2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications

A Survey of Security Network Coding toward Various Attacks

Shixiong Yao, Jing Chen, Ruiying Du, Lan Deng, Chiheng Wang
Department of Computer Science
Wuhan University
Wuhan, China
chenjing@whu.edu.cn

***Abstract*—As one of the emerging technologies with most potential for developing, Network Coding (NC) has gained significant momentum. Due to encode-and-forward model, NC has natural privacy in communication, but it also induces that attackers become more imperceptible and impact on NC caused by them becomes more far-reaching. This requires that the security schemes should consider the characteristics of different attacks in NC. In this paper, we provide a survey of secure network coding toward various attacks. First of all, we summarize four types of representative attacks in NC system including entropy attack, Byzantine attack, pollution attack and eavesdropping attack, and compare the differences of these attacks between in traditional store-and-forward mode and network coding mode. Secondly, we give a comprehensive investigation of numerous defense approaches and mechanisms classified by these attacks. Finally, for stimulating stream of thoughts about secure network coding schemes, several open issues are proposed and discussed.**

***Index Terms*—Network coding, four forms of attacks, defense schemes**

I. INTRODUCTION

It is well-known that the information stream cannot be overlaid in the traditional store-and-forward routing mechanism. However, In 2000, Rudolf Ahlswede [1] firstly proposed network coding (NC) and theoretically verified that the maximum throughput could be achieved in the NC system if the intermediate nodes encode the received packets in suitable method instead of only store and forward them. After that, they presented the linear network coding (LNC) scheme. In this scheme, the maximum throughput could be achieved practically. In order to solve the problem of NC in distributed network system, Tracey Ho [2] proposed random linear network coding (RLNC); Koetter Ralf [3] put forward algebraic framework of NC, and consequently the problems of the coefficient could be solved in algebraic method. Compared with the traditional routing mechanism, network coding has some advantages, such as, increasing network throughput, balancing network load, enhancing network robustness and so on. Thus NC has been applied in various fields, such as, P2P network [4], Ad hoc Network [5], Wireless Mesh Network [6], Wireless Senor Network [7], DTNs [8], Content Distribution Network [9] and so on.

Numerous research institutes have paid attention to this technology and proposed some state-of-art schemes [10-14]. As study continues, the results show that the advantages of NC are only reflected in the networks consisting of faithful nodes. However, not all the nodes are reliable in the real scenarios, where protection against malicious attacks remains to be a major challenge under various threats,

- In network coding mode, the impact of attacks becomes severer and wider than that in traditional mode. In intra-flow NC [15], if some packets are corrupted, they will affect the all the nodes in the path. And that is more serious in inter-flow NC [16], it not only influence one path, but also impact multiple paths from which the variants of the corrupted packets pass through.
- Malicious nodes become more imperceptible since they may launch various attacks through some innocent nodes which are far from the attack destination.

In this paper, we provide an overview of secure network coding toward various attacks. The main contributions are as follows,

- In the perspective of threats in network coding, we provide a classification principle with four types of representative attacks and compare the difference of these attacks in different transmitting mode.
- According to these types of attacks in network coding, we introduce various defense schemes with analyzing and comparing and the characteristics of them. From the results which are presented in TABLE III, a suitable scheme can be chosen to resist the corresponding attack.
- We propose some open problems in secure network coding in section IV, these issues may offer a guidance to follow-up researches.

The remainder of the paper is organized as follows. In Section 2, we introduce the primary form of attacks and the effect to the NC system. Four attacks and the defense schemes are analyzed respectively in detail in Section 3. Section 4 discusses several related research problems that remain open. In the last section, conclusion is provided.

II. THE PRIMARY FORMS OF ATTACKS

Unlike the traditional approaches which forward each input message directly, NC allows each intermediate node to combine multiple input messages into one or more encoded ones before forwarding them. Therefore, each output message sent to the downlink can be linear combination of the input messages received from the uplinks. Generally speaking, network coding system consists of the intermediate nodes which are responsible for transmitting, encoding, and the destination nodes which are responsible for decoding. But the open wireless channel would inevitably face various potential security threats, which may seriously affect the effectiveness of network coding. According to different classification principles, we can divide secure threats into different types, such as active attack and proactive attack, external attack and internal attack. In this paper, we summarize current works of secure network coding and propose a classification principle, in which four primary attack forms are classified. The definitions of these attacks are as follows.

978-1-4799-6513-7/14 $31.00 © 2014 IEEE
DOI 10.1109/TrustCom.2014.35

TABLE I. THE COMPARISON OF FOUR ATTACKS BETWEEN TRANDITIONAL MODE AND NETWORK CODING MODE

Types of attacks	Traditional Mode	Network Coding Mode
Entropy attack	Replay attacks. The attackers replay packets which were received by the sink node. The goal of replay attacks is to destroy the correctness of the authentication. Defense schemes are often based on timestamp, sequence number or the challenge- response model.	Because of mixture, traditional schemes such as timestamp and sequence are not suitable for NC. The defense schemes are based on cryptograph algorithm, watchdogs or algebra mechanism.
Byzantine attack	A compromised node which is an authenticated user can access all the information and resource of the whole system and has full control of a number of authenticated nodes. The defense schemes are based on voting mechanism or trust model.	The Byzantine attacks in traditional mode and NC mode are similar. But the influence is more serious in NC mode. The defense schemes are based on cryptology or watchdogs.
Pollution attack	An external node injects junk information into the network. Cryptographic algorithms such as digital signature and MAC can detect the pollution packets rapidly.	If this type of attacks is out of control, pollution could be propagated downstream rapidly. Thus its impact is more serious on NC. Homomorphism is effective and algebra mechanism is feasible in this mode.
Eavesdropping attack	An internal or external node could eavesdrop on communica-tion by wiretapping one or more links in wired network or using the high frequency antenna in wireless network. Encryption is an efficient way to defense against this type of attacks.	In NC system, an eavesdropper should wiretap more than one links to get meaningful information. Hence, eavesdroppers need stronger ability. The defense schemes are often based on permutation encryption or topology optimization.

- **Entropy Attack** [17]. This is a special replay attack. Malicious nodes create non-innovation coded packets which are linearly dependent with the coded packets stored at a downstream node. These packets waste resources since they induce useless information to increase the workload of receivers in the process of decoding original packets.
- **Byzantine Attack** [18]. Byzantine nodes are traitor nodes operating with a hidden intent to disable or impair the network. As trusted nodes, they locate along the multi-hop paths between source and destination nodes and have complete access to the information and resources of the network. Thus this attack is imperceptible.
- **Pollution Attack**. This attack is usually started by unauthorized nodes which inject polluted packets into the information flow. Due to the openness of wireless network, malicious nodes can launch it from arbitrary point in network.
- **Eavesdropping Attack** [19]. An eavesdropping attacker can either wiretap one or more links in the wired network, or use the high frequency antenna to acquire information within certain range of the intermediate node in the wireless network.

TABLE II. CLASSIFICATION ABOUT THE ATTACKS

	Active attacks	Passive attacks	External attacks	Internal attacks
Entropy attack	√			√
Byzantine attack	√			√
Pollution attack	√		√	
Eavesdropping attack		√	√	

Based on the previous definitions, the relationships between various classification principles are presented by TABLE II.

In spite of the characteristics which they have in common, the differences between them are obvious. Firstly, attackers no matter in entropy attack, Byzantine attack and pollution attack can inject packets into the network. However, the cryptographic schemes against pollution and Byzantine attacks may be inapplicable for entropy attacks, since packets produced by the entropy attackers look like legitimate. Secondly, the Byzantine attackers can implement pollution attack, but they are trusted internal nodes, the pollution attackers are often unauthorized external nodes.

These types of attacks also appear in traditional store-and-forward mode, but most of defense schemes are not compatible with network coding system. Compared with traditional mode, the impacts of these attacks on network coding system is more serious. For example, if a single corrupted block will pollute the network and prevent the receivers from decoding. Such pollutions can rapidly propagate in the network, and they result in performance sliding down significantly since the corrupted blocks will influence more paths. Thus we give a comparison of these attacks between in traditional mode and network coding mode as shown in TABLE I.

III. DEFENSE SCHEMES

In this section, we introduce the defense schemes according to the four attacks respectively in detail.

A Defense Schemes of Entropy Attack

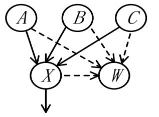
Entropy attack has received less attention than the other three attacks, but its impact on the whole system should not be ignored. Currently, there are three types of solutions mentioned in this paper to resist these attacks: (1) Watchdogs-based scheme; (2) Cryptographic-based schemes; (3) Algebraic-based scheme.

(1) Watchdogs-based scheme

Andrew Newell and Reza Curtmola [20] identified two variants of entropy attacks (local and global). They also proposed and evaluated several defenses which vary in detection capability and overhead. Based on the capabilities of the attackers, the authors subdivided entropy attacks into two categories: A local entropy attacker produces non-innovative coded packets for local neighboring nodes; A global entropy attacker produces coded packets that are seemingly innovative to local neighboring nodes but are non-innovative to one distant downstream node at least.

In order to deal with local entropy attacks, the author proposed non-innovative link adjustment (**NLA**) as a defense. The main idea of this defense is to determine which nodes are performing attack and remove them from the system. The simulation results show that the performance of the system has greatly improved with **NLA** under local entropy attacks.

For the global entropy attacks, the author put forward two ways: Upstream Buffer Propagation (**UBP**) and Buffer Monitoring (**BM**). The idea of **UBP** defense is easy to understand like that if a node has received a non-innovation packets, it means that a global entropy attacker may reside along any upstream path. In order to prevent this attack, the node which has received non-innovative coded packets propagates buffer information upstream until it reaches the attacker. When the attacker receives this information, it has a choice to continue performing the entropy attack and would be detected, or to start sending innovative coded packets. Both way, the entropy attack is mitigated. In the Buffer Monitoring (**BM**) defense, nodes monitor incoming and outgoing traffic of untrusted, neighbor nodes to immediately detect any coded packets created in a non-random fashion. For a watchdog to determine whether an output packet broadcasted by a watched node is random linear combinations of all its input packets, it must know about all input packets received by that forwarder for the generation. In Figure 1, in order to monitor *X*, the watchdog *W* should receive all the packets from *A*, *B*, *C*, *X*.



X is the node being monitored, *W* is the watchdog node, and *A*,*B*,*C* are upstream neighbors of ***X***. Solid lines indicate wireless links of the topology that are used for routing data. Dashed lines indicate wireless links of the topology to send data to the watchdog node.

Figure 1. BM statement

(2) Cryptographic scheme

Chi Cheng [21] presented an efficient symmetric key based authentication scheme for P2P live streaming system with network coding, to provide in-network detection against pollution attacks and entropy attacks simultaneously. They proposed a homomorphic message authentication code, called as **PMAC**, which has small key size and low computation overhead. Then, the PMAC and the delayed key disclosure technique are to make sure that the peers could not only detect the corrupted blocks, but also upload blocks in accordance with RLNC.

(3) Scheme based on algebraic mechanism

The cryptograph schemes are often computationally expensive or need extra secure channel. Yixin Jiang [22] proposed an efficient packet filtering scheme which was based on a novel self-adaptive probabilistic subset linear dependency detection (**S-PSLD**) algorithm against entropy attacks in NC. This scheme could rapidly filter out the resultant packets from entropy attacks since it verified the received packets probabilistically instead of exactly. What's more, in order to minimize the packet detection cost at forwarder while keeping the false positive rate at an expected low level, the self-adaptive algorithm was introduced such that each forwarder can dynamically tune the system security parameters according to the available bandwidth or the number of the received packets in buffer.

B Defense Schemes for Byzantine Attack

Trusted Byzantine nodes which have passed authentication are difficult to be detected, as these nodes have a hidden malicious component and have complete access to the network resources. In addition, the Byzantine nodes have abilities to eavesdrop and jam communication and can inject pollution packets into the network. Two types of schemes aiming at resisting them are analyzed in this part: (1) Cryptographic algorithm. (2) Watchdogs-based schemes.

(1) Cryptographic algorithm

Based on signature, homomorphic MAC, CBC schemes and so on are proposed to withstand Byzantine attacks.

Sidharth Jaggi *et al.* [23] designed distributed polynomial -time rate-optimal network codes that combat Byzantine adversaries. They presented three algorithms corresponding to adversaries with different strengths. The intuition underlying all of the algorithms is that the adversarial nodes can be regarded as a second source. The information received at the destination is a linear transform of the packets from the source and the adversary. Given enough linear coded packets, the destination can decode both sources. If the source's information satisfies certain constraints (such as particular checksum) but the attacker's data doesn't, the destination can distill out the source's information from the received mixture.

MinJi Kim *et al.* [24] proposed a homomorphic signature scheme that allowed packet-level Byzantine detection in P2P networks. This scheme is efficient and does not require a secure channel. Anh Le *et al.* [25] proposed a novel approach that could identify the precise location of all Byzantine attackers in systems with intra-session network coding. A key component of this approach is a homomorphic MAC scheme for expanding subspaces (SpaceMac) that allows to eliminate any uncertainty in identifying attackers via subspace properties.

Feng Tao *et al.* [26] presented a secure random network coding (RNC) model based on CBC (Cipher Block Chaining) for resisting Byzantine attacks. According to categories, the author defined different situations of Byzantine attacks appeared in the NC system and proposed the SRBA-RNC (Security-Resilient Byzantine Attack-Random Network Coding) model scheme which was based on Cipher Block Chaining.

(2) Watchdogs-based schemes

Obviously, watchdogs could monitor the illegal behavior or packets, so the schemes, based on watchdogs, could be used to defense Byzantine attacks.

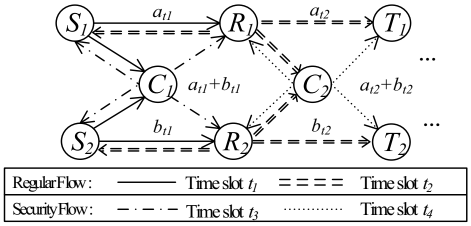


Figure 2. Two X common topologies

Vaibhav Pandit *et al.* [18] identified an inherent security benefit of Analog Network Coding which could be used to design a novel **Dual Non-adjacent Watchdog** scheme. This scheme assumes that the source and destination are trusted and the any other intermediate forwarder could potentially be a Byzantine attacker. Only neighboring nodes within each other's communication range can collude. The two X topology is considered in Figure 2. The flow of data from a source to a destination is called the Regular Flow. There are two regular flows: S_1 to T_1, S_2 to T_2. C_1 and C_2 are network coding nodes.

Every transmission is verified by a pair of watchdogs. The two watchdogs are not adjacent to each other, so they cannot collude. For example, in the regular flow of S_1 to T_1, S_1 and R_2 are the two watchdogs. S_1 is a part of the Regular Flow and R_2 is not adjacent to any nodes in this regular flow.

S_1 act as the first Watchdog by buffering a_{t1} after its transmission at t_1 and then comparing it to the overheard signal a_{t2} at t_2. The nonadjacent R_2 as the second Watchdog, at t_3, it receives the encoded signal (a_{t1} + b_{t1}) from NC node C_1. Now, R_2 possesses the signal (a_{t1} + b_{t1}) and the original signal b_{t1} at t_1 from S_2 to extract a_{t1}. At t_4, R_2 receives (a_{t2} + b_{t2}) from C_2. Similarly, R_2 can now use its prior knowledge of b_{t2} to extract a_{t2} from (a_{t2} + b_{t2}). Then R_2 can verify if a_{t1} = a_{t2}.

In this way, nodes S_1 and R_2 act as nonadjacent dual watchdogs for the transmission of a_{t2}. Similarly, for the transmission b_{t2}, S_2 and R_1 can act as non-adjacent dual watchdogs. Just one relay node in this example, this scheme can be applied to a longer chain of relay nodes.

C Defense Schemes for Pollution Attack

Pollution attacks are fatal to the NC system if they are out of control. The main methods to defense these attacks are based on cryptography and algebraic.

(1) Cryptography schemes

Cryptography has been well-recognized as one of the most effective approaches to address these security issues, and signatures and MACs are utilized most widely.

Following, the schemes based on Signature would be introduced. Jiang Yixin *et al.* [27] proposed signature scheme based on a dynamic-identity for NC. The scheme can rapidly detect/drop the packets that are generated from pollution attacks, and resist random forgery attack efficiently. By employing packet-based and generation-based batch verification approaches, a forwarder can verify multiple received packets synchronously with dramatically reduced total verification cost.

Traditional signature schemes based on hash function are unsuitable for NC, since the original source signatures would be destroyed by the subsequent encoding process, which is performed at each encoding forwarder. So homomorphic signature schemes have been proposed for NC. Yu Zhen [28] proposed an efficient signature-based scheme against pollution attacks on LNC systems. This scheme utilizes a homomorphic signature function, which allows forwarders to generate the signatures for their output signals from those of input signals. Each node appends the signatures to its output signals, so its downstream nodes can verify the received signals effectively and discard the polluted or forged ones.

Next, the schemes are based on MAC.

Message authentication codes (MACs) are used to guarantee the data integrity and origin, they could be utilized for defense pollution attacks. However, just like the traditional signature schemes, classical MAC is not compatible with the linear combination of NC. The homomorphic MAC scheme [29] for NC has been proposed for fast packets verification.

Li Qiming [30] applied Krohn's scheme [31] which was a homomorphic Hash scheme to detect pollution attacks. In this scheme, the content X is divided into n blocks $x_1, \cdots, x_n$, and then each block x_i is further divided into m subblocks $x_{i,1}, \cdots, x_{i,m}$. Each $x_{i,j}$ is an element in the multiplicative group $\mathbb{Z}_p^*$ for a large prime p. A hash function $\mathcal{H}$ is then applied on $x_1, \cdots, x_n$ to obtain the hash values $h_1, \cdots, h_n$. Particularly, m generators $g_1, \cdots, g_m \in \mathbb{Z}_p^*$ are be used in the hash function, and the h_i corresponding to $x_i$is computed as $h_i = \prod_{j=1}^{m} g_j^{x_{i,j}}\, mod\, p$. Obviously, the hash function $\mathcal{H}$ is homomorphic on account of that for any two blocks x_i and x_j , it holds that $\mathcal{H}(x_i)\mathcal{H}(x_j) = \mathcal{H}(x_i + x_j)$. These hash values are distributed to all the nodes reliably in advance. According to the homomorphic property of $\mathcal{H}$, a coded block x which is a linear combination of the original n blocks with coefficients $C = (c_1, \ldots, c_n)$ can be verified the integrity by using C and the hash values $h_1, \cdots, h_n$ at an intermediate node. In particular, the node checks if $\mathcal{H}(\mathrm{x}) = \prod_{i=1}^{n} {h_i}^{c_i}\, mod\, p$ is true. The whole scheme is shown in Figure 3.

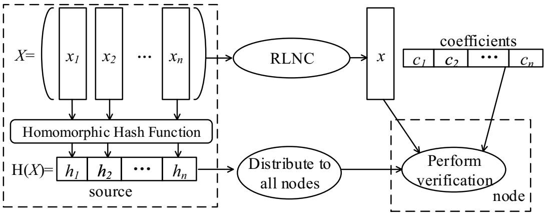


Figure 3. Homomorphic Hash and Verification

Anya Apavatjrut and Wassim Znaidi *et al.* [32] investigated the MAC acted as the core mechanism to defense pollution attacks. Nowadays, three families of MAC algorithms can be found: (1) HMAC [33], and MDx-MAC [34] rely on cryptographic hash functions; (2) CBC-MAC designs such as [35] depend on the implementation of block ciphers. (3) Finally, universal hash functions (UHFs) have been proposed. These three MAC designs can be utilized to thwart pollution attacks. Four modes of operation to protect the messages integrity are presented: Authenticate-XOR-MAC -Forward (AXMF), Authenticate-XOR-Forward (AXF), XOR -Authenticate-Forward (XAF) and XOR-Forward (XF). The last three modes, i.e. AXF, XAF and XF, are only available for linear/homomorphic MACs.

Previous approaches based on homomorphic MAC to resist the attacks either have an expensive computation overhead or don't locate the attackers. Yichao Xu *et al.* [36] proposed a scheme which integrated an efficient Homomorphic MAC with a lightweight non-repudiation transmission protocol. This scheme can not only detect the corrupted packets but also the location of the malicious nodes. The relationship between security and parameters of the protocol is provided, which could help to choose the better parameters satisfied the system requirements.

In energy-limited networks such as wireless network, the goal of further diminishing the computation cost is considered. After proposed an efficient signature-based scheme, Yu Zhen *et al.* [37] presented a multiple MACs scheme to defense pollution attacks. Multiple MACs for each message are generated by the source using its secret keys, where each MAC can authenticate only a part of the message and the parts authenticated by different MACs are overlapped. Every encoded message is attached with its MACs. Therefore, multiple subsequent forwarders can verify different parts of the encoded message collaboratively. By carefully controlling the degree of overlapping parts, this scheme can filter polluted messages in a few hops with high probability. Moreover, this scheme works not only for XOR NC, but also for normal NC.

Most of schemes mentioned above require such infrastructural elements as secure channels or public key infrastructural which are not provided in Delay Tolerant Networks (DTNs). In this scenarios, considering the adversary as another source, the question is whether the packets originate from the same source, and it doesn't matter who the source is. Based on this idea, L´aszl´o Czap [38] proposed the weak verification method that decides whether two packets have the same originator without accessing any public keys.

(2) Algebraic approaches

Except for the cryptographic schemes, algebra mechanism could also be used to defense pollution attacks.

Dai Bin [39] presented a polluted packets detection method based on multiple vectors orthogonal to the universal set of vectors for the transmitted NC packets. The main idea of this method is to calculate numbers of orthogonal vectors and send them to different forwarders and receivers which use them to judge whether the newly received packets are in the subspace based on the source packets. In this scheme, a transmission network which is an *n*-layer combination network is represented by a directed graph $G = (V, E)$. When the source node sends *h* packets with *N* symbols, the message matrix in NC can be presented as follows,

$$\begin{bmatrix} x_1 \\ \vdots \\ x_h \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 0 & x_{1,1} & x_{1,2} & \cdots & x_{1,N} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & x_{h,1} & x_{h,2} & \cdots & x_{h,N} \end{bmatrix}.$$

In this matrix, all $x_{i,j}$ are symbols in $\mathbb{F}_q$. Message transmitted in any edge *e* is the linear combination of $\boldsymbol{x}$, that is $\boldsymbol{y}(\boldsymbol{e}) = \sum_{i=1}^{h} \boldsymbol{g_i}(\boldsymbol{e})\boldsymbol{x_i}$. $\boldsymbol{g_i}(\boldsymbol{e})$ is the global encoding vector. So the vector $\boldsymbol{y}(\boldsymbol{e})$ is in $\mathbb{F}_q$. Since vectors $\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_h}$are linear independent vectors in $\mathbb{F}_q^{h+N}$, they can span a subspace $\boldsymbol{X}$. The vector $\boldsymbol{y}(\boldsymbol{e})$ is linear combination of $\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_h}$, so they are all in the subspace $\boldsymbol{X}$. Each vector in subspace $\boldsymbol{X}$ can be seen as a legal vector, which means that the vector is not fabricated. For this, orthogonal vectors can be introduced. According to the vectors $\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_h}$, a vector $\boldsymbol{v} = (\mathrm{v}_1, \mathrm{v}_2, \cdots, \mathrm{v}_{h+N}) \in \mathbb{F}_q^{h+N}$ could be found to satisfy the condition that $\boldsymbol{x_i} \cdot \boldsymbol{v} = 0,\ i = 1, 2, \cdots, h$. It means that the vector $\boldsymbol{v}$ is orthogonal to all $\boldsymbol{x}$, and $\boldsymbol{v}$ is orthogonal to the subspace $\boldsymbol{X}$. The vector $\boldsymbol{v}$ is distributed to the downstream nodes by the source. Then the downstream nodes check the vector $\boldsymbol{y}$ of each received packet whether it is orthogonal to $\boldsymbol{v}$ that is to compute whether $\boldsymbol{y} \cdot \boldsymbol{v} = 0$ is true. If the equation is not true, the packet is a polluted one.

If a malicious node wants to generate a packet that is not in subspace $\boldsymbol{X}$ but can pass the check without knowing the vector $\boldsymbol{v}$, on the assumption that it chooses a vector randomly, then this vector only has a probability 1/*q* to pass the check. In RLNC, *q* is always chosen to be 2^8 or 2^{16}, so the probability 1/*q* is very small and negligible. However, if the malicious node gets $\boldsymbol{v}$, the detection method will fail.

Elias Kehdi [40] presented a security algorithm based on the subspace properties of RLNC to detect and contain malicious attacks. The intermediate nodes verify the integrity of a block by checking if it belongs to the subspace spanned by the source blocks. This is feasible when every node has a vector orthogonal to all the combinations of the source blocks. These vectors, referred to as **Null Keys**, belong to the null space of the source blocks. The Null Keys algorithm allows nodes to rapidly detect corrupted blocks without decoding the blocks.

Jing dong [41] created non-cryptographic checksums and relied its security on the difference between the time when a packet was received and when the checksum was created. The checksums are used to verify the packets. The scheme is effective but requires time synchronization and delays packets before forwarding them. The scheme in [42] proposed Split Null Keys. In this scheme, the keys are split so that only a small portion of the key is updated periodically. The small updates are suitable for a scalable key distribution scheme that does not involve forwarders in creating keys and thus does not rely its security on constraints imposed on the network topology.

D Defense Schemes for Eavesdropping Attack

In the scenarios where sensitive information is communicated, the confidentiality is always the top security concern, and should be assured by all means. Existing schemes to resist these attacks are classified into two main categories. One is based on permutation encryption, the other is on topology optimization.

(1) Permutation encryption:

Yawen Wei [43] proposed two efficient coding schemes which were **weakly-secure** [44] against wiretapping attacks. Both schemes utilize a permutation function to randomize the message vectors sent by the source. The first inserts one random symbol into the source message; the second inserts no at all and thus achieves the maximum multicast capacity. Moreover, the second one does not consume more bandwidth to transmit symbols and encoding coefficients.

The first scheme inserts one random symbol into the source message. Assume the maximum multicast capacity is *n* in a coding system. So the message vector sent by the source is $X = (x_1, x_2, \cdots, x_{n-1}, w)^T$, where x_1 to x_{n-1} are information symbols and w is the random symbol. The source uses the h function and computes a vector $X' = (x_1', x_2', \cdots, x_n')^T$, where

$$\begin{cases} x_1' = x_1 + h(w) \\ x_2' = x_2 + h^2(w) \\ \cdots\ \cdots \\ x_{n-1}' = x_{n-1} + h^{n-1}(w) \\ x_n' = w \end{cases}$$

Here $h^2(w) = h\big(h(w)\big)$ means that *h* function is applied twice to parameter *w* and so forth. After the source calculated the vectorX', it multiplies a full rank matrix *C* of dimension *n*×*n* to X' and get $X'' = CX'$, then forwards it.

In the second scheme, no random symbol is inserted into the message vector. Instead, the source applied the *h* function to the information symbols directly. The message vector $X' = (x_1', x_2', \cdots, x_n')^T$ is:

$$\begin{cases} x_1' = x_1 + h(x_2) \\ x_2' = x_2 + h^2(x_3) \\ \cdots\ \cdots \\ x_{n-1}' = x_{n-1} + h^{n-1}(x_n) \\ x_n' = h(x_1') + \cdots + h(x_{n-1}') + x_n \end{cases}$$

Then the source can multicast X' to all receivers by any feasible insecure network code.

Peng zhang [45] proposed P-Coding, a novel security scheme against eavesdropping attacks in NC. With the lightweight permutation encryption performed on each message and its coding vector, P-Coding can efficiently thwart global eavesdroppers in a transparent way. The basic idea of P-Coding is to perform the permutation encryption on coded messages, as shown in Figure 4.

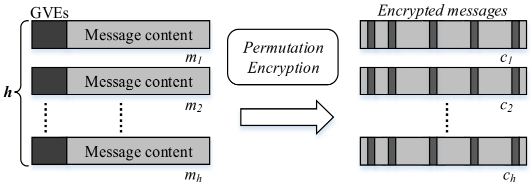


Figure 4. Permutation encryption on coded messages

After PEF (Permutation Encryption Function) operations, symbols of the messages and corresponding GEVs (Global Encoding Vectors) can be mixed and reordered together.

(2) Topology optimization

The idea of this type of schemes is to find the secure transmission topology, and then based on this topology, design a NC scheme to defense eavesdropping attacks.

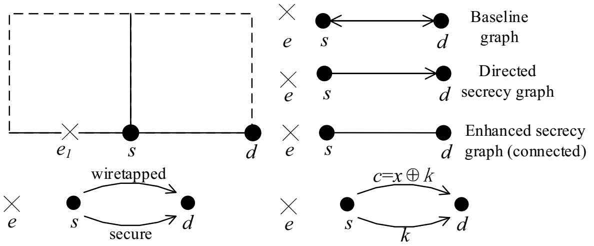


Figure 5. A simple two-way scheme

Cagatay Capar [46] presented a NC approach to help guarantee **information-theoretic secure** [19] (also be called Shannon secure) in the wireless environment. The main tools are described. The first is the wiretap network model [19], which is an abstract graph-based tool that allows a formal way to check whether secure communication is possible using NC. The second is the secrecy graph [47] which allows one to map a given physical wireless network topology to a graph. The authors gave three examples of small networks to demonstrate how NC helps wireless secrecy. The topology is a small square grid and the legitimate nodes are positioned on the corners with connections only to their nearest neighbors.

The first one is a simple two-way scheme. As shown in Figure 5, a secure connection to deliver a secret message from the source to the destination can be established, although the connection from the source to the destination is wiretapped.

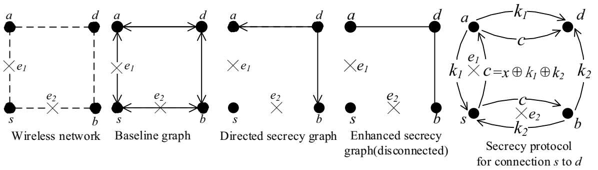

a
d
e_1
e_2
s
b
Wireless network
Baseline graph
Directed secrecy graph
Enhanced secrecy graph(disconnected)
k_1
c
$c=x\oplus k_1\oplus k_2$
k_2
Secrecy protocol for connection s to d

Figure 6 Non-collaborating eavesdroppers of known location

The second is Non-collaborating eavesdroppers of known location shown in Figure 6. The edges in both directions between *s* to *a* and *s* to *b* are wiretapped, hence the resulting secrecy graph is disconnected. Although the source is disconnected from both of its neighbors in both directions, with the aid of NC, delivering a secure message from *s* to *d* is possible.

The third is Eavesdroppers of unknown location shown in Figure 7. Compared with the second example, a very important difference in this one is that the locations of the eavesdroppers are unknown. Here, to check the secrecy condition, for each of the squares, we consider an optimally located eavesdropper. Regardless of eavesdroppers' locations, no eavesdropper in the network has any information about *x*.

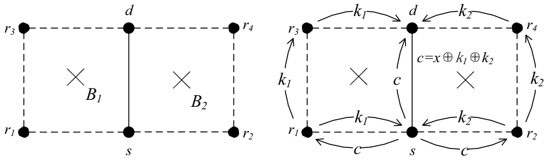


Figure 7. Eavesdroppers of Unknown Location

Xiangmao Chang [48] designed secure LNC against wiretap attacks. This scheme is proposed with the perspectives of network topology and NC. The authors first tried to find the transmission topology that is suited for NC. Based on the topology, they designed LNC scheme that is weakly-secure. Cai Ning [49] proposed a model called wiretap network. Danio Sliva *et al.* [50] proposed a coding scheme that can achieve the maximum possible rate of $k = n - \mu$ packets that are information-theoretically secure from the adversary. A distinctive feature of this scheme is that it is universal: it can be applied on top of any communication network without requiring knowledge of or any modifications on the underlying network code.

IV. Open issues

1 Cryptographic schemes are often computation expensive, which may be not suitable for the resource-constrained wireless network. For more practical, we should take computation and energy cost into account in algorithm design. For example, we can exploit some pre-configuration methods to reduce the communication and computation overhead. Or we design some efficient homomorphic algorithms inducing affordable computation overhead.

2 Most of existing defense schemes are only toward one or two attack forms and even underestimate the abilities of attackers. For instance, the defense schemes toward pollution attacks can't resist random forgery attacks and collusion attacks, and eavesdroppers are not cooperative. With the rapid development of attack technology, our secure schemes are expected to be more comprehensive to cope with various attacks.

3 Some schemes require special topologies or extra security channel, such as square gird network, two X network topology and so forth. In wireless network system, these requirements are probably too rigor. The more universal schemes are the trend in the future.

4 Due to the efficiency and robustness, NC can be applied in more and more emerging fields, such as cloud environment and Software Defined Network (SDN). The security of these applications is also a big challenge. The secure NC schemes satisfying their characteristics should be brought to attention.

TABLE III. A SUMMARY ABOUT THE MAIN ATTACKS AND DEFENSE SCHEMES

Attacks	Defense theory		Characteristics, Pros and cons	References
Entropy attacks	Cryptographic Algorithm		**PMAC** (homomorphic MAC), the key size is small and the computation overhead is low.	[21]
	Watchdogs		**Watchdogs** monitor the illegal behavior and packets. But the communication and computation overhead of the watchdogs could be expensive. Topology optimization makes the attackers out of the links, and tunes the topology within a unit time.	[20]
	Topology optimization			
	Algebra Mechanism		**S-PSLD** algorithm, its key idea is testing the linear-dependence for a given large number of vectors. It can rapidly filter out the resultant packets, minimize the packet detection cost, but verify the received packets in a probabilistic way instead of an exact way.	[22]
Byzantine attacks	Cryptographic	Checksum	1. In this scheme, **the attacker is regarded as another source**. If the sink node receives enough packets for decoding both sources, then it verifies the particular checksum, distills out the packet which does not satisfy constrains. The shortage of this scheme is that the junk packets just are distilled out at the sink node. 2. A **homomorphic signature** scheme allows packet-level Byzantine detection. The merit is efficient and no secure channel requirement. 3. A **homomorphic MAC** scheme can identify the precise location of all Byzantine attackers in intra-flow NC. And this scheme allows to eliminate any uncertainty in identifying attackers via subspace properties. 4. A secure random network coding model based on **CBC** is proposed to resist Byzantine attack. This scheme needs a low-rate secret channel to transmit coding coefficient.	[23],[24], [25],[26]
		Signature		
		Homomorphic MAC		
		CBC		
	Watchdogs		**Watchdogs** monitor all the illegal behavior. But the communication and computation overhead of the watchdogs could be expensive.	[18]
Pollution attacks	Cryptographic	1. Homomorphic Hash 2. Homomorpgic Signature 3. Homomorpgic MAC	1. Homomorphic Hash function is computation expensive. 2. Homomorphic signature schemes need no extra secure channels to transmit hashes. 3. Homomorphic MACs are compatible with NC, and they are employed to improve efficiency. However, some designs also have an expensive computation.	1.[30],[31] 2.[27],[28],[38] 3.[32],[33],[34],[35],[36],[37]
	Algebra Mechanism	Orthogonal Vectors	Message vectors $\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_h}$ are linear independent vectors in$\mathbb{F}_q^{h+N}$, they can span a subspace $\boldsymbol{X}$. A vector $\boldsymbol{v}$ is orthogonal to the subspace $\boldsymbol{X}$. In order to judge whether a packet is a pollution one, intermediate nodes just check the vector $\boldsymbol{y}$ of each received packet whether it is orthogonal to $\boldsymbol{v}$.	[40],[41], [42],[39]
		Null Space		
Eavesdropping attacks	Cryptographic (Permutation Encryption)		These schemes utilize a permutation function to randomize the messages and their coding vectors. These schemes have less communication cost than cryptographic schemes. But it needs secure channel to get random symbols or permutation keys.	[43],[45]
	Topology Optimization		These schemes integrate network topology design and NC design. Contrasted to the specific or arbitrary topology, this is a compromised idea to find the optimized secure topology.	[46],[48]

V. CONCLUSION

Network coding can achieve the maximum throughput of the network if the intermediate nodes are trusted. To some extent, since the information is encoded at intermediate node, it provides natural security in the communication. However, it is the mixture of packets that makes the network coding system vulnerable under many attacks, such as pollution attacks, entropy attacks and so on. In this paper, we give a survey about these attacks and defense schemes respectively. We summarize these approaches and mechanisms in TABLE III. Our work can provide a reference for future research.

ACKNOWLEDGMENT

This work was partially supported by the National Natural Science Foundation of China under Grant No. 61272451, 61173154, 61173175, 61100221. The corresponding author is Jing Chen.

REFERENCE

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, “Network information flow,” *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[2]Tracey Ho, Muriel Medard, “On Randomized Network Coding,” *Proceedings of the Annual Allerton Conference on Communication Control and Computing.* Vol.41, pp.11-20,2003

[3] Koetter Ralf, Muriel Médard, “An algebraic approach to network coding,” *IEEE/ACM Transactions on Networking.* Vol. 11, pp.782-795,2003

[4] A. M. Sheikh, A. Fiandrotti, and E. Magli, “Distributed scheduling for scalable P2P video streaming with network coding,” in *2013 Proceedings IEEE INFOCOM*, 2013, pp. 11–12.

[5] D. Annapurna, N. Tejas, K. B. Raja and K. R. Venugopal, “An energy efficient multicast algorithm for an Adhoc network using network coding and MAC scheduling,” in *2013 International Conference on Signal Processing and Communication (ICSC)*, 2013, pp. 62–67.

[6] S. Chieochan and E. Hossain, “Channel Assignment for Throughput Optimization in Multichannel Multiradio Wireless Mesh Networks Using Network Coding,” *IEEE Trans. Mob. Comput.*, vol. 12, no. 1, pp. 118–135, Jan. 2013.

[7] R. R. Rout and S. K. Ghosh, “Enhancement of Lifetime using Duty Cycle

and Network Coding in Wireless Sensor Networks," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 2, pp. 656–667, Feb. 2013.

[8] E. Altman, L. Sassatelli, and F. De Pellegrini, "Dynamic Control of Coding for Progressive Packet Arrivals in DTNs," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 2, pp. 725–735, Feb. 2013.

[9] Q. Yan, M. Li and Z. Yang, "Throughput Analysis of Cooperative Mobile Content Distribution in Vehicular Network using Symbol Level Network Coding," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 2, pp. 484–492, Feb. 2012.

[10] H. Yao, S. Jaggi, and M. Chen, "Passive Network Tomography for Erroneous Networks: A Network Coding Approach," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5922–5940, Sep. 2012.

[11] X. Shi, M. Medard, and D. E. Lucani, "Whether and where to code in the wireless packet erasure relay channel," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 8, pp. 1379–1389, Aug. 2013.

[12] A. Montanari and R. L. Urbanke, "Iterative Coding for Network Coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1563–1572, Mar. 2013.

[13]X. Yin, Y. Wang, X. Wang, X. Xue, and Z. Li, "A graph minor perspective to network coding: Connecting algebraic coding with network topologies," in *2013 Proceedings IEEE INFOCOM*, 2013, pp. 2364–2372.

[14] A. O. F. Atya, I. Broustis, S. Singh, D. Syrivelis, S. V. Krishnamurthy, and T. F. La Porta. Wireless network coding: Deciding when to flip the switch. *2013 Proceedings IEEE INFOCOM*, 2013: 260–264.

[15] S. Chachulski, M. Jennings, S. Katti, and D. Katabi, "Trading Structure for Randomness in Wireless Opportunistic Routing," in *Proceedings of the 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, New York, NY, USA, 2007, pp. 169–180.

[16] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "XORs in the Air: Practical Wireless Network Coding," in *Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, New York, NY, USA, 2006, pp. 243–254.

[17] Gkantsidis Christos, Rodriguez Pablo, "Cooperative Security for Network Coding File Distribution", *INFOCOM*. Vol. 3, pp.5, 2006.

[18] V. Pandit, J. H. Jun, and D. P. Agrawal, "Inherent Security Benefits of Analog Network Coding for the Detection of Byzantine Attacks in Multi-Hop Wireless Networks," in *2011 IEEE 8th International Conference on Mobile Adhoc and Sensor Systems (MASS)*, 2011, pp. 697–702.

[19] N. Cai and R. W. Yeung, "Secure network coding," in *2002 IEEE International Symposium on Information Theory, 2002. Proceedings*, 2002, p. 323–.

[20] A. J. Newell, R. Curtmola, and C. Nita-Rotaru, "Entropy Attacks and Countermeasures in Wireless Network Coding," in *Proceedings of the Fifth ACM Conference on Security and Privacy in Wireless and Mobile Networks*, New York, NY, USA, 2012, pp. 185–196.

[21] C. Cheng, T. Jiang, and Q. Zhang, "TESLA-Based Homomorphic MAC for Authentication in P2P System for Live Streaming with Network Coding," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 291–298, Sep. 2013.

[22] Y. Jiang, Y. Fan, X. (Sherman) Shen, and C. Lin, "A self-adaptive probabilistic packet filtering scheme against entropy attacks in network coding," *Comput. Netw.*, vol. 53, no. 18, pp. 3089–3101, Dec. 2009.

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

[24] M. Kim, L. Lima, F. Zhao, J. Barros, M. Medard, R. Koetter, T. Kalker, and K. J. Han, "On counteracting Byzantine attacks in network coded peer-to-peer networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 5, pp. 692–702, Jun. 2010.

[25] A. Le and A. Markopoulou, "Locating Byzantine Attackers in Intra-Session Network Coding Using SpaceMac," in *2010 IEEE International Symposium on Network Coding (NetCod)*, 2010, pp. 1–6.

[26] F. Tao, Z. Bingtao, and M. Jianfeng, "Security Random Network Coding Model against Byzantine Attack Based on CBC," 2011, pp. 1178–1181.

[27] Y. Jiang, H. Zhu, M. Shi, X. (Sherman) Shen, and C. Lin, "An efficient dynamic-identity based signature scheme for secure network coding," *Comput. Netw.*, vol. 54, no. 1, pp. 28–40, Jan. 2010.

[28]Z. Yu, Y. Wei, B. Ramkumar, and Y. Guan, "An Efficient Signature-Based Scheme for Securing Network Coding Against Pollution Attacks," in *IEEE INFOCOM 2008. The 27th Conference on Computer Communications*, 2008, pp.1409-1417.

[29] S. Agrawal and D. Boneh, "Homomorphic MACs: MAC-Based Integrity for Network Coding," in *Applied Cryptography and Network Security*, M. Abdalla, D. Pointcheval, P.-A. Fouque, and D. Vergnaud, Eds. Springer Berlin Heidelberg, 2009, pp. 292–305.

[30] Q. Li, J. C.-S. Lui, and D.-M. Chiu, "On the Security and Efficiency of Content Distribution via Network Coding," *IEEE Trans. Dependable Secure Comput.*, vol. 9, no. 2, pp. 211–221, Mar. 2012.

[31] M. N. Krohn, M. J. Freedman, and D. Mazi`eres, "On-the-fly verification of rateless erasure codes for efficient content distribution," in IEEE Symp. Security and Privacy, Oakland, CA, May 2004, pp. 226–240.

[32]A. Apavatjrut, W. Znaidi and A. Fraboulet, "Energy efficient authentication strategies for network coding," *Concurr. Comput. Pract. Exp.*, vol. 24, no. 10, pp. 1086–1107, Jul. 2012.

[33] H. Krawczyk, R. Canetti, and M. Bellare, "HMAC: Keyed-Hashing for Message Authentication." [Online].
Available: http://tools.ietf.org/html/rfc2104.

[34] B. Preneel and P. C. van Oorschot, "MDx-MAC and Building Fast MACs from Hash Functions," in *Advances in Cryptology — CRYPT0'95*, D. Coppersmith, Ed. Springer Berlin Heidelberg, 1995, pp. 1–14.

[35] J. Black and P. Rogaway, "CBC MACs for Arbitrary-Length Messages: The Three-Key Constructions," in *Advances in Cryptology — CRYPTO 2000*, M. Bellare, Ed. Springer Berlin Heidelberg, 2000, pp. 197–215.

[36] Y. Xu and K. Sakurai, "Cooperatively Securing Network Coding Against Pollution Attacks with Incentive Mechanism," in *Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication*, New York, NY, USA, 2012, pp. 52:1–52:10.

[37] Z. Yu, Y. Wei, B. Ramkumar, and Y. Guan, "An Efficient Scheme for Securing XOR Network Coding against Pollution Attacks," in *IEEE INFOCOM 2009*, 2009, pp. 406–414.

[38] L. Czap and I. Vajda, "Secure Network Coding in DTNs," *IEEE Commun. Lett.*, vol. 15, no. 1, pp. 28–30, Jan. 2011.

[39] D. Bin, S. Zhang, Y. Qu, J. Yang, and F. Wang, "Orthogonal vector based network coding against pollution attacks in n-layer combination networks,"in *2010 5th International ICST Conference on Communications and Networking in China (CHINACOM)*, 2010, pp.1–5.

[40] E. Kehdi and B. Li, "Null Keys: Limiting Malicious Attacks Via Null Space Properties of Network Coding," in *IEEE INFOCOM 2009*, 2009, pp. 1224–1232.

[41] J. Dong, R. Curtmola, and C. Nita-Rotaru, "Practical Defenses Against Pollution Attacks in Intra-flow Network Coding for Wireless Mesh Networks," in *Proceedings of the Second ACM Conference on Wireless Network Security*, New York, NY, USA, 2009, pp. 111–122.

[42] A. Newell and C. Nita-Rotaru, "Split Null Keys: A null space based defense for pollution attacks in wireless network coding," in *2012 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, 2012, pp. 479–487.

[43] Y. Wei, Z. Yu, and Y. Guan, "Efficient Weakly-Secure Network Coding Schemes against Wiretapping Attacks," in *2010 IEEE International Symposium on Network Coding (NetCod)*, 2010, pp. 1–6.

[44] Bhattad, Kapil, Narayanan, "Weakly secure network coding", NetCod, 2005

[45] P. Zhang, Y. Jiang, C. Lin, Y. Fan, and X. Shen, "P-Coding: Secure Network Coding against Eavesdropping Attacks," in *2010 Proceedings IEEE INFOCOM*, 2010, pp. 1–9.

[46] C. Capar and D. Goeckel, "Network coding for facilitating secrecy in large wireless networks," in *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, 2012, pp. 1–6.

[47] M. Haenggi, "The secrecy graph and some of its properties," in *IEEE International Symposium on Information Theory, 2008. ISIT 2008*, 2008, pp. 539–543.

[48] X. Chang, J. Wang, J. Wang, V. Lee, K. Lu, and Y. Yang, "On Achieving Maximum Secure Throughput Using Network Coding against Wiretap Attack," in *2010 IEEE 30th International Conference on Distributed Computing Systems (ICDCS)*, 2010, pp. 526–535.

[49] N. Cai and R. W. Yeung, "Secure Network Coding on a Wiretap Network," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 424–435, Jan. 2011.

[50] D. Silva and F. R. Kschischang, "Security for wiretap networks via rank-metric codes," in *IEEE International Symposium on Information Theory, 2008. ISIT 2008*, 2008, pp. 176–180.