Secure Network Coding against Content Pollution Attacks in Named Data Network

Tao Feng, Xiaomei Ma, Xian Guo, and Jing Wang

Abstract—Named Data Network (NDN) is one of the future Internet architecture, all nodes (i.e., hosts, routers) are allowed to have a local cache, used to satisfy incoming requests for content. However, depending on caching allows an adversary to perform attacks that are very effective and relatively easy to implement, such as content pollution attack. In this paper, we use a method of secure network coding based on homomorphic signature system to solve this problem. Firstly, we use a dynamic public key technique, our scheme for each generation authentication without updating the initial secret key used. Secondly, employing the homorphism of hash function, intermediate node and destination node verify the signature of the received message. In addition, when the network topology of NDN is simple and fixed, the code coefficients in our scheme are generated in a pseudorandom number generator in each node, so the distribution of the coefficients is also avoided. In short, our scheme not only can efficiently prevent against Intra/Inter-GPAs, but also can against the content poisoning attack in NDN.

Index Terms—Named data networking, content pollution attack, network coding signature.

I. INTRODUCTION

With the increasingly rich of applications and calculation modes, and with the enhancement of the social dependence on the Internet, Internet access and network function orientation has undergone tremendous change. In terms of scalability, mobility and security, TCP / IP architecture has been unable to satisfy the requirements of sustainable development of the Internet. Therefore, in order to better serve the future Internet, researchers trying to redesign the existing Internet architecture, and propose the Future Internet Architecture (FIA) [1]. Named Data Network (NDN) [2], [3] is one of the future Internet architecture, now is considered as a viable replacement for the current IP-based Internet, which uses the methods of name-based routing, forwarding and caching to optimize network transmission efficiency. In NDN, there is not the concept of connection, only the content publishers and content consumers. Content publishers need to encrypt and sign the content, content consumers need to decrypt and verify legal content. Moreover, all nodes (i.e., hosts, routers) are allowed to have a local cache, used to satisfy incoming requests for content. This makes NDN a good architecture for efficient large scale content distribution. However, depending on caching allows an adversary to perform attacks that are very effective and relatively easy to implement, such as content pollution attack [4], [5].

In theory, content signatures provide an effective and simple means for detecting content poisoning attacks in NDN [6], [7]. Currently, most of the signature algorithm are based on homomorphic hashing function. For example, Korhn et al. [8] using homomorphic hash function to construct the verification algorithm for the first time. However, the total size of the hash values of the scheme is proportional to the number of blocks, which could be very large and the cryptographic hash function is computationally expensive. Li et al. [9] employed a batch content distribution verification scheme, which reduced the computational cost of each node. Unfortunately, their scheme unable to realize the network max-flow and can bring delay to the sinks. Zhao et al. [10] from the perspective of linear space, presented a lower computational complexity signature scheme, but their scheme required long start-up latency. Yu et al. [11] proposed an RSA-based scheme, which firstly used the homorphism hash function to calculate the hash value of the source message, then used RSA algorithm to calculate signature of the hash value, then send message with signature. Rosario et al. [12] improved the Yu's scheme, they designed the RSA-based homomorphic signature scheme in integer field, and used the random oracle model to prove the security of the scheme.

With the deepening of the research, researchers found that the schemes of all the above only designed to prevent against Intra-Generation Pollution Attacks (intra-GPAs) do not work well for multiple generation propagations. In the case of multiple generation propagations, these schemes may exist inter-Generation Pollution Attacks (inter-GPAs). The so-called inter-GPAs, refers to that an adversary can take advantage of a valid message $W$ of the $k$-th generation, to act as the message of the $k+i$-th generation. Because this message does not belong to the $k+i$-th generation, so it is clearly a fake content. But the signature of this message is calculated by the source (calculated in the $k$-th generation), so this message can passed the verification of intermediate nodes, so that it can successfully forge a message of the $k+i$-th generation, and destroy the transmission of the $k+i$-th generation message. Guangjun L. et al. [13] used the dynamic public key technique, proposed a novel homomorphic signature scheme, in which the source does not update the system key frequently. This scheme can defeat Intra-GPAs and Inter-GPAs effectively.

In practice, however, the overhead of signature verification is a challenge in NDN router. The ref. [5] and [7] also show that the traditional signature verification have a high requirements of the router interface and computing power. So, when we choose signature scheme, not only consider the
security, but also consider the overhead. The ref. [14] presented a network coding signature algorithm of low communication overhead, which made the signature as redundant added into the source messages, then hashed the padded messages to the public parameters. This scheme has a lower communication overhead than other schemes. However, this method designed to prevent against Intra-GPAs, do not work well for multiple generations.

Inspired by the idea of ref. [13] and ref. [14], we propose a novel method with low communication overhead, particularly, that can efficiently prevent against Intra/Inter-GPAs. Eventually, we apply this method to named data network (NDN), to solve the content poisoning attack in NDN. Firstly, we use a dynamic public key technique, our scheme for each generation authentication without updating the initial secret key used. Secondly, employing the homomorphism of hash function, intermediate node and destination node verify the signature of the received message. In addition, when the network topology of NDN is simple and fixed, the code coefficients in our scheme are generated in a pseudorandom number generator in each node, so the distribution of the coefficients is also avoided.

II. HOMOMORPHIC HASH FUNCTION

Currently, most of the homomorphic signature algorithm are based on homomorphic hash function, Korhn et al. [8] using homomorphic hash function to construct the verification algorithm for the first time. The homomorphic hash function is as follows:

A trusted party globally generates a set of hash parameters $G = (p, q, g)$, where $p$ and $q$ are two large random primes such that $q | p - 1$. The hash parameter $g$ is a $1 \times n$ row-vector, composed of random elements of $F_p$ all order $q$. That is, $g = (g_1, \ldots, g_n)$, where $g_i^q = 1 \mod p \cdot g_i \in F_p$, $1 \leq i \leq n$.

Let $h_1, h_2, \ldots, h_m$ denote the hash value for messages $V_1, V_2, \ldots, V_m$, that is the hash value for messages $V_1, V_2, \ldots, V_m$, is generated as follows:

$$h_i = H(V_i) = \prod_{j=1}^{n} g_j^{v_{ij}} \mod p, 1 \leq i \leq m$$

This hash function is homomorphic. It has the additive homomorphism, because for arbitrary $X, Y \in F_q^n$, we have

$$H(X + Y) = \prod_{i=1}^{n} g_i^{(x_i + y_i)} = \left( \prod_{i=1}^{n} g_i^{x_i} \right) \cdot \left( \prod_{i=1}^{n} g_i^{y_i} \right) = H(X) \cdot H(Y)$$

Therefore, for arbitrary network node receives a message $W = (w_1, \ldots, w_n)$, $a_1, \ldots, a_m$ is a set of linearly independent vectors, then the hash value of the message $W = \sum_{j=1}^{m} a_j Y_j$ is generated as follows:

$$H(W) = \prod_{j=1}^{n} g_j^{v_{ij}} \mod p$$

$$= \prod_{i=1}^{n} g_i^{x_i + y_i} \mod p = \prod_{i=1}^{n} \prod_{j=1}^{m} g_j^{a_{ij}} \mod p$$

$$= \prod_{i=1}^{m} \left( \prod_{j=1}^{n} g_j^{v_{ij}} \mod p \right)^{a_{ij}} \mod p$$

$$= \prod_{i=1}^{m} h_i^{a_{ij}}$$

III. NDN CONTENT POISONING ATTACK

This section describes content poisoning attack scenarios. Content poisoning is an attack whereby adversary injects fake content into router caches. To be specific, the adversary attempts to inject fake content or modify the contents of the transmission in NDN, which will seriously interfere with the communications between the publisher and the receiver, in order to destroy the integrity of the content package. The purpose of this attack is to make the legitimate nodes can't identify the false content in the NDN. Due to the transmission mode of NDN, these polluted contents will be continuously spread, so that make the whole network traffic is contaminated.

In this paper, we consider a pro-active content poisoning, i.e., the adversary introducing malicious data into the contents, and injects fake content with the same name into router caches. Fake content can be injected into the network via malicious routers or end-nodes. For example, consider an adversary targeting a specific router $R$. Assuming that malicious producer sends a fake content to $R$ which is promptly cached. Consequently, $R$ is pre-polluted with fake content. If the router $R$ receives a real interest, it would send the fake content.

IV. SECURE NETWORK CODING AGAINST CONTENT POISONING ATTACKS IN NDN

In this paper, we consider a NDN network topology as shown in Fig. 1. It consists of $A$, $t_1, \ldots, t_k$ and a number of forwarders, where $A$ and $t_1, \ldots, t_k$ can be content publisher and content receiver in NDN. To facilitate the description, in this paper, $A$ is a content publisher, $t_1, \ldots, t_k$ are multiple content receivers. Content publishers produced messages in the form of the following:

$$v = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{pmatrix}$$

Let each source message $V_i$ is randomly picked from a finite field $F_q$ where prime $q$ is a pre-determined security parameter. So, each source message $V_i$ for $i=1, \ldots, m$ can be regarded as a row vector such as $V_i = (v_{i1}, \ldots, v_{in})$, where $v_{ij} \in F_q$ for $j=1, \ldots, n$ denote the codewords.
In this paper, our secure network coding based on homomorphic signature is mainly composed of four parts: Setup, Sign, Combine and Verify.

* Setup.

1) Given the system parameters $1', m, n, p, q, g$, where $p$ and $q$ are prime, and $q > 2^i$. Let $\alpha(g)$ denotes the order of $g$ in the field $F_p$ of prime order, $F_p$ is a subfield of $F_q$.

2) Choose $n+1$ random numbers $u_i(i=0, 1, \ldots, n)$ from $F_p$ such that

$$sk = \{ u_i \in F_q \mid i = 0, 1, 2, \ldots, n \}$$

and compute

$$pk = \{ p_i = g^{u_i} \mid i = 0, 1, 2, \ldots, n \}$$

* Sign.

To generate the signature tag using the secret key for the i-th basis vector $V_i = (v_{i1}, \ldots, v_{in}) \in F_q$ belonging to the forwarding generation, the content publisher works as follows:

1) Choose a unique generation identifier $id$ from $\{0, 1\}$ arbitrarily, which are used to scramble the above initial secret key $sk$ to $S$, i.e.,

$$S = (s_0, s_1, \ldots, s_n) \in F_q$$

where

$$s_0 = u_0$$

and

$$s_i = u_i + id \cdot u_0, (i = 1, 2, \ldots, n)$$

2) Compute the signature tag of $V_i$ by:

$$s_i = \sum_{j=1}^{m} s_j v_{ij} / s_0 \in F_q (1 \leq i \leq m)$$

That is, $V_i$ is signed as $v_i = (V_i, \sigma_i)$.

* Combine.

After receiving $l$ forwarding vectors $v_i = (V_i, \sigma_i)(i = 1, 2, \ldots, l)$, an intermediate node outputs:

$$\mathbf{v} = (W, \sigma_W) = \sum_{i=1}^{l} \alpha_i v_i = \sum_{i=1}^{l} \alpha_i V_i, \sum_{i=1}^{l} \alpha_i \sigma_i$$

where $\alpha_1, \alpha_2, \ldots, \alpha_l$ is code coefficients which are generated by the seed $\alpha_2$ of the node $X$ received (as show in Fig. 1), and $X$ send the seeds and message packets $\mathbf{v}$ to all descending node.

* Verify.

Given a forwarding vector $\mathbf{v} = (W, \sigma_W)$ belonging to the generation denoted by $id$, $W = \sum_{i=1}^{l} \alpha_i V_i$, a verifier (including the receivers) computes

$$H_1 = g_0^{\sigma_W} \cdot H(W)$$

$$H_2 = \prod_{i=1}^{l} g_i^{\alpha_i} \mod p$$

and checks if

$$\frac{H_1}{H_2} = 1$$

where

$$g_0 = p_0, g_1 = p_1 \cdot p_0^i (i = 1, 2, \ldots, n)$$

If $\frac{H_1}{H_2} = 1$ holds, the message is through the verification, otherwise, the verifier discard it.

V. CORRECTNESS AND SECURITY

A. Correctness

This section shows the correctness property of the constructed scheme.

Theorem 1: $W = (w_1, \ldots, w_n)$ is a combination of the received packets in a node, i.e.,

$$W = \sum_{i=1}^{l} \alpha_i V_i$$

and

$$H_1 = g_0^{\sigma_W} \cdot H(W), H_2 = \prod_{i=1}^{l} g_i^{\alpha_i} \mod p$$

If $\frac{H_1}{H_2} = 1$, $W$ is not polluted.

Proof: Let

$$H(W) = \prod_{i=1}^{l} g_i^{\alpha_i} \mod p$$

We have

$$g_0^{\sigma_W} \cdot H(W) = g_0^{\sigma_W} \cdot \prod_{i=1}^{l} g_i^{\alpha_i} \mod p$$

$$= g^{\sum_{i=1}^{l} \alpha_i v_{ij}} \mod p$$

$$= g^{\sum_{i=1}^{l} \alpha_i v_{ij}} \mod p$$

$$= g^{s_{ij}} \mod p$$

Due to $W = \sum_{i=1}^{l} \alpha_i V_i$, thus

$$H(W) = \prod_{i=1}^{l} g_i^{\alpha_i} \mod p$$

By Theorem 1, the message is not polluted.
Given the fake content $W = (w_1, \ldots, w_n)$ that make the equation $H_1/H_2 = 1$ holds, then we have:

$$g_0^{\sigma_W} \cdot \prod_{i=1}^n g_i^{w_i} = \prod_{i=1}^l g_i^{\alpha_i}$$

This is equivalent to solve the problem of discrete logarithm.

2) Given $\sigma_W$, computes $W \not\in \text{span}\{V_1^{(id)}, V_2^{(id)}, \ldots, V_m^{(id)}\}$ that make the equation holds, i.e.,

$$g_0^{\sigma_W} \cdot \prod_{i=1}^n g_i^{w_i} = \prod_{i=1}^l g_i^{\alpha_i}$$

We assume that the n-1 elements before $W = (w_1, \ldots, w_n)$ have been confirmed, thus the last element $w_n$ is computed as follows:

$$g_n^{w_n} = \prod_{i=1}^l g_i^{\alpha_i} / g_0^{\sigma_W} \cdot \prod_{i=1}^l g_i^{\alpha_i}$$

Obviously, if we want to compute $w_n$, we still solve the discrete logarithm problem. Therefore, the adversary $A$ cannot construct a message $W \not\in \text{span}\{V_1^{(id)}, V_2^{(id)}, \ldots, V_m^{(id)}\}$, so that $g_0^{\sigma_W} \cdot \prod_{i=1}^n g_i^{w_i} = \prod_{i=1}^l g_i^{\alpha_i}$. So, as long as $g_0^{\sigma_W} \cdot \prod_{i=1}^n g_i^{w_i} = \prod_{i=1}^l g_i^{\alpha_i}$, we can consider $W$ as a unpolished message.

D. The Security for Inter-GPAs

The adversary $A$ can take advantage of a valid message $W$ of the $k$-th generation, to act as the message of the $k+i$-th generation. Because this message does not belong to the $k+i$-th generation, so it is clearly a fake content. But the signature of this message is calculated by the source (calculated in the $k$-th generation), so this message can passed the verification of intermediate nodes, so that it can successfully forge a message of the $k+i$-th generation, and destroy the transmission of the $k+i$-th generation message. Next, we will prove this method does not pose a threat for our scheme.

We consider a message $W$ of the $k$-th generation, where $\sigma_W^{(k)}$ denote the signature of the message $W$ in the $k$-th generation. By the construction of the signature in Section IV, we have

$$\sigma_W^{(k)} = (s_i - \sum_{j=1}^{m} w_j s_j^{(k)}) \mod q, 1 \leq i \leq m$$

where $s_i^{(k)} = u_i + k \cdot u_i$, $s_i$ denote the private key of the $k$-th generation, so we have:

$$S^{(k)} = (s_0^{(k)}, s_1^{(k)}, \ldots, s_m^{(k)}) = (u_0^{(k)} + k \cdot u_0, u_1^{(k)} + k \cdot u_1, \ldots, u_m^{(k)} + k \cdot u_m)$$

If an adversary $A$ take advantage of this message $W$ of the $k$-th generation, to forge a message of the $id$-th generation, assume that the private key of the $id$-th generation...
\[ s_i^{(k)} (i = 0,1, \ldots , n) \]  
and 
\[ S^{(id)} = (s_0^{(id)}, s_1^{(id)}, \ldots , s_n^{(id)}) \]  
\[ = (u_0^{(id)} + id \cdot u_0, u_1^{(id)} + id \cdot u_1, \ldots , u_n^{(id)} + id \cdot u_n) \]

Obviously, \( S^{(id)} \neq S^{(k)} \), therefore, when the intermediate nodes verify this message, have:
\[ g_0^{(id)} \prod_{i=1}^{n} g_i^{(id)} = g_0^{(k)} \prod_{i=1}^{n} g_i^{(k)} \]
\[ = g \cdot g \]
\[ = g^{(1)} = g^{(1)} \cdot g^{\text{id}_{\text{uid}}} \]

For the given \( id \), by the above proof, the possibility of the equation \( g_0^{(id)} \prod_{i=1}^{n} g_i^{(id)} = g_0^{(k)} \prod_{i=1}^{n} g_i^{(k)} \) can be ignored. As a result, we consider \( g_0^{(id)} \cdot g^{\text{id}_{\text{uid}}} \neq \prod_{i=1}^{l} g_i^{(id)} \), i.e., the message \( W \) cannot be verified.

In conclusion, our scheme is secure against Intra/Inter-GPAs respectively.

VI. CONCLUSION

Some researches [5]-[7] shows that the signature is an effective means to solve the content poisoning attack in NDN. In this paper, we try to integrating network coding to NDN, using a secure network coding method which is based on homomorphic signature to solve a potential attacks — content poisoning attack in NDN, while enhancing the data transmission efficiency, improve the security of network data. Of course, we proved content signatures provide an effective means to detect content poisoning attack only in theoretically.

Currently NDN network and network coding have carried out extensive research coding in theory and practice [15]-[18], and have confirmed from the theoretical model that using network coding can improve network performance, but the verification steps or simulated environment, mostly based on assumptions or idealized model still have a gap with the actual application environment, so some conclusions have limitations. In addition, the process is mostly quantitative research. Therefore, building application evaluation model and practical system under NDN network combined with NC, remains to be further researched.

ACKNOWLEDGMENT

The research work was supported by National Natural Science Foundation of China under Grant No. 61461027 and No.61462060, and supported by Science and technology program of Gansu Province under Grant No. 1308RJZA277 and No.145RJZA078.

REFERENCES