Secure Network Coding-Based Named Data Network Mutual Anonymity communication Protocol

FENG Tao[1,a], XING Fei[2,b], Lu Ye[3,c], Fang Jun Li[4,d]

School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China

[a] feng@lut.cn,[b]sosubduded@163.com,[c]luye528@126.com

Keywords: NDN;mutual anonymity; anonymous routing;network coding ;authentication mechanism

Abstract. NDN is a kind of future Internet architecture. Due to the naming data network(NDN) design introduces four privacy challenges, Many research institutions began to care about the privacy issues of NDN. In this paper, this paper are in view of the major communication privacy issues of NDN to investigate privacy protection, then put forwards more effectively anonymous transfer policy for NDN. Firstly, using network coding anonymity technology of based on information segmentation, we propose network coding-based NDN anonymity communication protocol. Secondly, we add interest package authentication mechanism in the network coding of this protocol and encrypt the coding coefficient, security of this protocol is improved by this way. Finally, we proof the proposed anonymous transfer protocol security and anonymity.

The Introduction

Focus on the problems of the challenge of mobility and content distribution that the TCP/IP network architecture faces, academic proposes a new idea of resigning architecture, which is named data network (NDN).It is one of the most popular architecture in the future Internet architecture (FIA)[1].Security mechanism is added in the NDN design, but the researchers find that NDN design faces four privacy challenges: name privacy,content privacy, signature privacy, cache privacy [2].

The solutions to the privacy challenges above are mostly for the information itself.For example,name privacy,some uses bloom filter [3] to improve the user's interest packet privacy; content privacy, Pailler homomorphism encryption system[4]or broadcast encryption mechanisms[5] is proposed to encrypt packet;signature privacy, research suggest to use the group signature, ring signature [6] and so on to solve the signature privacy. However, technology of these information encryption and signature cannot protect the communication relationship between the sender and the receiver.Therefore,in order to protect the communication relationship of sender and the receiver in NDN network,anonymity technology can be factored into an overall plan.

TCP network anonymity scheme according to anonymity message forwarding way can be divided into three types: the first is based on the single message forwarding mode,such as based on anonymous communication path-Mix[7],Tor [8]; The second is based on replicated message-based. although they can provide a higher anonymity guarantee, they incur a huge traffic and waste the bandwidth; The third is based on the information splitting[9,10]. This way mainly adopts slicing a message into multiple pieces to implement the anonymity communication, but as long as the leaking node on each disjoint paths, attackers might get communication information. To solve the problems, some research consider using the network coding method to cope with the aforementioned problems. Later someone uses encoding confusion [11] to improve the ability against the attacks of the anonymous communication, but the application scope is small. Again, some suggest the anonymous communication based on the network coding and information division[12], this way is better suitable for distributed network and wireless network.

NDN’s basic anonymity method is to use trusted anonymous proxy. Uzun first proposes NDN onion routing (NDor) [13] anonymous scheme,On this basis, the author also puts forward a more complex methods that establish on NDN infrastructure layer networks -ANDaNA [14],it is further improvement on the Tor. But ANDaNA anonymity communication network increases the encryption system cost,and make the pledge that we must use at least one unattacked node. To reduce the encryption and decryption operating costs and improve the ability to resist attacks of anonymous communication,we consider adopting mentioned network coding anonymity communication.

We find NDN routing adopts interest driven way that is similar to publish/subscribe routing MP2P.We use anonymity technology of based on network coding referencing anonymous protocol of MP2P to design NDN network anonymity protocol. Because there is flooding attack interest in NDN network,we combine SANC [15] with coding scheme for source authentication, and encrypt coding coefficient to improve the ability of resisting collision attack.

This article structure arrangement as follows: The second section introduces the preliminary knowledge:secure network coding. The third section describes in detail based on secure network coding for NDN mutual anonymity protocol. The fourth section analyze this protocol's security and anonymity. The fifth section is summarizes.

Secure Network Coding

Network coding that we use in this scheme is given. Forwarding information is operated by making use of random network coding,so security and anonymity of the system are guaranteed. we not only encode in the source node,but also the encode in intermediate node,to ensure that the information transmission is untraceability and improve the capability of resistance flow analysis attack.

The most common attack of NDN network is sending fake information. In order to solve the Interest flooding attack,we provide network coding algorithm combining SANC[15]to ensure that the information from trusted users. Before authentication. We assume user and proxy acquire share key A_{SD} by KDC. Because there is a linear relationship in the global coding vectors of network coding,the user node uses homomorphic encryption to prevent coding vectors from being wiretapped.

The operation of the user node:

The user divides original data into batches of k packets. Packet format is defined as a binary form, including the batch ID number and the data segment. Batch ID uses for intermediate node to identify the data from the different precursor but belong to the same batche.

Data segment of one packet selects from the finite field F_q ,one packet is represented by the vector of length k , In this way, a batch can be denoted as the form of matrix X :

$$X = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix} = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_k \end{bmatrix} \tag{1}$$

where k denoted the number of packet in a batch, i th row of matrix X denoted i th packet.

(1)user encoding

user left multiply encoding vector to k packets to form coded packet.

$$x_i^{(0)} = \sum_{j=0}^{k} c_{ij} B_j \tag{2}$$

where B_j denotes the ith row of matrix X , c_{ij} denotes the coefficients of original packets' random linear combination. To embed the authentication key to global coding matrix C ,particular way is: coding coefficient c_{ij} denoted as the parity mapping function:

$c_{ij} = \begin{cases} Randn \in \{2Z\}, & A_{SD}^{i} = 0 \\ Randn \in \{2Z+1\}, & A_{SD}^{i} = 1 \end{cases}$, A_{SD}^{i} is ith authentication bit, *Randn* generates random numbers according to an arbitrary distribution from Galois field. According to the above linear encoding, the source node generates k code packages:

$$\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ c_{21} & c_{22} & \cdots & c_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k1} & c_{k2} & \cdots & c_{kk} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_k \end{bmatrix} = \begin{bmatrix} c_{11}B_1 + \cdots c_{1k}B_k \\ c_{21}B_1 + \cdots c_{2k}B_k \\ \vdots \\ c_{k1}B1 + \cdots c_{kk}B_k \end{bmatrix} = \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \\ \vdots \\ x_k^{(0)} \end{bmatrix} \tag{3}$$

We use homomorphic encryption scheme to encrypt global coded vector. User randomly displaces and recalculate public key of proxy node,to get public of user. k coded vector are encryption by public keys of user. $c_i^{'}s$ the encrypted structure $c_i^{'} = E_{pki}(c_{i1}...c_{ik}) = [c_1(i)...c_k(i)]$, i is i th encoding packet. Due to the homogeneity, random linear network coding can directly operates on encrypted coding vector. Transmission of the source message is as follows:

$$\{x_i^{(0)}, c_i^{'}\}, (i = 1, \cdots, k) \tag{4}$$

Finally, the users flooding the k packets to its neighbor nodes.

(2) intermediate node encoding

As the packets traverse the network, the intermediate nodes apply a random linear combination to the packets.Meanwhile,for protecting the structure of GEV,we mix the odd number of packets,as follow:

$x^{(i)} = \begin{cases} \sum_{j=1}^{m} \beta_j x_j & \mathrm{m} \in \{2Z+1\} \\ \beta_1 x_1 + \sum_{j=1}^{m} \beta_j x_j & \mathrm{m} \in \{2Z\} \end{cases}$,where β_j is j th bit of local coding vector.Intermediate nodes to the operation of the coefficient is:

$$c^{(i)} = \beta_1 c_1^{'} + \cdots \beta_m c_m^{'} = \beta_1(c_1(1)\cdots c_k(1)) + \cdots \beta_m(c_1(m)\cdots c_k(m)) \tag{5}$$

using homomorphic properties

$$E_{pk}(a_1 \oplus a_2) = E_{pki}(a_1) \oplus E_{pkj}(a_2) \tag{6}$$

$c^{(i)}$ can be simplified to

$$c^{(i)} = E\left(\sum_{i=1}^{m} \alpha_i c_{i1} \cdots \sum_{i=1}^{m} \alpha_i c_{ik} \right) \tag{7}$$

Afterwords, forward the packets to next node.

(3) proxy node decoding

The destination node receives at least k packet. Firstly decrypt received coding coefficients using homomorphic properties, proxy node uses private key sk to decrypt the coding coefficient.Then we define an inverse mapping function $x = f^{-1}(y) = \begin{cases} 0 & y \in \{2Z\} \\ 1 & y \in \{2Z+1\} \end{cases}$,Function returns zero if the corresponding coefficient is even and returns one if the corresponding coefficient is odd.

Destination authenticates the source by checking whether decryption GEV matches the authentication key at hand.It checks whether received package is new or not by checking the rank of global coding matrix, if global coding rank is k, the destination calculates the matrix inverse and uses Gaussian elimination method to decodes the data:

$$X^{(n)} = [G, GX] \longrightarrow [I, X] \tag{8}$$

Network Coding-based Mutual Anonymity Communication Protocol for NDN(NC-MANDN)

NDN anonymous communication protocol description.

Now some researchers have introduced network coding in NDN network, and the network coding combining NDN effectively protect cache privacy. We design NDN network mutual anonymity protocol using network coding technology,the anonymity protocol not only realizes the anonymous communication between the user and content nodes, but also avoids the single point failure. NC-MANDN anonymous communication process is constitute of anonymously inquiry, anonymously publish and anonymously content transfer. NC-MANDN Implements content publishers anonymously publish content, user anonymously request content communication process.

Anonymously query issuance.

Before user send interest packet message, randomly select a router in the network as a proxy node, and select different proxy for every time require issuance. User initiates an interest query by following three phases:(1)user encodes the interest packet message; （2）proxy restore the interest packet; （3）proxy forwards the interest packet in the network.we describe the working process in figure1,where user I adopts secure network coding that we mentioned in last chapter to encode interest message and sends them to next node.These pieces are transmitted in the network until proxy node Id can get enough interest pieces to restore the original package.The proxy Id forwards the interest message to all neighbor nodes. Then the interest message is transmitted in the network by random walk mechanism until this message can reach content nod.The detailed description as follows:

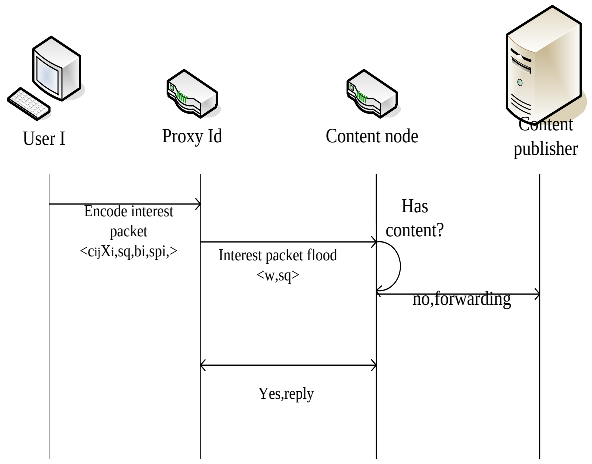

Figure3.1 Anonymously queries

(1)user I encoding interest query message

User I divides the interest message into one batch of k packets .Then the batch of k packets are encoded by adopting secure network coding mentioned in previous chapter. Before sending message,the user I creates a sequence number sq for each interest,where sq is a 160bit hash value. Meanwhile,For each coding package,I also attaches a batch identifier bi and sequence number sp. Encoded interest packet format is $\left\langle c_i^{'} x_i, sq, bi, sp_i \right\rangle$, I sends packets to its neighbor nodes. Later,the neighbor nodes in random walk way send coding packet to its immediate successor nodes.

(2)Id decoding the message

Each intermediate node adopts a random linear code to them according to the introduction of the previous section. Because user node encrypt code coefficient, so the intermediate nodes even received k pieces, also don't know the source node coefficient then can't know clear text of interest. Proxy node Id receive k linearly independent coding vector and decrypt coding coefficient using its private key sk ,then Id can decode the encoding packets and get the original interest packet.

(3)interest message flooding

Id restores the original interest packet,then flooding the message to neighbor node,the message format is $\langle I, sq \rangle$， where I is original interest packet.Interest message is sent in the network in random walk way.When finding content nodes having corresponding content,interest query process will stop.

Anonymously content publish.

Before NDN communication, content publisher need release content package to router in the network, in order router does not know who published content to it, content packet can be released anonymously in the way similar to query issuance. The detailed description as follows:

(1)content publisher encoding content packet

Content publisher adopts secure network coding mentioned in previous chapter to encode content pieces.Before sending, sq is added to encoded packet. The sq is same with sq of interest packet having same name. Meanwhile,For each coding package,content publisher also attaches a batch identifier di and sequence number sp. Encoded content packet format is $\langle c_i^{'} x_i, sq, di, sp_i \rangle$, content publisher sends packets to its neighbor nodes. Later,the neighbor nodes in random walk way send coding packet to its immediate successor nodes.

(2)router decoding the message

Each intermediate node adopts a random linear code to encoded pieces according to the introduction of the previous section. Because content publisher encrypt code coefficient, so the intermediate nodes even received k pieces, also don't know the coding coefficient then can't know clear text of content. Router receive k linearly independent coding vector and decrypt coding coefficient using its private key sk ,then router can decode the encoded content packets, get the original content packet and cache the content .

Anonymously content packet delivery.

If content node or content publishers have the content of the user requires, it is destination node R_i. Destination node R_i splits content package, carries out content confusion in accordance with the above security network coding, along the multi-path forwards to proxy node Id.Then, proxy Id according to sq records in the NDN router FIB table, the encoded pieces are forwarded along the opposite direction of interest packet to user I.The diagram below description:

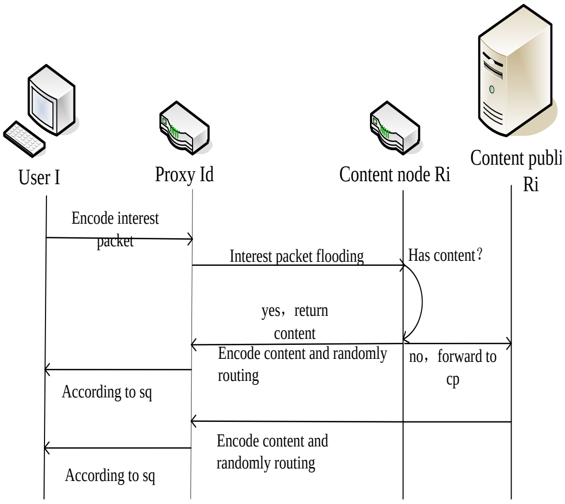

Figure3.2 Anonymously content delivery

(1)content splitting

First of all, R_i divides content package into k packets. We get a hash table recording all packets hash value. The hash table is used to check the integrity and confidentiality of the packets. Secondly, hash table and all the content packets are encoded. Finally, the sequence number sq inserted into the header of encoded packets and these packets are sent to network.

(2)Id delivery content to user

These encoded packets along multi-path are sent to the Id in randomly route. Proxy Id authenticate identity of content publisher,if the content is legitimate, proxy Id encode content packets and hash table through liner network coding. Then, Id according to the sq in FIB forwards sends these encoded packets to its neighbors. Neighbor nodes continue along the opposite path according to sq records to send packets to the user I. User I receives k hash table packets and data packets, and can get the content,then hash value of packets are compared with hash value in the table to verify the correctness. If it is not equal,user requires R_i to retransmit wrong content again.

Anonymity and security analysis

Anonymity analysis.

This section mainly analyzes this agreement's anonymity degree under different scenarios. Anonymity Degree(AD) is the definition by the probability can't be able to guess correctly the participants' (the user and the content nodes) the identity.

(1) From the outside attacker's point of view, the user' or content nodes' AD is (N - 2)/(N - 1).

(2) The node Id and I forward message by random routing. If the user is not Id's neighbor node, the identity of user's the probability of being found is 1 / (N - N_{Id}), where N_{Id} node denote Id's neighbor nodes.

(3) Under the collusion attack,user and AD of content nodes analysis

Firstly, using information entropy quantitatively analyses agreement's AD.The symbols used in the analysis formula describe as follow. x represents the malicious nodes that nearest from user, M represents the number of malicious nodes, l represents the path length between user and agent receiving coded packets from malicious node x, S represents node set receiving coded packets from the user.

$H_k(k \geq 1)$ represents the event of the first malicious nodes x on the communication path occupy k location. Then, calculate the probability $P(I \mid H_{1+})$, which is the probability that attacker can correctly guess the user's identity. Finally, get the user's AD which is probability $P(\overline{I \mid H_{1+}})$.The probability of $H_k(k \geq 1)$ is

$$\Pr[H_k] = (\frac{N-M}{N})^{k-1} \cdot \frac{M}{N} \tag{9}$$

At the same time, Assume $p_r(w)$ represents the probability that conspiracy malicious nodes receive $w(w \geq k)$ coded data packets, and p_f represents the probability that a node which is willing to forward received packets, $p_d(k)$ represents the probability that exist k nodes between the user and the first x. We concluded that attacker correctly guesses probability of user's identity, as follow:

$$\begin{aligned} P(I \mid H_{1+}) &= \sum_{k=1}^{l} (\frac{N-M}{N})^{k-1} \cdot \frac{M}{N} \cdot p_f^{k-1} \cdot p_d(k) \cdot p_e(w) \\ &\leq (1-(1-\frac{M}{N})^{l+1}) \cdot (\frac{|S|}{N-1})^{w} \end{aligned} \tag{10}$$

User's AD is:

$$\begin{aligned} P(\overline{I \mid H_{1+}}) &= 1 - P(I \mid H_{1+}) \\ &\geq 1-(1-(1-\frac{M}{N})^{l+1}) \cdot (\frac{|S|}{N-1})^{w} \end{aligned} \tag{11}$$

Content node's AD during content delivery,analysis method is similar to user's and content node's AD is:

$$\begin{aligned} P(\overline{R \mid H_{1+}}) &= 1 - P(R \mid H_{1+}) \\ &\geq 1-\left(1-(1-\frac{M}{N})^{l+1}\right) \cdot (\frac{|F_p|}{N-1})^{w} \end{aligned} \tag{12}$$

Where l represents the path length between agent node that receives coded fragmentations from malicious nodes x and content node . F_p represents node sets obtaining fragmentations from content nodes.

Security analysis.

This section we discuss several attacks of the NC - MANDN protocol, and analyses the security of the protocol.

(1) The collusion attack

Interest query stage, only when the adjacent malicious nodes through collaboration receive at least k coded packet, the attacker can guess the user correctly. Because the query phase, the data forwarding is random routing, which makes the malicious nodes collect at least k coded packet is difficult,unless there is a lot of malicious nodes in a network collaboration.

(2)The time attack

NDN’s cache attack is checking request time to detect the sensitive information of the cached data and predict the identity of the user or the content nodes. Because our agreement is operated on the logical overlay network, and forwarding in random routing. In this way, using the query time can't find the user and the true locations of the content nodes.

(3) Precursor node attack

Because in the agreement decoded interest random routing, node Id like "destination node" in the precursor attack,attackers get the identities of the user and content content nodes by node Id.While agent node Id is randomly distributed in the network, the opponent can not predict the true location of the agent.

(4) trace attack

Trace attack is attackers track direction of same sq the data flows that content nodes return. If the same sq flows point to the same node, the adversary can confirm the node is user. But in this agreement between the user and the agent node Id is random routing. Unless an opponent controls enough nodes to look back, the opponent is very difficult to learn the identity of the user.

Summary

This paper puts forwards a kind of mutual anonymous routing protocol based on network coding in NDN.In addition, we add source authentication mechanism to network coding we use to verify interest user sends. Finally, we prove security and anonymity of the protocol. But considering when there are a number of malicious nodes, anonymity degree is low, we consider combining network coding with other anonymous methods improve the anonymity.

Acknowledgements

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.

Reference

[1] Pan J, Paul S, Jain R. A survey of the research on future internet architectures. Communications Magazine, IEEE(2011), 49(7): 26-36.

[2] L. Zhang et al. Named data networking (ndn) project. Technical Report NDN-0001, October (2010).

[3] Massawe E A, Du S, Zhu H. A Scalable and Privacy-Preserving Named Data Networking Architecture Based on Bloom Filters.Distributed Computing Systems Workshops (ICDCSW), 2013 IEEE 33rd International Conference on. IEEE, (2013), 22-26.

[4] Nabeel, N. Shang, and E. Bertino. Efficient privacy preserving content based publish subscribe systems. In SACMAT ,(2012).

[5] D. Boneh, C. Gentry, and B. Waters. Collusion-resistant Broadcast Encryption with Short Ciphertexts and Private Keys. In CRYPTO,(2005).

[6] Chaabane A, De Cristofaro E, Kaafar M A, et al. Privacy in Content-Oriented Networking: Threats and Countermeasures. arXiv preprint arXiv:1211.5183, (2012).

[7] Chaum D. Untraceable Electronic Mail Return Addresses, and Digital Pseudonyms. Communications of the ACM(1981),24 (2): 84-90.

[8] R. Dingledine, N. Mathewsonn, and P. Syverson.Tor: The second-generation onion router. In The13th USENIX Security Symposium (2004).

[9] Katti S, Katabi D, Puchala K. Slicing the onion: Anonymousrouting without PKI, MIT-CSAIL-TR-2005-053. Cam-bridge: Massachusetts Institute of Technology, |(2005).

[10] Katti S, Cohen J, Katabi D.Information slicing: Anonymityusing unreliable overlays.Proceedings of the 4th USE-NIX Symposium on Network Systems Design and Imple-mentation(NSDI).New York:IEEE(2007), 43-56.

[11] Wu Zhenqiang,Ma Yalei.A Novel Anonymous communication Model:coding mix.Journal of wuhan university: science(2011),57(5)

[12] Fan K,Wei XLong D. A load-balanced route selection for network coding in wireless mesh networks. Proceedings of the 2009 IEEE international conference on Communications (ICC)(2009),5335-5340.

[13] Uzun E, DiBenedetto S, Tsudik G, et al. Anonymous Named Data Networking Application.CCNxCon, CCN Community Meeting(2011).

[14] DiBenedetto S, Gasti P, Tsudik G, et al. ANDaNA: Anonymous named data networking application. arXiv preprint arXiv:1112.2205(2011).

[15] Fathy A, ElBatt T, Youssef M. A source authentication scheme using network coding[J]. International Journal of Security and Networks(2011), 6(2): 101-111.