RESEARCH Open Access

# Efficient weakly secure network coding scheme against node conspiracy attack based on network segmentation

Rong Du<sup>1\*</sup>, Chenglin Zhao<sup>2</sup>, Shenghong Li<sup>1</sup> and Jian Li<sup>1</sup>

#### **Abstract**

In this paper, we consider the problem of building a secure network against node conspiracy attack that based on network segmentation. As we know, network coding has demonstrated its great application prospects in wireless sensor network (WSN) transmission. At the same time, it is facing a variety of security threats, especially conspiracy attack. In existing research, secure coding design strategies are much more than secure topological structure. In this background, a weakly secure scheme is proposed from the perspective of topology and network segmentation. Based on the network segmentation and topology design, the network coding transmission is weakly secure. We conduct a simulation to show that the proposed scheme can efficiently prevent conspiracy attack.

Keywords: Network segmentation; Weakly secure; Network coding; Conspiracy attack

#### 1. Introduction

In 2003, Li [1] demonstrated that with a finite field size, the maximum flow from the single source to sinks can be achieved by linear network coding [2,3]. Based on this theory, network coding technology, network coding has demonstrated its great application prospects in both wired networks and wireless networks. With large-scale application of network coding, it faces a growing number of security issues.

There are many studies on the safety of network coding. Based on secure models, it can be separated into two groups in previous, Shannon secure and weakly secure. The difference of these two classes is that Shannon secure disallows any information leakage and weakly secure disallows any meaningful information leakages. For example, given two data streams x and y, based on weakly secure requirements, the attacks allow to get the combination value of  $x \oplus y$ , but not x or y alone, while in Shannon secure, the attacks disallow learning neither of them. In this paper, we focus on the weakly secure topological structure.

Based on the attack models, there are mainly two attack models, polluting attacks (active attack) [4-8] and

Most existing researches are mostly from a coding perspective with given topologies, which propose different coding algorithms in different network environments, but there is almost no research in secure topology design. Topology design strategies do not need complicated encoding and decoding process which save a lot of memory and computing time. Topology design also has certain failure rate, but when the cooperative eavesdroppers are relatively

<sup>&</sup>lt;sup>1</sup>School of Electronic Information and Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China Full list of author information is available at the end of the article



wiretapping attacks (passive attacks) [9-18]. In this paper, we focus on wiretapping attacks, defined by Cai and Yeung in [9] and proposed a multicast network coding against wiretapping attacks in [10,11]. Feldman et al. [12] proposed a coding scheme in small infinite field at the expense of a small amount of bandwidth. Chan [13] gave the boundaries of the multicast capacity in secure network coding. Bhattad and Narayanan [14] proposed a weakly secure network coding system. On the basis of [14], Silva [15] proposed a general weakly secure network coding system. When the calculation ability of eavesdroppers is limited, Jain [16] designed a weakly secure networking system using one-way function. In [17,18], the authors discussed the security issues in the light of the different conditions and different other safety requirements in wireless sensor network (WSN). Fancsali et al. [19-21] did the corresponding research and gave the respective security coding system.

<sup>\*</sup> Correspondence: durongorc@163.com

fewer and allow a certain error rate, the topology design scheme reflects its advantage.

In view of the fact, the secure topology design is worthy of study. In this paper, we propose a weakly secure network topology algorithm based on topology design and network segmentation. The rest of the paper is organized as follows: We first discuss related work and the security goals we aim to in Section 2. Then, we discuss the problem and present an algorithm for secure network topology design and network segmentation in Section 3. Finally, the results and discussions are addressed in Section 4, and the paper is concluded in Section 5.

#### 2. Problem statement

In this section, we first summarize network coding and then we introduce the system model. Finally, we introduce the threat model and security goals to be used in this paper.

## 2.1 Network coding

Unlike traditional communication networks, network coding is a new technique that allows intermediate nodes to encode multiple input messages together to form multiple output messages. We can use the following example to show how network coding can provide high transmission rate. Figure 1 is a classic example of network coding, each link has unit capacity. Node W encodes the message that is transmitted from W to X through a linear combination  $b1 \oplus b2$ . Thus, the source S-transmitted two bit streams b1 and b2 can multicast to the nodes Y and Z simultaneously; the transmission rate can achieve the multicast rate of 2 bits per unit time.

### 2.2 System model

In this paper, a directed acyclic graph  $G = \langle V, E \rangle$  is considered, where V and E are the node set and the edge set, respectively.  $C_{\min}(G)$  is the minimum cut of G, and the capacity  $C_{\min}(G)$  is the maximal possible information

rate of network G. Each edge has one data stream unit per time slot. The source node s generates and sends out an n symbol message vector  $X = (x_1, x_2, ..., x_n)^T$  in a finite field  $F_q$ . In linear coding systems, the messages on outgoing edges of node  $v_n$  are linear combinations of messages on its incoming edges. It can be understood as that each edge of the network carries an equation of source symbols.

#### 2.3 Threat model and security goals

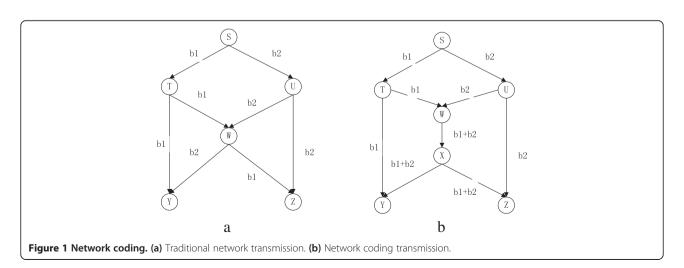
In wiretapping attack, the eavesdroppers are able to gain access to the information transmitted on these nodes, suppose the positions of malicious nodes are known. Also, they can cooperate with each other to decode the packet sent from the source S. Precisely, they can wiretap on a collection of  $M = \langle M_1, M_2, ..., M_K \rangle$ , where M represents a set of malicious nodes; accordingly, they can gain the data stream carried by the incoming links of these malicious nodes, suppose  $E = \langle E_1, E_2, ..., E_K \rangle$  is the incoming link of these malicious nodes. In this paper, we focus on weakly secure. We disallow any meaningful information leakages transmitted from the source node to the sink node.

# 3. Problem description and analysis

#### 3.1 Related definitions

Assume  $C_G(s,t)=k$ , for any intermediate node  $v_i$  in G; if the In  $(v_i)$  is less than the capacity of the graph  $C_G(s,d)$ , then for sufficiently large size q, the generated network code is said to be secure with high probability, because the intermediate node  $v_i$  cannot recover any of the k symbols based on k-1 or fewer linear equations. On the other hand, if  $C_G \leq |\ln(v_i)|$ , the security is said to be topology dependent, and the network is considered secure if and only if  $\operatorname{rank}(\operatorname{in}(v_i)) < C_G$ .

In [22], the sibling work of this paper, we analyzed how the topology design influenced the security of networks, and we proposed a secure strategy against node conspiracy attack by topology design. This method is



suitable for the small network environment; when in a large network, the wiretapping nodes become more, increasing the number of links that needs to be removed. Therefore, we propose a strategy of network coding against wiretapping attack based on network segmentation.

#### 3.2 Secure network segmentation algorithm

Figure 2 is a directed acyclic graph, and each link has a unit capacity. The source node s wants to transmit some information to the destination node t without leaking meaningful information to the eavesdropper. Suppose there are m malicious nodes, each node has n incoming links  $C_G(s,t)=k$ . We need to remove mn-k links to ensure the network security. If we divide the network into two sub-networks, assuming that malicious nodes are uniformly distributed, we just need to remove approximately  $\frac{mn}{s}-k/2$  links to the network security.

In Figure 2, we randomly generate a 50-node network diagram; after path enforcement, we get a directed graph G(V,E). The entire network is divided into two subnetworks  $G_1(V_1,E_1)$  and  $G_2(V_2,E_2)$  by the red line; the dashed line is the link to be removed.  $C(G_1) = C(G_2) = 3$ , any one of the sub-network is safe and leads the whole network security.

How to find the best split routing is the problem that we are mainly faced with, and we get two objective functions.

We know that the min-cut sum of the two sub-networks is no more than the min-cut of the whole network:

$$C_{G_1} + C_{G_2} \le C_G \tag{1}$$

To ensure the throughput of the network, the divided maximum flow is as close as possible to the original maximum flow. We find the split routing and remove the dashed links; the removed links are  $E_{p_1}$ . Then, we pick one sub-network, with the algorithm of [22]; the removed links are  $E_{p_2}$ :

$$E_p = E_{p_1} + E_{p_2}. (2)$$

The two objective functions are max  $(C_{G_1} + C_{G_2})$  and min  $E_p$ . Such an algorithm is referred to as secure network segmentation (SNS) algorithm.

# 3.3 An improved scheme based on the network segmentation and topology design

In [22], we proposed a secure strategy against node conspiracy attack by topology design (ISTD). Here, we use Figure 3 to describe the method generally.

We recommend a conception. As node i,

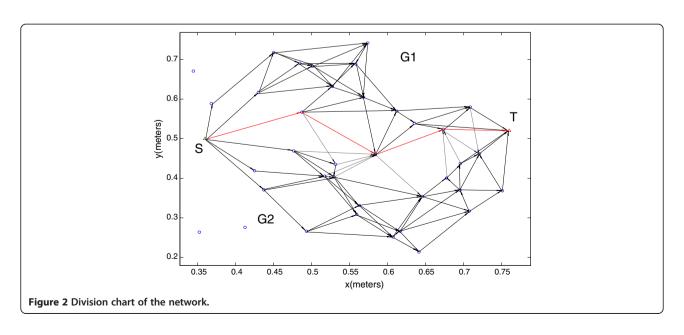
Case 1. in(i) = 1, out(i) = 1, the outgoing message is no change with incoming message (shown in Figure 3a)

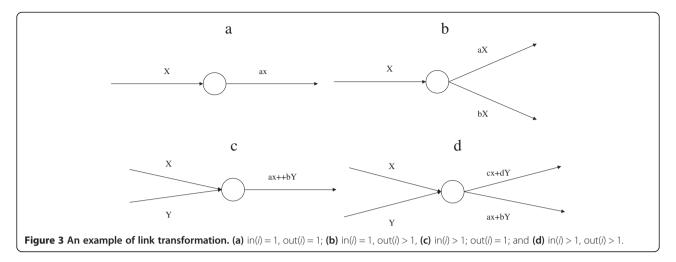
Case 2. in(i) = 1, out(i) > 1, the outgoing messages are linear correlation of incoming message (shown in Figure 3b)

Case 3. in(i) > 1, out(i) = 1, the outgoing message is a combination of the incoming messages (shown in Figure 3c)

Case 4.  $\operatorname{in}(i) > 1$ ,  $\operatorname{out}(i) > 1$ , the outgoing messages are not linear correlation of incoming messages. We can mark these no linear correlation messages as  $G_{i,j} = (x_1, x_2, ..., x_l)^N$  (shown in Figure 3d)

In Figure 4a, suppose nodes 3, 5, and 8 are malicious nodes. Links  $\{2 \rightarrow 3, 9 \rightarrow 3, 4 \rightarrow 5, 10 \rightarrow 5, 1 \rightarrow 8, 7 \rightarrow 8\}$  are polluted links. They carry the different messages  $(x, y)^1$ , z,  $(x, y, z)^1$ , z, y, and x. Each data stream x, y, z appears three times in these six polluted links. To satisfy the security requirements, we need to remove three polluted links at least. Using the STD algorithm, removing  $\{9 \rightarrow 3, 4 \rightarrow 5, 10 \rightarrow 10, 10 \rightarrow 10,$ 





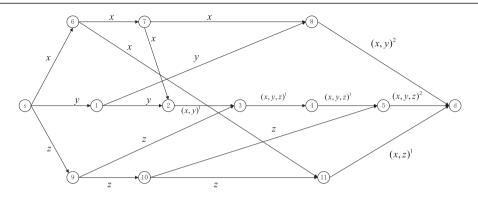
 $10 \rightarrow 5$ } is the best result. Only one additional link  $\{3 \rightarrow 4\}$  needs to be removed. In Figure 4b, we do not remove any polluted links. Instead, we remove  $\{6 \rightarrow 7\}$  and re-create the network topology; we will find that the topology is already weakly secure. Eavesdroppers cannot get data stream x, which means the data stream x of  $\{2 \rightarrow 3, 4 \rightarrow 5, 7 \rightarrow 8\}$  comes from  $\{6 \rightarrow 7\}$ ; the sink node d receives data stream x from another way. It tells us that if we can get every data source of polluted links, we may find an advanced scheme to solve the security problem.

In Figure 4c, we mark the original network topology in another way. Node 6 gets the data stream x from the source node s and sends it to node 7 and node 11. We define the source node is a parent node of node 6, and node 7 and node 11 are child nodes of node 6. We mark the data stream x carried by  $\{6 \rightarrow 7\}$  and  $\{6 \rightarrow 11\}$  as  $x^1$  and  $x^2$ , node 7 gets the data stream from node 6 and points to nodes 2 and 8, then the data stream carried by  $\{7 \rightarrow 8\}$  and  $\{7 \rightarrow 2\}$  can be marked as  $x^{1.1}$  and  $x^{1.2}$ ; here,  $x^{a,b,c,d\cdots}$  represents different sources of data x. For node 5, it gets the data stream  $(x^{1,2}, y^2, z^1)^1$  and  $z^{2,1}$ .

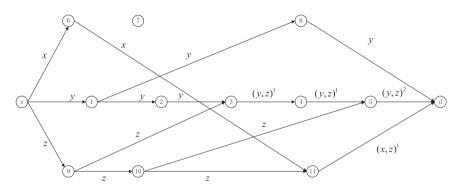
From the above transmission matrix G, polluted links  $\{2 \rightarrow 3, 4 \rightarrow 5\}$  carry the data stream  $x^{1,2}$ ,  $\{7 \rightarrow 8\}$  carries the data stream  $x^1$ , which both come from the data stream  $x^1$ , and the sink node d receives both  $x^1$  and  $x^2$ . Once link  $\{6 \rightarrow 7\}$  is removed, the eavesdroppers cannot get the data stream x, while the sink node d can get complete information from the source node s.

It needs to be mentioned that the same data streams need to be combined. For example, the data  $x^{1,2,3,4}$  and  $x^{1,2,3,5}$  can be combined into  $x^{1,2,3}$ , and the data  $x^{1,2,3,4}$  and  $x^{1,1,2,3}$  are combined into  $x^1$ ; the combination is to find the maximum number of occurrences of the original data. In Figure 4,  $\{5 \rightarrow d\}$  carries the data  $(x^{1,2}, y^2, (z^1, z^{2,1}))^1$  where  $(z^1, z^{2,1})$  means the data stream z comes from both  $\{9 \rightarrow 3\}$  and  $\{9 \rightarrow 10\}$ , and it can be grouped into z. The biggest features of the ISTD algorithm are finding the sources of each polluted edges and removing these relatively few source edges to make the network secure.

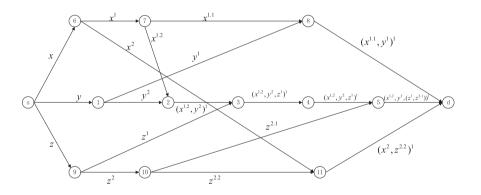
Combining the two algorithms of SNS and ISTD, we get an improved scheme based on topology design and



# a. Network topology



# b. Secure topology when node 3,5 and 8 are malicious nodes



c. Re-creation mark of network topology

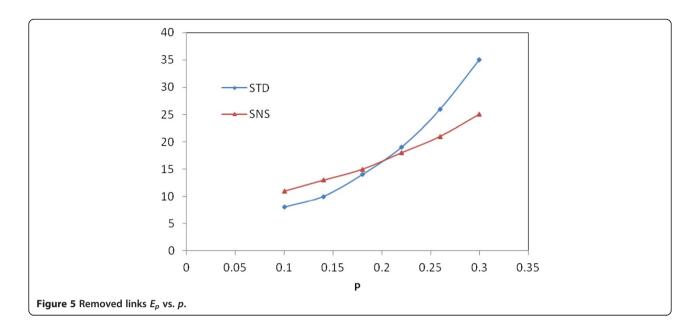
**Figure 4** Advanced scheme of transmission topology design with transmission rate of 3 in WSN system. (a) Network topology. (b) Secure topology when nodes 3, 5, and 8 are malicious nodes. (c) Re-creation mark of network topology.

network segmentation (ISNS). We summarize this method into the following steps:

Step 1: Given a directed acyclic graph  $G = \langle V, E \rangle$ , after path enforcement, suppose the positions of malicious nodes are known, the minimum cut of G

is  $C_G$ .  $V^{'}=\left(v_1^{'},v_2^{'},...,v_m^{'}\right)$  is the set of malicious nodes.  $E^{'}=\left(e_1^{'},e_2^{'},...,e_n^{'}\right)$  is the set of the incoming edges of  $V^{'}$ . We call it polluted edges.

Step 2: Mark all links of the network topology, fill the transmission matrix G with the data stream X, and combine the source of information flow.



Step 3: Remove the same linear correlation of incoming message. We define E'' is the set of the remaining edges which we called it polluted edges.

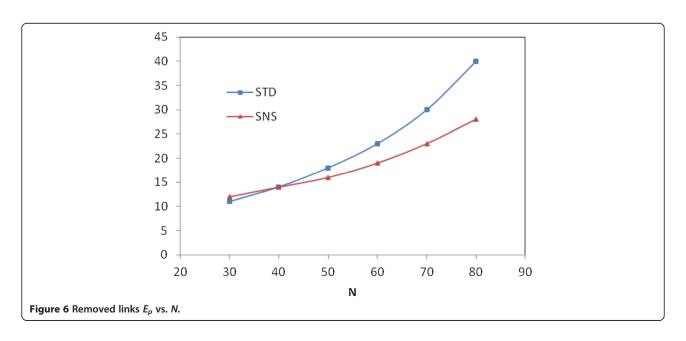
Step 4: Get a forward routing which contains the polluted edges. The total number of the forward routing cannot be more than the total polluted edges. The forward routing is the split routing. The entire network is divided into two sub-networks  $G_1(V_1, E_1)$  and  $G_2(V_2, E_2)$ . Suppose the split routing is  $E_p$ . Step 5: We calculate the number of polluted links of the two sub-networks  $E_1^r$  and  $E_2^r$ . Suppose the less polluted links is  $E_1^r$ . For safety requirements, all the links that need to be removed is  $E_1^r + E_p$ .

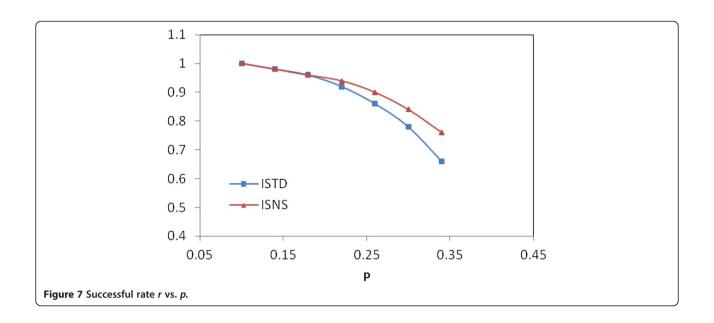
Step 6: Get the solution of the topology design and network segmentation. The least links to be removed is the optimal solution, if there is no such a topology that satisfies the condition. Then, the message leakage is unavoidable.

#### 4. Simulation and discussion

# 4.1 The performance and discussion of SNS

In this section, simulations are conducted based on ns-2 simulator and MATLAB to evaluate the effectiveness of the proposed algorithm. The network is defined by these parameters, the number of nodes, N, (the number of edges,  $E_{\rm all}$ ), the probability of malicious nodes in





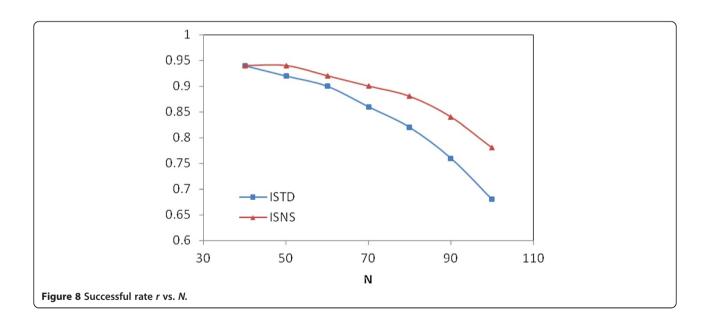


Table 1 Performance comparison of coding design strategies and topology design strategies

	Encoding and decoding cost	Security	Success rate	
Coding design strategies	$o(h^2) - o(h^3)$	Shannon or weakly		≈1
Topology design strategies (ISNS)	o(h)	Weakly	Wiretappers' rate less than 0.1	Wiretappers' rate less than 0.2, more than 0.1
			≈1	>0.9

intermediate nodes, p, and the removed links  $E_p$ . The algorithm in [22] is the STD algorithm, and the SNS algorithm is the basic scheme in this paper. For each combination of parameters, we generate 50 instances.

In Figure 5, we set N = 50 and vary p in the range of (0.1 - 0.3) to calculate  $E_p$ . In Figure 6, we set p = 0.2 and vary N in the range of (30-70).

We can see from Figures 5 and 6 that, with the increase of p and N, the SNS algorithm removes few links than the STD algorithm. It improves the link utilization, when faced with a large number of wiretapping nodes, and performs particularly well. Compared to the STD algorithm, the SNS algorithm in this paper has been greatly improved, especially suitable for larger networks.

## 4.2 Improved SNS algorithm (ISNS)

STD is the proposed scheme in [22]. We proposed an advanced scheme ISTD, relative to STD. ISTD was greatly increased efficiently. In this paper, the ISTD algorithm will be integrated into SNS. ISNS is the improved scheme based on the network segmentation and topology design. Compared with ISTD, the efficiency of ISNS algorithm has been greatly improved. It removed less polluted links; the small change of the transmission topology improves the successful rate.

The network is defined by four parameters, the number of nodes, N, the probability of malicious nodes in intermediate nodes, p, the largest degree of each node D (the largest amount of incoming links), and the successful rate of transmission r. For each combination of parameters, we generate 50 instances.

In Figure 7, we set N = 50, D = 7 and vary p in the range of (0.1 - 0.35) to calculate the r. In Figure 8, we set p = 0.2, D = 7 and vary N in the range of (50-100) to calculate the r. With the increase of p and N, the ISNS algorithm performs better than ISTD. From the simulation results, we can see that ISNS can cope with larger structures and more malicious node network conspiracy attack.

# 4.3 Performance comparison between coding design strategies and topology design strategies

Compared with secure coding design strategies, the topology design scheme has its own advantages and disadvantages. From the ISTD algorithm and ISNS algorithm, we can see that topology design strategies do not need complicated encoding and decoding processes; they use linear network coding which save a lot of memory and computing time. The proposed algorithm also has certain failure rate, but when the cooperative eavesdroppers are relatively fewer (not more than 0.2) and the network allows a certain error rate, the topology design scheme reflects its advantage. In Table 1, we give the comparison of coding design strategies and topology design strategies.

It is especially suitable for the network with low percentage of malicious nodes and allows certain error rate.

#### 5. Conclusion

In this paper, we have investigated the topology design and network segmentation issue for weakly secure against node conspiracy attack. We analyzed how the network segmentation and topology design influenced the security of networks. We proposed a secure strategy against node conspiracy attack by network segmentation and topology design. We compared the ISTD and ISNS strategies. Simulations showed that the proposed routing algorithm ISNS achieved good performance. It can cope with larger structures and more malicious node network than ISTD. As a future research, we will study the secure topology design strategy under a large number of malicious nodes and a larger structure.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

The work presented here was carried out in collaboration among all authors (RD, CZ, SL, and JL). RD, CZ, and SL defined the research theme. RD and SL designed the methods and experiments, carried out the laboratory experiments, analyzed the data, interpreted the results, and wrote the paper. JL co-designed the experiments and discussed the analyses, interpretation, and presentation of data. All authors read and approved the final manuscript.

#### Acknowledgements

This work is funded by the National Science Foundation of China (61271316, 61071152, 61271180), 973 Program (2010CB731403, 2010CB731406, 2013CB329605) of China, Chinese National Twelfth Five-Year' Plan for Science & Technology Support (2012BAH38 B04), Key Laboratory for Shanghai Integrated Information Security Management Technology Research, and Chinese National Engineering Laboratory for Information Content Analysis Technology.

## Author details

<sup>1</sup>School of Electronic Information and Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China. <sup>2</sup>School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, People's Republic of China.

Received: 12 November 2013 Accepted: 30 December 2013 Published: 13 January 2014

#### References

- S Li, R Yeung, N Cai, Linear network coding. IEEE Trans Inf Theory 49(2), 371381 (2003)
- 2. R Ahlswede, N Cai, S-YR Li, RW Yeung, Network information flow. IEEE Trans Inf Theory 46(4), 1204–1216 (2000)
- R Koetter, M Medard, An algebraic approach to network coding. IEEE/ACM Transactions on Networking 11(5), 782–795 (2003)
- Z Yu, Y Wei, B Ramkumar, Y Guan, An efficient signature-based scheme for securing network coding against pollution attacks, in *Proceedings of* the 27th IEEE Conference on Computer Communication, INFOCOM 2008, (Phoenix, 13–18 Apr 2008), pp. 1409–1417
- T Ho, B Leong, R Koetter, M Medard, M Effros, D Karger, Byzantine modification detection in multicast networks using randomized network coding, in *Proceedings of IEEE International Symposium on Information Theory* (ISIT), (Chicago, 27 June–2 July 2004), p. 144
- S Jaggi, M Langberg, S Katti, T Ho, D Katabi, M Medard, Resilient network coding in the presence of Byzantine adversaries, in *Proceedings of the* 26th IEEE Conference on Computer Communications, INFOCOM 2007, (Barcelona, 6–12 May 2007), pp. 616–624

- M Krohn, M Freedman, D Mazieres, On-the-fly verification of rateless erasure codes for efficient content distribution, in *Proceedings of IEEE Symposium on* Security and Privacy, (Berkeley, 9–12 May 2004), pp. 226–240
- C Gkantsidis, PR Rodriguez, Cooperative security for network coding file distribution, in *Proceedings of the 25th IEEE International Conference on Computer Communications, INFOCOM 2006*, (Barcelona, 23–29 Apr 2006), pp. 1–13
- N Cai, R Yeung, Secure network coding, in Proceedings of IEEE International Symposium on Information Theory (ISIT), (Lausanne, 30 June–5 July 2002), p. 323
- N Cai, RW Yeung, A security condition for multi-source linear network coding, in *IEEE International Symposium on Information Theory, Nice*, (24–29 June 2007), pp. 561–565
- Z Zhang, RW Yeung. A general security condition for multi-source linear network coding, 2009, in *IEEE International Symposium on Information Theory* (ISIT). (IEEE, 2009) pp. 1155-1158
- J Feldman, T Malkin, C Stein, On the capacity of secure network coding, in Proceedings of the 42nd Annual Allerton Conference on Communication, Control, and Computing, (Monticello, 29 Sept-1 Oct 2004)
- 13. T Chan, Capacity bounds for secure network coding, in *IEEE Communication Theory Workshop*, (Christchurch, 30 Jan–1 Feb 2008), pp. 95–100
- K Bhattad, KR Narayanan, Weakly secure network coding, in First Workshop on Network Coding, Theory and Applications, (Riva del Garda, 7 Apr 2005)
- D Silva, FR Kschischang, Universal weakly secure network coding, in Information Theory Workshop on Networking and Information Theory, (Volos, 10–12 June 2009), pp. 281–285
- K Jain, Security based on network topology against the wiretapping attack. IEEE Wirel Comm 1(1), 68–71 (2004)
- D Jing, R Curtmola, R Sethi, Toward secure network coding in wireless networks: threats and challenges, in 4th Workshop on Secure Network Protocols, (Orlando, 19–22 Oct 2008), pp. 33–38
- A Mills, B Smith, T Clancy, On secure communication over wireless erasure networks, in *IEEE International Symposium on Information Theory*, (Toronto, 6–11 July 2008), pp. 161–165
- S Fancsali, LP Ligeti, Some applications of finite geometry for secure network coding. J. Math. Crypt. 2(3), 1862–2984 (2008)
- MM Hassanzadeh, M Ravanbakhsh, O Ytrehus, Two layer secure network coding-(2-LSNC), in *IEEE International Symposium on Telecommunications*, (Tehran, 27–28 Aug 2008), pp. 7–12
- K Harada, H Yamamoto, Strongly secure linear network coding. IEICE Transactions on Fundamentals E91-A(10), 2720–2728 (2008)
- R Du, C Zhao, F Zhao, S Li, Strategies of network coding against nodes conspiracy attack. Security and Communication Networks, (2013). doi:10.1002/Sec.753

#### doi:10.1186/1687-1499-2014-5

Cite this article as: Du et al.: Efficient weakly secure network coding scheme against node conspiracy attack based on network segmentation. EURASIP Journal on Wireless Communications and Networking 2014 2014:5.

# Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Immediate publication on acceptance
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com