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/070507

A Method for Low-overhead Secure Network Coding

Song Fei[*] *and Cui Zhe*[*]

Chengdu Institute of Computer Application, Chinese Academy of Sciences, China

Received: 26 Dec. 2012, Revised: 28 Apr. 2013, Accepted: 29 Apr. 2013
Published online: 1 Sep. 2013

Abstract: This paper presents a low-overhead secure network coding scheme. The randomness of chaotic sequence is extremely sensitive to initial conditions characteristic of the program, the combination of the chaotic sequence with the original source message vector to construct a new cryptosystem to achieve the perfect secrecy of the encoding scheme. The program only in the original random network coding system based on the source changes, the intermediate nodes remain unchanged; added only an interference signal at the source to guarantee the security of the encoding scheme. The analysis results show that the scheme can guarantee perfect secrecy, and send signals with minimal overhead.

Keywords: Network coding, low-overhead, Chaos Optimization Algorithm

1 Introduction

2000s, Ahlwede [1] based on the concept of network information flow, first proposed the idea of network coding. I.e., in a network, each intermediate node has the coding capacity to the data packet it received and then transfer out the processed information. Network coding is divided into two main categories:Linear network coding and non-linear network coding. Linear network coding structure is simple and practical advantages, so this paper focus on research and discuss linear network coding. The initial purpose of network coding to achieve maximum network flow and improve network throughput. However, after deeper research found a good way of network coding is also a secure network transmission [1,2]. In 2002, Cai [3] first given a communication model that on network eavesdropping, put forward a necessary and sufficient condition for building the linear safety network coding and constructed a secure coding of network information theory. J.Feldman [4,5] discussed in detail and described the link between security network coding alphabet, multicast rate and coding overhead. Zhang [6] mixed the original source information and the vector of the same length with randomly assigned, and transmitted the mixed random vector in order to ensure the safety of the source message. In above-described program the most secure network coding scheme are assumed that the attacker can eavesdrop on a separate channel. Therefore, in order to establish a secure network transmission system, it need to select the k random vectors from the same number of field coding with the k source information vector to be encoded. In this way, the transmission must be allocated a portion of the bandwidth to transmit interference information (that means there are k random vectors), So, occurred a considerable amount of coding overhead. Secure network coding disguised data during execution, and effectively carried the data, but at the same time it increases the coding overhead and complexity of the network nodes.

In order to improve the robustness of the network coding, while reducing the overhead, [7]. focus on analysis for the characteristics of the mobile random network and unreliable network and the noise network, proposed a secure network coding that using the training sequence embedded in the source data and combined with the channel encoder. A low-complexity secure network coding that proposed by Xu [8] using of the nature of the sparse matrix to reduce the complexity of the node, however the program still added a considerable amount of confidential redundant and coding overhead is not reduced. Adeli [9] put forward a small overhead secure network coding based on Hash function, Although in his text, the security of the scheme and the security conditions has been discussed, but the program is still exist important security vulnerabilities. A security theorem proposed by K. Jain [10], they proved messages are not malicious attacker to obtain using a pseudo-random function. In summary, the paper first

[*] Corresponding author e-mail: asfei1031@gmail.com, Zhecui@163.com

discusses and select the appropriate chaotic map and combining cryptography and then added to the initial information into the initial information of the source.

Through theoretical analysis and a security proof of the theorem, the coding scheme using a unique random number and a chaotic sequence with the source, and then encoded by a suitable network reaches perfect secrecy, it can be ensured secure communications.

2 Knowledge and definition

This paper mainly focuses on a class of acyclic, delay-free single-source multi-sink multicast communication network, in order to simplify the analysis of the problem. For an acyclic Multicast Network $G=(V,E)$ V Is a collection of points, E is a collection of channel, for constructing the network code, assume $GF(q)$ is a q rank finite field (q is a large prime number).

Definition 1. (Network coding: Assume $G=(V,E)$ is a directed network, the n-dimensional linear network code on the G will assign a vector $u(e)$ to each edge $e \in E$ of G , and make it has the following properties. (called $u(e)$ "Global encoding nuclear" of channel e).

(1) Allocation n-dimensional vector space $\tau(X)=GF_n(q)$ to the information source X;

(2) The vector of edge $e \in E$ is a linear combination of end node $tail(e)$ which reached edge e ; If information source X is a end node of e , then the vectors of e will select from $\tau(X)$;

(3) There must be has a specific decoding function operation that can be able to recover the source information at every nodes of information source.

Assume the vector space of node S is $\tau(S)$, for any non-source node S,$\tau(S)$ is the collection of all the linear combination of the vectors which arrived each edge of S . So for $e \in E$ there is $u(e) \in (tail(e))$.

Definition 2. Network coding: Information source X generating the non-compressed data information marked as m . All of m consisted of the vector n that expressed as $m=(m_1,m_2,...,m_n)^T$. Then the information will be encoded make use of the coding algorithm and these data information will be sent to the information sink through the network. (Encoding algorithm mentioned above will be described in Section 4).

Definition 3. Network coding: For information sink T: There are t sinks in the network. The purpose of the network coding is the secure transmission the messages sent by the source to the sink node.

Definition 4. Network coding: Attacker C : Attacker C has a limited eavesdropping capability, it can be able to eavesdrop on k channel. $\Omega=(\alpha_1,\alpha_2,...,\alpha_k)$ indicate the collection of tapped channels.

According to the above definitions and knowledge, we can construct a linear network coding. Each edge of this network has a coding vector v_i , Where, i represents a number of edges in the network, that i from 1 to $N=||E||$, then

$$v_i=(v_{i1},v_{i2},...v_{in})^T \tag{1}$$

These encoding vectors are selected from a same coding domain. The information that transmitted on channel is represented as x_i , x_i is a linear combination of all the information, then

$$x_i=v_i^T m, i=1,2,...,N \tag{2}$$

3 The original secure network coding scheme

Assume the attacker can eavesdrop up k independent edges (The independent edges mean the encoding vector of each edge is linear independent), In order to prevent an attacker to obtain meaningful information, Select α random number from the encoding domain add to the information source, expressed as:

$$\bar{m}=(m_1,m_2,...,m_{n-\alpha},z_1,z_2,...,z_\alpha) \tag{3}$$

Review paper [3] to [6] have been proposed to achieve the perfect secrecy secure network coding scheme. Their basic approach is to ensure that the attacker will make redundant information adding to information is not invalid. In order to achieve this purpose, they assume:

$$V\bar{m}=[v_1,v_2,...v_k]^T\bar{m}=(x_1,x_2,...,x_k)^T \tag{4}$$

Among Expression (4), v_i is corresponding n -dimensional coding vectors of edge i . V is a matrix of these encoding vector combination.

If x is divided into two sub-matrix, then $A_{k\times(n-\alpha)}$ corresponds to the original information source $\bar{m}$, corresponds to the redundant information, so,

$$[A_{k\times(n-\alpha)}|B_{k\times\alpha}]\bar{m}=(x_1,x_2,...,x_k)^T \tag{5}$$

To make the attacker can not get useful information by eavesdropping, so requirements all the columns in the sub-matrix are linearly independent, then $\alpha \geq k$. So, the number of redundant information in $\bar{m}$ should be not less than k .

The existing scheme is added a random number into the encoding vector and then make it linear combination with information. This approach makes network secure coding exist a lot of coding overhead. Moreover, for the information added redundant, the attacker can undermine its security by take more information of independent side. In this paper The basic method of low overhead, secure network coding is: chaotic sequence and network coding combination structure "one-time pad" encryption system.

4 chaotic sequence

Chaotic sequence is a pseudo-random sequence, in formal simply performance to a disordered state. However, the chaotic sequence is a kind of ordered structure of a rich internal level, just did not show obvious periodicity and symmetry.

Chaotic sequence has the following characteristics mainly [11]:

(1) The chaotic sequence is generated by determining the differential form or discrete form of equation, it can generate and receive control by the user.

(2) Chaos in nonlinear mapping or nonlinear systems in order to produce, and Chaos mapping status is formed under the repeated separation and folding, so the chaotic mapping relationship is not reversible certainly.

(3) Chaotic systems are extremely sensitive to initial values. That replace and modify the chaotic sequence can increase the system's capacity of resistance to decipher.

Fully understand the characteristics of chaotic sequence more than several, it can be applied to the field of secure communication as a method of encryption. Shannon [12] and some domestic scholars has been applied chaos theory to cryptography, but how to select the chaotic mapping that meet the requirements of cryptographic properties is a key issues need to solved.

There are three main chaotic map: Logistic map, improved Logistic map and Chebyshev map. Their expression is shown in Table 1.

Table 1. Several main chaotic map.

chaotic map	expression
Logistic map	$x_{n+1}=\lambda\cdot x_n(1-x_n),\ \ x_n\in(0,1)$
improved Logistic map	$x_{n+1}=1-2\cdot x_n^2,\ \ -1<x_n<1$
Chebyshev map	$x_{n+1}=\cos(g\cdot\arccos x_n),\ \ -1<x_n<1$

Among them, λ is a fractal parameter in the Logistic map, when $\lambda\in(3.5699,4]$, the system is in a chaotic state; g is a mapping order of Chebyshev map, when $g>2$, the system is in a chaotic state.

The main purpose of this paper is to construct a low-overhead secure network coding scheme, to avoid increasing the network transmission parameters, so I chose the modified logistic map to generate chaotic sequence.. The simulation for improved logistic mapping to the initial value sensitivity is given in figure 1.

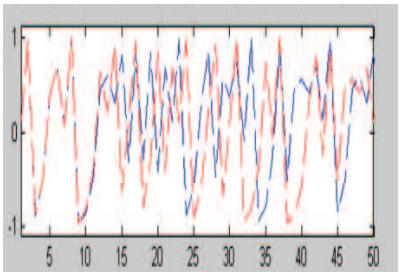

Figure. 1 Improved Logistic mapping chaotic sequence.

In Figure 1, the abscissa represents the sequence length, the vertical coordinates represent the real value of improved logistic, the blue line represents the initial value of the chaotic sequence is 0.1234, the red lines represent the initial value is 0.1235 chaotic sequence, sequence length is 1024.

After digitized, the change rate of improved Logistic map sequence 0/1 is 50.68%. This indicates that the improved Logistic mapping chaotic sequences will generate strong initial sensitivity and difficult to decipher, so it is suitable for secure communications. Chaos mapping is irreversible, That can be seen as a one-way cryptography, and is very suitable for the one-time pad cryptosystem.

5 Low-cost secure network coding

The purpose of the security network coding scheme of this paper is ensure security and minimize overhead of the code transmission signal at the same time. This program uses the chaotic sequence and by the nature of the chaotic sequence shows that the chaotic sequence can be generated by the user control, I.e. can generate a sequence of fixed length. If chaotic mapping and initial value are given the chaotic sequence can be calculated; Moreover, chaotic sequence is pseudo-random, there can be no duplicate values occur within a certain length. Furthermorethe one of the main reasons using chaotic sequence is that it can produce different random signals just uses one redundancy β . Its security equivalent to $n-1$ information of $\bar{m}$ will be encrypted using $n-1$ independent vectors .

5.1 Coding scheme

In the single-source multi-sink multicast network $G=(V,E)$, assume the information source will send $n-1$ character $x_1,x_2,...,x_{n-1}\in F_q$ .Set $Y(.)$ represents a chaotic sequence $y_1,y_2,...,y_n$ and $y_n=1-y_{n-1}^2,(-1<y_n<1)$. The chaotic sequence $Y(.)$ that generated by the improved Logistic mapping and the random number β can generate a message vector, as shown in formula (6):

$$\bar{m}=((x_1+Y(\beta),x_2+Y(x_1,\beta),x_3+Y(x_1,x_2,\beta)...,\\ x_{n-1}+Y(x_1,x_2,...,x_{n-2},\beta)),\beta)^T \quad (6)$$

Comma separator between the variables to various will combination each parts and then use it as a input initial value of the chaotic map. For simplicity, referred to as:

$$\bar{m} = (e_1, e_2, ..., e_{n-2}, \beta)^T \tag{7}$$

And, $e_1, e_2, ..., e_{n-2}$ are well-distributed and $q > max\{t, |\Omega|\}$. Take note of the initial value of chaotic sequence $Y(.)$ in formula (1) are not the same (similar to the one-time pad encryption). So, even if the attacker has been gotten the first $n-1$ data of $\bar{m}$, as long as the random number is secure, then the attacker can not obtain any information about the $x_1, x_2, ..., x_{n-1}$ in the time of the polynomial $\bar{m}$. So this program only use a random signal to complete the purpose of hiding information. The remaining coding process is: linearly assemble $\bar{m}$ (Linear network coding and then transmission to the respective information sink through the network).

5.2 Decoding scheme

Improved Logistic map in this program is open to any party (including the attacker). The degree of difficulty or easy to information sink T recovery $\bar{m}$ depends on the complexity of the basic linear network coding.

As the sink is the recipient of the information, the sink T know the mapping of improved Logistic. So, we just need restore $\bar{m}$ first and then utilize redundant β that has the ability of generating a random signal, all the information can be restored. Thereby the sink successfully decode and get the original source message vector.

6 Safety and feasibility analysis

6.1 The security of one-time pad encryption system

Theorem 1 ”one-time pad” cryptosystem has perfect secrecy.

Proof:

Assume:

u : Plaintext sequence emitted by the information source $u = (u_1, u_2, ..., u_n)$;

d: The ciphertext sequence received by the information sink $d = (d_1, d_2, ..., d_n)$;

P_u: The probability that information source issue expressly sequence u;

r_d: The probability that information sink received ciphertext sequence d ;

q_k: The probability of cryptographic key that information source encrypt the plaintext.

To prove perfect secrecy, is to prove:

$$P_r(u|d) = P_u \tag{8}$$

Owing to there is only one cryptographic key and the key is made up of two independent random variables, so there is:

$$q_k() = 1/2^n \tag{9}$$

as a result

$$r_d = \sum_{u \in P} \frac{P_u}{2^n} \tag{10}$$

and because the select key is independent of the plaintext, so

$$P_r(u \cap d) = \frac{P_u}{2^n} \tag{11}$$

get a final result

$$P_r(u|d) = \frac{P_r(u \cap d)}{r_d} = \frac{\frac{P_u}{2^n}}{1/2^n} \tag{12}$$

so, formula (8) established, the perfect secrecy of the ”one-time pad” encryption system has been proved.

6.2 Feasibility and the general security of the encoding algorithm

Theorem 2 In the single source multi-sink multicast network $G = (V, E)$, if $q > t$, the appropriate linear network coding can be constructed in polynomial time [13], so that the source can send a message to go to more than one information sink in the network.

Proof: Set h is a maximum number of data packets that can be transferred between the information source and sink, that is network capacity. When we construct the random linear network coding, we need to use time $O(|E|)$ to find a path from source to sink. And we need to use the time $O(|T|)$ to allocated encoding vector for each edge, and the spend time is $O(h \cdot |T|)$ that for each sink test the linear independence of the assigned coding vector. So, for all sink test the linear independence of assigned encoding vector can be completed within in the time $O(|T| \cdot h^2)$.

Therefore, if in the coding region to select the appropriate encoding vector to form a network coding global coding nucleus, thereby conformation the linear network coding can be completed within in the time $O(|E| \cdot |T| \cdot h^2)$.

Theorem 3 In the single source multi-sink multicast network $G = (V, E)$, the $\Omega = \{\alpha_1, ..., \alpha_w\}$ is a collection that consist by possible eavesdropped channel [14]. The encoding scheme as described in section 4, if $e_1, ..., e_{n-1}$ obey well-distributed and $q > max\{t, |\Omega|\}$, then the secure communication can be achieved in the multicast network

Proof: We can see from the text [3], the information that the attacker gained from eavesdropping channel constituted a linearly independent vectors team when $e_1, ..., e_{n-1}$ obey well-distributed. Thus eavesdropper can not get any information about β the eavesdropper can not know the chaotic sequence, the encryption system is a ”one-time pad”. By theorem 1 we know the secure network coding scheme is perfect secrecy. From theorem 2, when $q > t$, the linear network coding that meet the conditions can be constructed, the low cost secure network coding present by this paper can achieve the secure communication.

7 Conclusion

In this paper, a low-overhead secure network coding scheme that based on chaotic sequence has been presented. The paper take advantage of chaotic sequence and the characteristics of a ”one-time pad” encryption system, combined with the linear random network coding and finally changed the original source message vector. The secure network coding scheme does not need to know the network topology, and does not require any new features in the internal nodes, and applies to both wired and wireless networks. Through the characteristics analysis in theory and the proof on the security theorem, improved secure network coding scheme can be able to achieve the information security requirements and effectively reduces the overhead.

Acknowledgement

The first author acknowledges the financial support by National 863 Project (2008AAO1Z402): low redundancy data disaster tolerance.
The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

[1] Ahlswede RCai N. Network information flow[J]. IEEE Transactions on Information Theory, **46**, 1204-1216 (2000).
[2] Cai, N. Chan, T. ”Theory of Secure Network Coding” **99**, (2011).
[3] Cai NR. W. Yeung, Secure network coding[C]. IEEE Int. Symp Inf Theory,Lcusanne: IEEE Press, **323**, (2002).
[4] J. Fedman, T. Malkin, C. Stein, et al. On the capacity of secure network coding[C] Proceedings of 42nd Annual Allerton Conference on Communication, Control and Computing, Monticello, IL,USA (2004).
[5] Yeung R W. Zhang Z. Distributed source coding for satellite communications [J]. IEEE Transactions on Information Theory, **45**, 1111-1120 (1999).
[6] Zhang S, Liew S, Lam P. ”Physical layer network coding”. In: ACM MobiCom, LosAngeles, (2006).
[7] Bhadra S, Shakkottai S, Gupta P. Min-cost selfish multicast with network coding[C]. In: 4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, **128**, (2006).
[8] Xu Guangxian, Fu Xiao. ”Low-complexity Secure Network Coding Algorithm Based on Sparse Matrix”. Computer Engineering, **38**, (2012).
[9] Zhukabayeva Tamara, Sembiyev Ordabay, Khu Ven-Tsen, Research of cryptosystems resistance on cascade codes to nonalgebraic decoding attacks, Applied mathematics & Information Sciences, **6**, (2012).
[10] K. Jain, ”Security base on network topology against the wiretapping attack”. IEEE Wireless Communications, 68-71 (2004).
[11] Liu, N. Guo, D. and Parr, G. ”Complexity of chaotic binary sequence and precision of its numerical simulation”. Nonlinear dynamics, **67**, (2012).
[12] C. E. Shannon, ”CommunicationTheoryofSecrecySystems”. BellSystem Technology Journal, **28**, 656-715 (1949).
[13] Y. Shang, Efficient strategies for attack via partial information in scale-free networks, Information Sciences Letters, **1**, 1-5 (2012).
[14] Emina soljanin, ”Network Multicast with Network Coding”, IEEE signal processing magazine, 109-112, (2008)
[15] K. Sivaselvan, C. Vijayalakshmi, Enactment of Stochastic Model Stream Mechanism on Multi class Queueing Network, International Journal of Computing and Network Technology, **1**, 153-161 (2013).

Song Fei is a doctorial student in Chengdu Institute of Computer Application, Chinese Academy of Sciences, China. He has published more than two articles in reputed international journals and International Conferences. His research interests are in the areas of Data Disaster Recovery, Reliability Engineering and Network coding.

Cui Zhe is currently a professor in Chengdu Institute of Computer Application, Chinese Academy of Sciences, China. His research interests lie in Trusted Computing, Embedded Systems and Reliability Engineering.