A Method of Network Coding to Ensure Public Network Security

Zihe Liang
Beijing University of Technology, Chaoyang District, Beijing, China
chunchenxiao@cas-harbour.org

Abstract. When connecting to the wireless network in public places, users’ personal information will be exposed in the public network, which is easy to be stolen and tampered with by third parties. In this paper, the author proposes a method to protect user information when connecting to public wireless networks in order to protect users’ privacy. Random linear network coding is used to encrypt the accessed user serial number, which is then forwarded to the upper layer. When transmitted to a specific node, the encrypted information is transmitted to the server for information decryption. After decoding the user serial number, the node is accessed. In this way, we can encrypt user information, reduce the delay time of forwarding, and improve the transmission efficiency. The time complexity is $O(n^2)$.

1. Introduction
The information transmitted in traditional routing mechanism cannot be superimposed. Intermediate nodes (such as router switches) store and forward data packets only. In multicast transmission environment, the maximum transmission capacity determined by the maximum flow minimum cut theorem is usually not achieved. Network coding technology, a new technology proposed by researchers in 2000, has attracted the attention of many researchers. Ho et al. and Jaggi et al. proposed random linear network coding and deterministic linear network coding respectively.

However, each node in the network processes the information which are received on each channel linearly or non-linearly by using network coding, and then forwards it to the downstream node. The intermediate node acts as an encoder or signal processor. In this way, the maximum flow bound of multicast routing transmission can be achieved, and the transmission efficiency of information can be improved. The main advantages are as follows: 1) Enhancing network throughput: whether wireless or wired network, the greater the number of network nodes, the more obvious the advantage of network coding in throughput; 2) balancing network load and distributing network traffic over a wider network; 3) improving bandwidth utilization; 4) saving node energy consumption in wireless network.

With the development of mobile Internet, Location-based service (LBS) becomes more and more important. The technology has been widely used in social networking, navigation, food and other fields. However, more and more facts have proved that LBS not only brings convenience to people, but also poses a great threat to users’ privacy and security. Therefore, in recent years, researchers have been looking for a way to enjoy LBS convenience while minimizing the exposure of personal information. Security network coding can be used to solve this problem and the location of every user should be encoded by linear operation. When server asks for one user’s location, this location does linear or other operations before sent to another node whose node can be decided by closest distance. Finally, the location received by server has been encoded for many times and every user’s location is also encoded to protect user information safety.
2. Network Coding

2.1 Basic Knowledge
Taking the classical butterfly network as an example, the basic idea of network coding is illustrated. In the communication network $G = (V, E)$ shown in Figure 1: $V$ represents the node set in the network; $E$ represents the edge set; $S$ is the source node; $C$ and $D$ are the destination nodes; other nodes are the intermediate nodes; channel capacity is the unit capacity; $S$ sends messages $a$ and $b$ to the two destination nodes $C$ and $D$. Both (a) and (b) are traditional routes for data transmission, while (c) is network coding for data transmission. Because channel capacity is unit capacity, it is easy to find that at $U$ node, traditional routes can only choose to send $a$ or $b$ to $V$ within the same time.

Only one of destination nodes can receive $a$ and $b$ at the same time. If the network coding transmission mode is adopted, the $U$ node performs exclusive or operation $a \oplus b$ on the received $a$ and $b$, and then sends $a \oplus b$ to the subsequent nodes. When $C$ receives $a$ and $a \oplus b$, it can get $b$ through $a \oplus a \oplus b = b$. Similarly, $D$ can also get $a$, so that $C$ and $D$ can receive $a$ and $b$ at the same time. This is the network originally proposed. The coding theory can realize the maximum flow of the network.

![Fig.1 Basic idea of network coding](image)

2.2 Characteristic

2.2.1 Improve throughput rate. From the above description, it can be seen that network coding can improve network throughput, which is one of its obvious characteristics. In detail, network coding can improve the throughput of broadcast, unicast and multicast networks. Unicast refers to one-to-one relationship in which a source sends messages to one destination; multicast refers to one-to-many
relationship in which one source node sends messages to multiple destination nodes; broadcast refers to one-to-all relationship in which one source node sends messages to all destination nodes.

2.2.2 Balanced flow. Network coding can also equalize traffic as shown in Figure 2. Each link has a capacity of 2 bits. Source nodes send messages m and n to intermediate nodes via multicast method. Routing is still compared with network coding. In the routing mode, the capacity of each link in the graphical multicast tree is 2 bits. As shown in Figure 2(a), the source S transmits the message m and n to the intermediate node and transmits them in the graphical mode, so only 5 of them are utilized. Using network coding method, as shown in Figure 2(b), where node D receives messages m and n directly, intermediate node R1 can decode n according to m and m ⊕ n, intermediate node R2 can decode m from and m ⊕ n. In this case, each link transmits 1 bit data, and all links are used, which is the role of traffic balancing.

![Fig. 2 Balanced flow of networking](a) (b)

2.2.3 Improve reliability. Network coding can also improve the reliability of the network. For example, in distributed storage, as shown in Figure 3, source S sends data m and n to nodes A and B respectively. Node U serves as a backup node here, and its function is to store m or n. If node D needs all the data of the source, A and B can send the data directly to D. At this point, if A or B has problems, D can get the required data through either U or A ⊕ B.

If network coding is not used, D may not get the required data, because U can only store m or n. Assuming that U stores m at this time, when B has a problem, D can not get the required data through U, because U can not get the data n. Similarly, if U stores N and A fails at this time, D cannot get m.

This kind of situation can be avoided by using network coding. Let U store m and n at this time. If something goes wrong with B, as shown in Figure 3 (b), D can get m and n from U, and then decode it to get n. If A makes an error as shown in Figure 3 (c), D can get m n from U, and then decode to get M. It can be seen that network coding can improve the reliability of the network.

![Fig. 3 Improved reliability](a) (b)
2.3 Classification

2.3.1 Linear network coding. The key problem of constructing linear network coding is to determine the coefficients of the coding function. When the intermediate node receives the information of each link, it calculates the received information with its own specific coding coefficient matrix. These coding coefficients are determined and distributed centrally. When all the intermediate coding nodes on a transmission link encode the data matrix. After mutual operation, if the destination node still satisfies the full rank condition, it can ensure that the destination node can decode correctly and obtain the source information. The precondition of centralized coding scheme is that the source node knows the whole network topology, so as to assign coding coefficients to the intermediate node, and the network topology is static.

2.3.2 Random network coding. When the relevant data is received, the corresponding data are linearly combined by randomly selected coefficients, which is called random network coding, and the linear equations are solved by Gauss elimination method when decoding.

(1) Coding principle. As shown in Figure 4, the coefficients \((s_1, s_2, \ldots, s_i)\) on each link are randomly selected from the finite field, and the intermediate nodes in the graph transmit the correlation coefficients and combinations to the corresponding next nodes. The source transmits the linear combination \(s_1m + s_2n\) and coefficients to node A, then node A transmits them to U. At the same time, source S transmits linear combination \(s_3m + s_4n\) and coefficients to node B, and then B transmits them to U. After receiving two message combinations \(M_2\) and \(M_3\), U makes linear combination operation and obtains \(M_3\). Then U transmits the coefficients \(s_5\) and \(s_6\) to \(R_1\) and \(R_2\). Finally, destination nodes get two equations from the parent nodes. It can travel the receiving message to server.
(2) Decoding principle. Here the author discusses it privately for the sink and lodging nodes. Based on the data packet it receives, two equations are made, as follows:

$$M_3 = s_5M_1 + s_6M_2$$  \hspace{1cm} (1)  
$$M_1 = s_1m + s_2n$$  \hspace{1cm} (2)

As long as the coefficient matrix of $M$ and $N$ is full rank, $m$ and $N$ can be obtained by simultaneous equations. The same is true for $R_2$. As long as the coefficients of other intermediate nodes (except the sink and sink nodes) are randomly selected from a large enough finite field and the sink and sink nodes receive the full rank of the coefficient matrix of the equation, the sink and sink nodes can solve the corresponding data through the simultaneous equations, which is called the Gauss elimination method. It can be seen from the above that the idea of random network coding is concise, easy to implement, and can also improve network throughput. These are the principles of random network coding.

3. Personal Information Protection Based on Network Coding

3.1 Scheme Description

With the rapid development of Internet technology, the Internet has become an indispensable part of life. People get information, collect information, social entertainment on the Internet. Wi-Fi, a wireless network, allows users to browse web information conveniently at anytime and anywhere. It is a cost-effective and fast way to access the Internet. But Wi-Fi, which connects public places, also hides the risk of information being stolen and tampered with. Public Wi-Fi has security risks. People with ulterior motives can eavesdrop on user information through public networks. This paper studies a method of information encryption using linear network coding, which can decode automatically on the server side and prevent the risk of information eavesdropping during transmission.

In this method, each user's information is regarded as a node, all nodes automatically form a topological network structure, the edges between nodes are given weights, weights are assigned according to the distance between users, and a tree is formed according to the minimum spanning tree algorithm. When the server requests to query a user's information, the user sends the information to the parent node after linear encoding. The user who receives the information will reorganize the received information together with his own information and send it to the parent node after random linear network encoding. The root node knows to send all the linear combination to the server and the server itself decode.

3.2 Algorithmic Description

(1) Random selection of a node as the root node, according to the prim algorithm to generate a tree (select the smallest edge to join the tree), if there are $N$ nodes, then give each node a random number in $0...N-1$ as the ordinal number of the node, and the parent node needs to record the ordinal number assigned to the child node.

(2) If the ordinal number of a node component is $i$, its data is recorded as $m_i$, multiplied by a non-zero random number $a_i$, and the encoding data $m_i$ is obtained. The encoding coefficient is written before the encoding data, and the encoding coefficient is:

$$M_i = [0, 0 ..., a_i, ..., 0]$$  \hspace{1cm} (3)

Formula: $M_i$ is a row vector, except column $i$ component is $a_i$, the other components are all 0.

(3) Suppose that a node has $m$ descendant nodes. When it receives all $m$ data and adds its own data, a random matrix of $(m+1)*(m+1)$ is used to linearly combine the $m+1$ data, and then the newly generated $m+1$ data is sent to its parent node.

(4) When the root node completes the linear mixing of $n$ user data, it sends it to the server.

(5) Server decoding. After the server receives enough data, it decodes the data using Gauss elimination method. According to the decoding result, the serial number of these request files can be obtained. The server sends the serial number directly to the root node, which arrives at the node where
the request is made and returns the information of the node according to the sequence number information of the child node stored by each parent node.

4. Performance analysis

4.1 Algorithmic complexity evaluation
In this method, the author firstly uses Prim algorithm to generate a tree among user nodes. This algorithm searches for the minimum spanning tree in weighted connected graphs. It means that the tree constructed by the edge subset searched by the algorithm includes not only all vertices in the connected graph, but also the smallest sum of weights of all edges. The details could be seen as following:

| Prim algorithm | Total time complexity |
|----------------|-----------------------|
| Adjacency matrix | O(V^2) |
| Binary heap | O((V+E)log(V))=O(E log(V)) |
| Adjacency table | O(E + V log(V)) |

Table 1. Prim complexity analysis

Look through all the algorithm, the author generates a Prim tree, and then assigns numbers to each node. When transfer the numbers to next nodes, linear operations are made among this number. The author analyzes the time and space complexity about our algorithm as following:

| Complexity | Generate forwarding tree | Assign number | User node encode | Server decode |
|-------------|--------------------------|---------------|-----------------|--------------|
| Time        | O(n^2)                  | O(n)           | O(n^2)          | O(n^2)       |
| Space       | O(n)                    | O(n)           | O(n^2)          | O(n^2)       |

Table 2. User information encoding based network coding complexity analysis

4.2 Data forwarding and transmission performance analysis
In this method, the determination of edge weights is determined by the distance between straight lines.

<ds, obj (0)> = (x1-x2)^2 + (y1-y2)^2  \tag{4}

This paper uses computational weights and spanning trees to determine the next transmission object, which improves the transmission efficiency and reduces the forwarding delay. Using random linear network coding, the coding coefficients are randomly selected in the smallest finite field at each time of coding, which can effectively reduce the time of selecting the coefficients.

5. Conclusion
In this paper, a method of encrypting user information connected to the network in public network is proposed. Random linear network coding is used to encrypt user information. All users are weighted by distance. Minimum spanning tree is generated by prim algorithm. After linear encoding the user serial number requested to access, it is transmitted to the parent node until it is transmitted to the designated node. The node transmits all the received linear coded information to the server and decodes it on the server side. This method encrypts user information, ensures user information security, reduces forwarding delay time and improves transmission efficiency.

References
[1] Rudolf Ahlswede; Ning Cai, Shuo-Yen Robert Li, Raymond W. Yeung. Network information flow (PDF). IEEE TRANSACTIONS ON INFORMATION THEORY. 2000.
[2] Zhang H, Moura JMF, Krogh B. Estimation in sensor networks: a graph approach [C] // Proceedings of the 4th international symposium on information processing in sensor networks.: IEEE Press, 2005.
[3] Ho T, Médard M, Koetter R, et al. A random linear network coding approach to multicast[J]. IEEE Transactions on Information Theory, 2006, 52 (10):4413–4430.
[4] Jaggi S, Sanders P, Chou P A, et al. Polynomial time algorithms for multicast network code construction[J]. IEEE Transactions on Information Theory, 2005, 51(6):1973–1982.
[5] Pan Liqiang, Li Jianzhong. A missing value estimation algorithm based on multiple regression model in sensor networks [J]. Computer research and development, 2009, 46 (12): 2101-2110.

[6] Ahlswede R, Cai N, Li SY, et al. Network information flow[J]. IEEE Transactions on Information Theory, 2000, 46(4):1204−1216.

[7] Szymon Chachulski, Michael Jennings, Sachin Katti. Trading Structure for Randomness in Wireless Opportunistic Routing [C] // ACM. Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications. Kyoto, Japan: ACM, 2007: 169-180.

[8] S. Katti, H. Rahul, W. Hu and D. Katabi. Xors in the air: practical wireless network coding [C] // ACM. Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications. Pisa, Italy: ACM, 2006: 243-254.

[9] Qiming Li, John C.S. Lui, Dah-Ming Chiu. On the Security and Efficiency of Content Distribution Via Network Coding [J]. IEEE Transactions on Dependable and Secure Computing, 2012(9):211-221.