A Privacy Preserving Network Coding Signature Scheme Based on Lattice

Bin Wu¹,², Caifen Wang²,*

¹College of Mathematics & Statistics, Northwest Normal University, 730070, Lanzhou, RP China
²College of Big Data and Internet, Shenzhen Technology University, 518118, Shenzhen, RP China
²wb1226@lzre.edu.cn

Abstract. Network coding is a desirable method to optimize network throughput and improve routing reliability, and is widely used in distributed Internet of Things systems. However, the packet mixing characteristic of network coding makes the transmission vulnerable to pollution attacks, which may hinder the reconstruction of the original data. Moreover, network coding has the potential to resist privacy threats such as eavesdropping attacks since intermediate nodes encourage coding/mixing operations, however, once adversary collects enough packets, simple network coding deployment can not achieve this purpose. In this paper we first present a privacy-preserving network coding signature scheme using lattice theory, which can be viewed as a promising tool for designing post-quantum cryptographic protocols. We blind the global coefficient matrix by encrypting the original encoding vector so as to resist eavesdropping attack, while preventing both intra/inter-generation pollution attacks by simultaneously signing the data packet and the generation identifier. We showed that our scheme is correct, and also proved that our scheme is secure against both eavesdropping and pollution attacks under the security of pseudo-random function and the small integer solution (SIS) assumption in standard lattices.

1. Introduction
The rapid development of the Internet and communication technology has promoted the Internet of Things (IoT) to become a leading technology that brings convenience to people's production and life. More and more smart terminals are connected to the Internet, and various information has been exchanged and shared among these terminals [1]. According to the International Data Corporation (IDC) report, by 2020, there will be nearly 28 billion IoT devices installed [2]. Given the the scale of IoT’s expansion accompanied by a surge in shared data, it is very essential to increase the throughput in such a huge decentralized network. The initial motivation of network coding was to increase the throughput of decentralized networks and indeed, this technology is considered to be a good approach to improve data sharing in peer-to-peer networks and digital content distribution over the Internet [3]. In short, network coding is a message-switching technique with both routing and coding functions. The core idea is to allow intermediate nodes to perform (non-)linear operations over some received data blocks and then forward it to downstream nodes. This technology not only optimizes the network throughput [4], but also improves the robustness of the network [5]. However, due to the information-mixing nature of network coding, the system deployed with network coding is highly susceptible to pollution attacks, where an outside adversary or internal malicious nodes can inject some modified or
fake packets into the network. This problem is particularly serious because even a single polluted packets may propagate and contaminate the entire data, resulting in destination decoding failure, thus reducing network throughput. Up to now, all existing method for thwarting this kind attacks can be categorized into two types: information theoretic and cryptographic approaches.

**Information theoretic approaches:** With the help of error-correcting techniques, the information theory method is proposed to resist pollution attacks [6, 7]. The work of [6] extended traditional error correction codes to network error correction codes. Refs [7] provided network error correction code with the minimum rank. However, Existing information theoretic schemes do not rely on any computational assumptions and only provide relatively limited pollution detection / correction in an end-to-end manner [8, 9]. Therefore, it does not help the transponder prevent the transmission of invalid pollution packets, in addition, the communication overhead of these approaches is heavier.

**Cryptographic approaches:** On the contrary, the scheme based on cryptography enables the transponder to verify the validity of the packets it receives on the route, thus rejecting malformed packets before polluting the downstream stream. Hence, in recent years, cryptographic solutions have attracted researchers’ attention, including Symmetric-Key Cryptography (SKC) based approaches, such as homomorphic(broadcast) MAC [10], and Public-Key Cryptography (PKC) based approaches, such as homomorphic signature and homomorphic hashing [11].

However, almost all above-mentioned linearly homomorphic signature schemes are based on traditional number-theoretic primitives (e.g., the integer factorization problem and the discrete logarithm problem). These constructions are vulnerable to quantum cryptanalysis [12].

### 1.1 Our Contributions

We noticed that the existing PKC-based NCS schemes either cannot resist quantum computer attacks (based on some traditional number-theoretic primitives) or none of the post-quantum cryptographic schemes take into account eavesdropping attacks, which may lead to private information leakage to unauthorized users. In order to solve both eavesdropping and pollution attacks in an architecture with network coding deployed, in this paper, we propose a post-quantum secure network coding signature scheme with privacy preserving in the random oracle. Note that in this paper the pollution attacks studied include both intra-GPAs and inter-GPAs. Our contributions mainly consist of the following aspects:

1. We present a lattice-based network coding signature scheme with privacy preserving. Specifically, we blind the global coefficient matrix by encrypting the original encoding vector so as to resist eavesdropping attack, while preventing both intra/inter-generation pollution attacks by simultaneously signing the data packet and the generation identifier.

2. We prove that our scheme is secure against both eavesdropping and pollution attacks under the security of pseudo-random function and the small integer solution (SIS) assumption in standard lattices.

### 2. System model and Threats model

#### 2.1. A privacy-preserving network coding model

In linear network coding, the messages to be transmitted is divided into multiple generation, and each generation is uniquely identified by a generation identifier. Furthermore, a file is viewed as a sequence of $n$-dimensional vectors $\mathbf{v}_1, \cdots, \mathbf{v}_m \in F_p^n$ over the finite field $F_p$, where $p$ is prime. Individual vectors are usually called blocks or packets.

In a network coding model, three stages are performed to complete the file transmission:
Figure 1: Demonstration of encryption and signature process of source node.

1) **Source node.** Before transmission, the source node creates $m$ augmented vectors $(v_1, \cdots, v_m)$, where each $v_i$ is obtained by prepending to $\bar{v}_i$ a $m$-dimensional unit vector $e_i$, with $1$ in position $i$ and $0$ elsewhere, i.e., $v_i = (\bar{v}_i, 0, \cdots, 0, 1, 0, \cdots, 0) \in F_p^{m+1}$. This way, the vectors $v_1, \cdots, v_m$ form a basis of a subspace $V \subset F_p^{m+1}$.

Regarding privacy protection in network coding, an intuitive way to solve the privacy problem is though encryption, only encrypting the coding vector is sufficient to prevent eavesdroppers from retrieving the original messages. This is because the global coefficient matrix is encrypted, the eavesdropper cannot get the correct coefficient, and thus cannot solve the linear equations to recover the original message.

Hence, the source node encrypts each augmented vector $v_i$ to generate the corresponding encrypted block $v_i^{(E)} = (\bar{v}_i, 0, \cdots, 0, c_i, 0, \cdots, 0) = (\bar{v}_i, c_i)$.

where $c_i$ is the encryption result of the $m$-dimensional unit vector $e_i$, using the secret key and the generation identifier $id$. Finally, the source node generates a coded data block $w^{(E)} = \sum_{i=1}^{m} a_i v_i^{(E)}$, where $a_i$ is randomly selected from $F_p$. Hence, a complete block to send can be written as $(id, w^{(E)}, \sigma)$, where $\sigma$ is the signature on $w^{(E)}$ and can be computed by calling the evaluation algorithm of our proposed homomorphic network coding signature scheme.

2) **Intermediate node.** Whenever a intermediate node receives the blocks $\{ (id, w_i^{(E)}, \sigma_i) \}_{i=1}^{k}$ on its $k$ incoming edges, it computes a linear combination of the received blocks, e.g., $w^{(E)} = \sum_{i=1}^{k} a_i w_i^{(E)}$, where $a_i \in F_p$. The resulting vector $w^{(E)}$ along with its signature $\sigma$ derived form $\{ \sigma_i \}_{i=1}^{k}$ are then transmitted through its outgoing edges. Thus, in this network, all blocks transmitted on any edge are linear combinations of the original encrypted blocks $v_1^{(E)}, \cdots, v_m^{(E)}$.

3) **Destination node.** Upon receiving valid blocks, the destination node first decrypts the data block, if the node has collected $m$ linearly independent decrypted vectors $w_1, \cdots, w_m$ of the form $w_i = (w_i^L, w_i^R)$, where $w_i^R (w_i^L)$ denotes the right-most ($m$ left-most $n$) positions of the vector. The node then computes the invertible matrix $G$ of the global coefficient matrix. Finally, the original messages can be recovered by computing...
2.2. Threats model

We consider the following two types of attacks.

- Pollution attack: A malicious node may inject invalid blocks into the network, so the original messages cannot be rebuilt. This is particularly sensitive because a single error introduced by a node can be propagated by honest nodes. Furthermore, as stated in [21], the pollution attacks can be divided into two categories, that is, intra-GPAs and inter-GPAs, which are described in detail as bellow.
- Eavesdropping attacking: This attacker can be considered as a passive eavesdropper who has the ability to observe all network links. This attacker attempts to recover the original messages by collecting linearly independent data blocks in transit and solving linear equations.

3. Our construction

In this section, we present a concrete privacy-preserving network coding signature scheme, which is composed of six polynomial time algorithms (Setup, Encrypt, Sign, Combine, Verify, Decrypt) with the following functionality:

- Setup \((1^n, m)\): Taking as input a security parameter \(n\) and a positive integer \(m\), this algorithm preforms the following steps:
  1. Choose two primes \(p, q\) with \(q \geq ((n+m)kp)^2\). Define \(l = \left\lfloor \frac{(n+m)}{6 \log q} \right\rfloor\).
  2. Set \(\Lambda_1 = p\mathbb{Z}^{n+m}\).
  3. Sample one matrix with associated trapdoor \((A, T_A) \leftarrow \text{TrapGen}(n+m, q, l)\), here \(A \in F^{l \times (n+m)}_q\) alone with a short basis \(T_A\) of \(\Lambda_q^\perp(A)\). Define \(\Lambda_2 = \Lambda_q^\perp(A)\) and \(T = p \cdot T_A\). Note that \(T\) is a basis of \(\Lambda_1 \cap \Lambda_2 = p \cdot \Lambda_2\).
  4. Set \(v = p \cdot \sqrt{(n+m) \cdot \log q \cdot \log(n+m)}\).
  5. Let \(H : \{0,1\}^* \rightarrow F^l_q\) be a hash function.
  6. Let \(F : K \times \{0,1\}^* \times \{0,1\}^* \rightarrow F^k_p\) be a keyed pseudo-random function (PRF) and \(k \in \mathbb{K}\).

- Encrypt \((id, k, w)\): Let \((e_1, \cdots, e_m)\) represents the canonical basis of \(\mathbb{Z}^m\), that is \(e_i\) is the \(i\)-th unitary vector. Based on a PRF, the source node computes \(c_{id,i} = F_k(e_i, id)\), for \(i \in [m]\), and it generates the encryption matrix

\[
E_{id} = \begin{pmatrix}
  c_{id,1} & 0 & \cdots & 0 \\
  0 & c_{id,2} & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \cdots & c_{id,m}
\end{pmatrix}
\]

Since the dataset identifier is introduced as an input in the encryption algorithm, the encryption matrix of each generation is guaranteed to be different. For each augmented block \(v_i = (\overline{v}, 0, \cdots, 0, 1, 0, \cdots, 0), i \in [m]\) identified by \(id\), the source node generates the corresponding encrypted block as below

\[
v_{i}(E) = Encrypt(k, id, v_i) = (\overline{v}_i, 0, \cdots, 0, c_{id,j}, 0, \cdots, 0)
\]
Hence, for any data block \( w \in \text{Span}(v_1, \ldots, v_m) \), let \( w = (w^L, w^R) = (w_1, \ldots, w_n, w_{n+1}, \ldots, w_{n+m}) \), we have \( w^{(E)} = \text{Encrypt}(k, id, w) = (w^L, w^R^{(E)}) = (w_1, \ldots, w_n, w_{n+1}^{(E)}, \ldots, w_{n+m}^{(E)}) \), where \( w^{R(E)} = w^R \cdot E_{id} = (w_{n+1}^{c_{id,1}}, \ldots, w_{n+m}^{c_{id,m}}) \).

- **Sign \((T, id, v^{(E)}, i)\)**: On input secret key \( T \), a identifier \( id \in \{0,1\}^n \), an encrypted vector \( v^{(E)} \in F_{p+m}^n \), and an index \( i \), do:
  1. Compute \( a_i = H(id||i) \in F_p^l \).
  2. Compute \( t \in Z^{r+m} \) such that \( t \mod p = v^{(E)} \) and \( A \cdot t \mod q = c_{id,i} \).
  3. Output \( \sigma \leftarrow \text{SamplePre}(\Lambda_1 \cap \Lambda_2, T, t, v) \in (\Lambda_1 \cap \Lambda_2, T) + t \).

- **Combine \((id, \{(a_i, w_i^{(E)}, \sigma_i)\}_{i=1}^k)\)**: Given an identifier \( id \) and a set of tuples \( \{(a_i, w_i^{(E)}, \sigma_i)\}_{i=1}^k \) with \( a_i \in Z_p \), this algorithm computes \( \sigma = \sum_{i=1}^k a_i \sigma_i \), \( w^{(E)} = \sum_{i=1}^k a_i w_i^{(E)} \). Hence, the combined packed block to forward is given by \((id, w^{(E)}, \sigma)\).

- **Verify \((PK, id, w^{(E)}, \sigma)\)**: On input a public key \( PK \), a identifier \( id \), and a data block \( w^{(E)} = (w_1, \ldots, w_n, w_{n+1}^{(E)}, \ldots, w_{n+m}^{(E)}) \) as well as its signature \( \sigma \), do: If all of the following three conditions holds, output 1 (accept), otherwise 0 (reject):
  1. \( \| \sigma \| \leq mpv \sqrt{n+m} \).
  2. \( \sigma \mod p = w^{(E)} \).
  3. \( A \cdot \sigma \mod q = \sum_{i=1}^m w_{n+i}^{(E)} a_i \).

- **Decrypt \((k, id, w^{(E)})\)**: Given the keyed pseudo-random function \( F \) and the secret key \( k \), the destination node computes \( \{c_{id,i} = F_k(e_i, id)\}_{i=1}^k \) and generated the decryption matrix

\[
D_{id} = \begin{pmatrix}
    c_{id,1}^{-1} & 0 & \cdots & 0 \\
    0 & c_{id,2}^{-1} & \cdots & 0 \\
    \vdots & & & \vdots \\
    0 & 0 & \cdots & c_{id,m}^{-1}
\end{pmatrix}
\]

Let \( w^{(E)} = (w^L, w^{R(E)}) \), where \( w^{R(E)} = (w_{n+1}^{(E)}, \ldots, w_{n+m}^{(E)}) \) denotes the coding vector of \( w^{(E)} \). Then the decrypted coding vector is given by \( w^R = w^{R(E)} \cdot D_{id} \).

Hence, the decrypted block can be written as \( w = \text{Decrypt}(k, id, w^{(E)}) = (w^L, w^R) \).

**Correctness.** We require that for each key pair \( w^{(E)} \) $$(PK, SK)$$ output by **Setup**, the following hold:

1. Decryption correctness. For all \( id \in \{0,1\}^n \) and any \( w \in \text{Span}\{v_i\}_{i=1}^m \), \( \text{Decrypt}(k, id, \text{Encrypt}(k, id, w)) = w \).

2. Authentication correctness. For all \( id \in \{0,1\}^n \) and \( v^{(E)} \in F_{p+m}^n \), if \( \sigma \leftarrow \text{Sign}(id, T, v^{(E)}, i) \), then \( \text{Verify}(PK, id, v^{(E)}, \sigma) = 1 \).
(3) Evaluation correctness. For all \( id \in \{0,1\}^n \) and all sets of triples \( \{(a_i, \sigma_i, w_i(E))\}_{i=1}^k \), if \( \text{Verify} (PK, id, w_i(E), \sigma_i) = 1 \) for each \( i \in [k] \), then \( \text{Verify} (PK, id, \sum_{i=1}^k a_i w_i(E)) \), Combine \( (id, \{(a_i, w_i(E), \sigma_i)\}_{i=1}^k) \) = 1.

**Theorem 1.** The proposed privacy-preserving network coding signature scheme is correct.

### 3.1. Security analysis

In this subsection, we analyze the security of the above concrete construction.

#### 3.1.1. Resistance to pollution attacks

The unforgeability of a homomorphic network coding signature scheme against adaptive chosen dataset attacks can be formalized by the following game between challenger and adversary \( A \).

- **Setup:** The challenger runs **Setup** \((1^n, m)\) to generate the public/secret key pair \((PK, SK)\). It then gives \( PK \) to adversary \( A \) while keeping \( SK \) secret.

- **Queries:** Proceeding adaptively, \( A \) specifies a sequence of queries for the vector subspace \( V_i^{(E)} \subseteq F_{p^{n+m}} \), which is described by encrypted augmented basis vectors \( \{v_j^{(E)}\}_{j=1}^m \). For each \( i \), the challenger chooses a generation identifier \( iid \) uniformly from \( \{0,1\}^n \), and return \( iid \) and \( \text{Sign}(SK, iid, v_j^{(E)}) \) to \( A \).

- **Output:** \( A \) outputs an identifier \( id^* \), a nonzero vector \( v^* \in F_{p^{n+m}} \), and a signature \( \sigma^* \).

The adversary wins if \( \text{Verify}(id^*, PK, v^*, \sigma^*) = 1 \), and one of the following conditions is met

1. \( id^* \neq id_i \) for all \( id_i \) that appear in the signing queries (Type 1 forgery).
2. \( id^* = id_i \) for some \( i \), but \( v^* \notin V_i^{(E)} \) (Type 2 forgery).

**Definition 3.** (Unforgeability) The proposed signature scheme is unforgeable against adaptive chosen dataset attacks for polynomial-time adversaries \( A \) if the advantage of \( A \) winning the above game is negligible in security parameter.

**Theorem 2.** Assuming the hardness of the \( \text{SIS}_{q,n,\beta} \) for \( \beta = m \cdot p^n \cdot n \log n \sqrt{\log q} \), our homomorphic network coding signature scheme is unforgeable in random oracle model.

#### 3.1.2. Resistance to eavesdropping attacks

**Definition 4.** The proposed scheme is said to be resistant to the eavesdropping attack if for any batch of \( k \) data blocks of the \( i \)-th generation identified by \( id_i \), and collected by adversary \( A \), denoted by \( \{w_i^{(E)}\}_{i=1}^k \), the probability that the \( A \) can recover the original messages from \( \{w_i^{(E)}\}_{i=1}^k \) is a negligible function of the security parameter.

**Theorem 3.** The proposed scheme is resistant to the eavesdropping attack.

### 4. Conclusion

In this paper, we first propose a lattice-based network coding signature scheme with privacy preserving, which can mitigate intra-generation/inter-generation pollution attacks, while using pseudo-random functions and the inherent security nature of random network coding to prevent eavesdropping attacks. We showed that the construction of our scheme is correct, and also proved that our scheme is resistant against both eavesdropping attacks and pollution attacks by demonstrating that the probabilities of an adversary recovering the original messages and forging a valid signature for a forged packet are both negligible functions of the security parameter, under the security of pseudo-
random function and the small integer solution (SIS) assumption in standard lattices.

Acknowledgments

References

[1] Atzori L, Iera A, Morabito G. The Internet of Things: A survey. COMPUT NETW 54:2787-2805 (2010).
[2] Alaybeyi SB. Pragmatic Strategies to Improve Industrial IoT Security. tech rep Gartner (2016).
[3] Ren H, Li H, Dai Y, Yang K, Lin X. Querying in Internet of Things with Privacy Preserving: Challenges, Solutions and Opportunities. IEEE Network 32(6):144-151 (2018).
[4] Krohn M, Freedman M, Mazieres D. On the-fly verification of rateless erasure codes for efficient content distribution. In: Proc. IEEE Symposium on Security and Privacy, Berkeley, CA, USA, pp.226-240 (2004).
[5] Jin J-Q, Ho T, Viswanathan H. Comparison of network coding and nonnetwork coding schemes for multi-hop wireless networks. In: Proc. 2006 IEEE International Symposium on Information Theory (ISIT), Seattle, WA, USA, pp.197-201 (2006).
[6] Cai N, Yeung R W. Network coding and error correction. In: Proc. of IEEE International Symposium on Information Theory (ISIT), pp.119-122 (2002).
[7] Zhang Z. Network error correction coding in packetized networks. In: IEEE Information Theory Workshop, pp.433-437 (2006).
[8] Ho T, Leong B, KoeetterOR, et al. Byzantine Modification Detection in Multicast Networks Using Randomized Network Coding. IEEE Transactions on Information Theory 54(6):2798-2803 (2008).
[9] Jaggi S, Langberg M, Katti S, et al. Resilient Network Coding in the Presence of Byzantine Adversaries. IEEE Transactions on Information Theory 54(6):2596-2603 (2008).
[10] Agrawal S, Boneh D. Homomorphic MACs: MAC-based integrity for network coding. In: Proc. International Conference on Applied Cryptography and Network Security (ACNS), Springer, NY, USA, pp.292-305 (2009).
[11] Boneh D, Freeman D, Katz J, Waters j. Signing a Linear Subspace: Signature Schemes for Network Coding. In: Proc. International Workshop on Public Key Cryptography (PKC), vol 5443. Springer, Berlin, Heidelberg, pp.68-87 (2009).
[12] Shor PW. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM Journal on Computing 26(5):1484-1509 (1997).