Ideas for Realizing Secure Low-latency Anonymity Network using Network Coding

Babatunde Ojetunde[1,a)]   Naoki Shibata[1,b)]   Juntao Gao[1,c)]

Abstract: With the advent of Internet of Things (IoT), billions of devices are not only connected to each other but constantly exchanged information. This has brought about the development of various applications that support it and are current discussion in the research community. However, security of devices and privacy are still an open issue. In this paper, we propose a new anonymous network to ensure the privacy of users or devices in a communication. First, we surveyed various works on anonymity network, then provide a brief overview of our proposed anonymous network. Specifically, our method adopts a distributed hash table (DHT) and network coding to achieve a peer-to-peer (P2P) network anonymity and secure user's identity in a communication. Each source node creates a packet with two layers, onion-like structure, the first layer is encrypted with the destination node public key while the second and outer layer is encrypted with the exit node public key. Our proposed method will guarantee the secure communication and ensure anonymity of the two parties at the end of the communication.

1. Research Background

Communication between two end-to-end users are often secured by the means of encryption or other cryptographic mechanism. While the messages being exchanged are fully secured by these methods, the identity of the users on the other hand, however, are not always hidden except when a secure private network such as TOR network is utilized. In this type of network users identities are concealed and cannot be linked to any of its online activities. Hence been an anonymous user.

Issues relating to users' anonymity have been a continuous discussion of many research works, however, most of such work has never been realized. Anonymity networks are broadly classified into two basic categories: (1) high-latency anonymity networks, (2) low-latency anonymity networks [1]. In the high-latency anonymity, proxies are used to relay messages between the source user and the destination. A proxy need to first collect messages in a batch and then add delays before the messages are sent out of the batch in a random version. Thereby preventing an attacker from being able to analyze the traffic. However, this is only effective in communication that are not affected by delays. In low-latency anonymity, messages are sent to destination through a single or multiple intermediate hops, each intermediate hop can only identify its predecessor and successor hops. Therefore, no single hop can connect the sender and the destination of the message. While high-latency anonymity can only support delay tolerant applications, low-latency anonymity supports interactive applications such as instant messaging. In addition, low-latency anonymity network are not designed to withstand traffic analysis attacks.

Despite many works on user's anonymity, problems such as identity theft, user impersonation, DoS attack, and so on still persist. In this paper, we survey various works on anonymity network and propose a method to achieve a new anonymous network. Specifically, we adopt a distributed hash table (DHT) and network coding method to secure user's identity in a communication. The source node randomly selects four nodes from the network to form its path to destination. The information of nodes in the network can be accessed through a DHT, which will adopt to maintain a decentralized node information management. The source node then selects two nodes to serve as an entry and exit nodes from the chosen nodes. After which the source node forms a two layer packet, the first layer is formed as the main packet that is to be sent to the destination node. This packet is encrypted with the public key of the destination node and also contains the contact information of the destination node. The second layer packet is encrypted with the public key of the exit node. The source node sends the message to the entry node which apply a network coding to the message and broadcast the coded message to other nodes on the selected path. Then each node on the path after receiving the broadcast packet, first decode the message and check to see if the packet is intended for them. The exit node will be able to decrypt the outer layer of the message to get the contact information of the destination node and relay the packet accordingly.

The contact information of the source node is not known to other nodes on the path except for the entry node while the contact information of the destination node is only known to the exit node alone. Each node on the selected path cannot link the message to both the source node and the destination node. Therefore,

[1] Nara Institute of Science and Technology
[a)] ojetunde.babatunde.nq3@is.naist.jp
[b)] n-sibata@is.naist.jp
[c)] jtgao@is.naist.jp

our proposed scheme will guarantee the secure communication and ensure anonymity of the two parties at the end of the communication. unlike Tor network, in our method the source node has all the information about the routes and paths to use to send its packet.

2. Related Work

In this section, we review previous work on anonymity network. Mashael *et. al.* [1] surveyed the performance and security issues of the Tor network. Firstly, they classified anonymity networks into low latency and high latency anonymity network, then discussed features of the two classifications. Secondly, they explained the design features of Tor, the most common anonymity network that is widely used and its weaknesses. Then they further reviewed previous works that have been proposed to improve the performance of Tor networks based on their categorization of such works. Previous works on Tor are categorized based on the following, with the aim of improving the network: (1) reducing congestion, (2) improving router selection, (3) scalability, (4) better security, (5) reduces overhead such as communication and computational cost. Also, they point out their advantages and weaknesses in terms of anonymity, implementation and feasibility. Lastly, they discussed unresolved issues and future direction of anonymity network.

Roger *et. al.* [2] discussed the second-generation onion router which addressed the shortcomings of the original onion routing design in terms of congestion control, forward secrecy, server discovery, integrity checking, exit policies and hidden services. They further highlighted various attacks and how they design overcome such attacks. The attacks discussed are divided into (i) passive attacks such as observing user traffic pattern, observing user content, end-to-end timing correlation, website fingerprinting and so on; (ii) active attacks such as compromise keys, DoS non-observed nodes, smear attacks, replay attacks, and so on; (iii) directory attacks such as destroy directory servers, subvert the directory server and so on; and (iv) Attacks against rendezvous points. Finally, they explained open questions in low-latency anonymity network and future directions.

Similarly, Haraty *et. al.* [3] surveyed the implementation of Tor focusing on the its features, benefits and drawbacks. First, they give a brief background on the history of onion routing development starting from the Chaum mixes process [4] to the second generation onion routing. According to Haraty *et. al.*, the Chaum mixes uses a series of private and public keys trusted to a single entity to hide the sender's identity from the receiving party. The second generation onion routing: Tor on the other hand, uses three hops Tor nodes and a bridge (a bridge is a Tor node which serves as an entry node into the Tor network) to relay a packet from the source node to a destination node. The packet goes through a multi-level encryption and the Tor nodes can only encrypt/decrypt a single layer of the packet. The source node and the destination node are not part of the Tor network and only the bridge knows the identity of the source node. They further described the features of Tor and finally identify advantages and disadvantages of Tor networks. Edman *et. al.* [5] also surveyed anonymous communication systems. They explained the mainly concepts and technologies employed in anonymous networks. They highlighted the adversarial models for anonymous networks and showed the properties (e.g. capability, visibility, mobility and participation) that describe the strength of such adversary. Then they further explained the overview of anonymous networks and classified the designs for anonymous communication into high-latency and low latency anonymity. They reviewed previous works based on different designs, for example in high-latency anonymity, they highlighted works based on mixes and mix networks and showed the deployable systems such as Penet remailers, Cyberpunk remailers, and Mixminion. In low-latency on the other hand, works such as Anonymizer.com, Onion routing, PipeNet, Crowds, Tarzan, Tor etc. were described. Finally, they explained various traffic analysis attacks such as website fingerprinting, timing attacks, predecessor attacks, disclosure attacks and stated the future directions of anonymous networks.

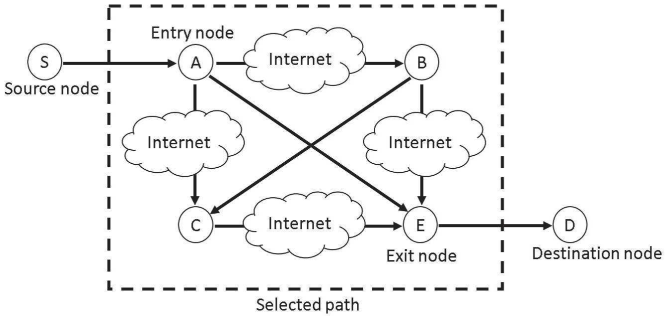


Fig. 1: Network design of proposed anonymous network

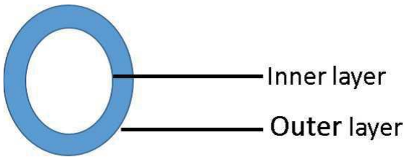


Fig. 2: Layer structure of packet

Zhanghua *et. al.* [6] proposed a secure communication with network coding based on the idea that network coding can be used to mix different data flows by an intermediate node using algebraic combinations of multiple datagrams. Then adopt confidential cryptosystem to encrypt the packet. The secret key is added as part of the cryptogram which prevents an attacker from knowing the key to decrypt the packet being forwarded to a destination node. Their scheme does not need the packet to be transmitted on a private channel.

Mittal *et. al.* [7] proposed a low-latency peer-to-peer (P2P) anonymous communication system to solve the scalability issues introduced in Tor and other systems already proposed. Their idea is based on random walk over redundant structured topologies, in which shadow nodes validate other nodes routing table. The validation is used to confirm the steps of a random walk and this helps in preventing information leak attacks. Each node in the network maintains several shadows which keep records of the node's neighbor information. The source node uses the information provided by its direct neighbor to form a route to a destination node, this information is already verified by a shadow. If the source node contact any of the shadow nodes directly, this will break its anonymity. In addition, a stabilization protocol is used to constantly ensure that information of a new node is broadcasted to other neighbor nodes. Finally, each node periodically performs a secure lookup to determine the identity of nodes they are shadowing.

Similar to our approach Chang *et. al.* [8] proposed a peer-to-peer anonymous routing based on network coding. Their scheme uses coding instead of public key infrastructure which allows the source node to anonymously send a secret message to the destination node. The source node randomly chooses a group of nodes from the network form a route which it uses to anonymously send its message. The packet contains the last-hop flag, the next-hop flag, a secret key and the packet to be forwarded. Each next-hop node on the route remove its ID in the received packet and find the sum of the packet's content to decode its message. Also, their scheme uses packet padding to maintain constant packet size which prevents an attacker from knowing the location of nodes in the route. Their scheme also adopts network coding in the data transfer phase. In addition to the normal row operation of network coding, they proposed an additional column operation where each relaying node performs a column operation by right-multiplying with a matrix which is an instruction from the source node.

According to the survey papers reviewed, some of the current issues in Tor are highlighted as follows:

- **Performance** — improving performance of the network is still an ongoing problem as the number of volunteers supporting Tor network are low (scalability issues).
- P2P approach is not easily adaptable as the Lookup process does not protect the identity of relay nodes.
- Byzantine like types of attacks are still a major concern in the earlier proposed P2P methods and Tor network.
- Incentive-based scheme proposed for Tor still faces challenges of relay nodes bandwidth usage measurement.
- **End-to-end Traffic and Timing Analysis attacks** — most solutions proposed to solve these types of attacks result in high bandwidth and latency costs. New approaches that have low bandwidth usage and latency are required.
- Tor network access blocking using deep packet inspection (DPI) requires new and improve methods that are more resistant to DPI.
- Tor network Security needs more efficient solutions and improvements.

Our contribution in this paper is the introduction of a new idea on a P2P-like approach to anonymous network. Our proposed idea ensures that the bottleneck introduced in the Tor network by utilizing centralized directory servers to manage Tor network nodes information is prevented. In our method a DHT is used by a source node to access the information of other nodes in the network. In addition, we introduce a network coding to reduce the amount of traffic in the network by ensuring that a relay node serving as an entry node applies network coding to all packets before broadcasting it to other nodes on the selected path to destination. Unlike many solutions already proposed, our method only uses two onion layers where the inner layer of the packet is encrypted with the destination node public key and the outer layer is encrypted with the exit node public key.

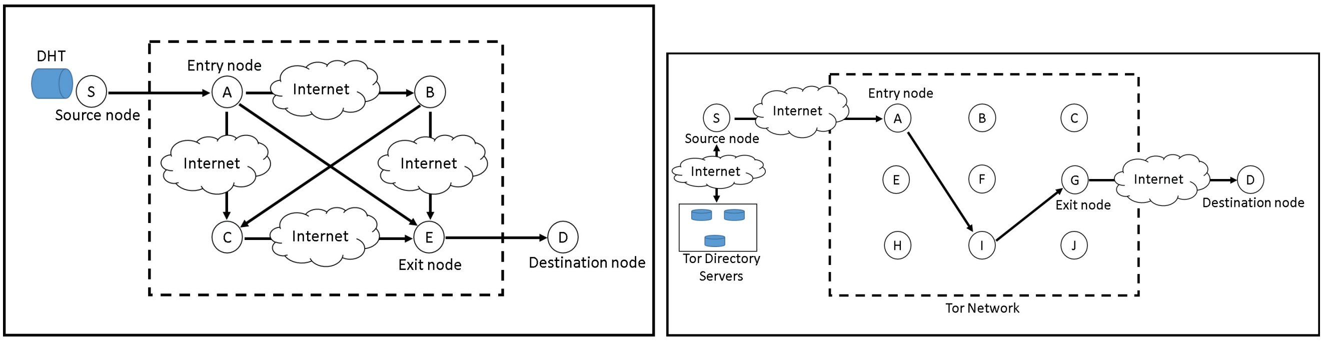
3. Ideas on a New Anonymous Communication using Network Coding

In this section we explain our idea of anonymous communication between a source node and a destination node using network coding. First, we explain how distributed hash table (DHT) is used to achieve a decentralized anonymous network, then describe how our method achieves network anonymity using network coding.

Our idea on anonymous communication leverage on the network coding approach and a DHT to achieve a peer-to-peer anonymous network. In our method, we use DHT [9] to store the contact information (such as IP address) of nodes in the network. A DHT is a decentralized system that can be used to provide a lookup of nodes in a network. According to the system, a keyspace partitioning (e.g. a keyspace is a set of all possible keys that can be used to initialize it) is used to share the ownership of the keyspace between all the nodes in the network. Also, a consistent hashing algorithm is used to map keys to nodes and each node in the network is responsible for maintaining the information that is mapped to its DHT space. There are various forms of DHT that has been developed over the years. We will focus on a slightly modified version of Chord DHT in our explanation.

To create a DHT, each node in the network needs to maintain information about other nodes such their preceding and succeeding nodes. Therefore, first we need to maintain an order of nodes. According to Chord DHT method, a random ID of k bits size is assigned to each node. Then the nodes are arranged in a ring form to set the IDs in a clockwise increasing order. The next node for each node is a node with the closest ID that is greater than the current node's ID except in the case of a node with a greater ID but its succeeding node has the smallest ID. The next step is to determine a node that is in charge of a particular key. To do this, a key and the given ID of a node is hashed to generate another key of exactly k bits. For example, let's say there is a node with ID n_1 using key k_1, to generate a key k bits we use: $k = h(n_1, k_1)$. The key k generated is used to map a node's data to the responding node's keyspace that it matches.

In our method, when a new node joins the network, it uses its IP address to generate the key. Each node key k is mapped to the DHT and distributed to nodes in the network. The key is used to locate a node that have a matching keyspace ID on the DHT and the new node can now forward its data which contains contact information of the node that can be used by other nodes in the network when choosing a path to route their packets. The data are then stored in the keyspace of the DHT. With this approach, we can achieve a decentralized management of information of nodes in the network, unlike in Tor where a central server is used to access this information. However, this approach cannot fully guarantee the source and destination nodes anonymity as relay nodes can still determine the two parties.



(a) Proposed anonymous network design

(b) Tor Network design

Fig. 3: Comparison of our proposed anonymous network to Tor design.

To achieve anonymity, we employ the use of network coding. Network coding is a technique which is used to transmit packet where a relay node merge (encoding) two packets in a single packet and the result is forwarded to a destination node. The destination node after receiving the merged packet will decode the packet using the same coding algorithm. This improves the network throughput, performance and scalability.

When a source node wants to send a packet anonymously to a destination node, the node will randomly choose four relay nodes to form its path to destination. Also, the source node selects one of the four nodes to serve as the entry node and another node to serve as the exit node. Figure 1 shows a typical network for sending an anonymous message from a source node to a destination. After selecting the entry and exit nodes, the source node creates its packet. Then encrypt the packet, and the ID of the destination node. The message and the ID of the exit node is further encrypted, as a second layer (as shown in Figure 2). Unlike in the Tor network where the source node and the destination node may not be part of the Tor network, in our method the source node and destination node are part of the anonymous network which helps improve the scalability of the network.

The source node sends the message to the entry node which then broadcast it to other nodes that are part of the four nodes selected as relay nodes. When any node on the path receives the message they check to see if they can decrypt the message. Only the exit node will be able to decrypt the message to remove the first layer of the message. After decrypting the message, the exit node can now get the information that it can use to forward the message to the destination node. The contact information of the source node is not known to other nodes on the path except for the entry node while the contact information of the destination node is only known to the exit node alone. We assume that the entry node and exit node are not colluding. When an attacker cannot detect the source node and the destination nodes as the message is broadcast by the entry node to other nodes. In a situation where the packet is captured by an attacker, the attacker will not be able to guess the exit and destination keys to decrypt the first and second layer of the message.

Using the flooding approach to send the message in our anonymous network increase the amount of traffic in the network, therefore we employ the use of network coding to suppress the message size. According to this method, each entry node will apply network coding to the message received from the source node before broadcasting it to other nodes on the path. This will improve the overall throughput of our anonymous network. We assume that the bandwidth usage in our anonymous network is always constant. Figure 3 shows the network design of our proposed anonymous network versus the Tor network design.

Our approach differs to Tor and other previous works on the anonymous network with the following:

(1) **No Central authority** — Unlike Tor network in which nodes depend on central directory servers to download information of available nodes in Tor network, which is used to randomly select relay nodes. Our method does not depend on any central server or authority. Each node has access the all nodes information through the DHT, thereby preventing the bottleneck that may occur in Tor network when the directory servers are blocked.

(2) **Two onion layers** — Our method uses only two onion layers.

(3) **Scalability** — Our method adopts a P2P approach to improve the scalability issues in Tor.

(4) **Bandwidth Usage** — Bandwidth usage is still an ongoing issue in most of the previous solutions, our method focuses on keeping the bandwidth usage constant.

(5) **Attack Model** — Tor network does not assume a strong attack adversary while our method can protect against strong adversary.

Acknowledgment

This work is supported by the JSPS KAKENHI Grant Numbers 16K01288, 16K12421, 15K15981 and Otsuka Foundation scholarship reference number 16-S38.

References

[1] Mashael Alsabah and Ian Goldberg. Performance and Security Improvements for Tor: A Survey. *ACM Comput. Surv.*, Vol. 49, No. 2, pp. 1-36, September 2016.

[2] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: the second-generation onion router. In Proceedings of the 13th conference on USENIX Security Symposium *(SSYM'04)*, Vol. 13. USENIX Association, Berkeley, CA, USA, pp. 21-21, 2004.

[3] R. A. Haraty and B. Zantout. The TOR data communication system: A survey, 2014 IEEE Symposium on Computers and Communications *(ISCC)*, Funchal, pp. 1-6, 2014.

[4] David L. Chaum. Untraceable electronic mail, return addresses, and

digital pseudonyms. *Commun. ACM* Vol. 24, No. 2, pp. 84-90, February 1981.

[5] Matthew Edman and Blent Yener. On anonymity in an electronic society: A survey of anonymous communication systems. *ACM Comput. Surv.* Vol. 42, No. 1, Article 5, pp. 1-35, December 2009.

[6] Zhanghua C., Yuansheng T., and Jinquan L., Secure communication with network coding, In Proceedings of International Conference on Applied Physics and Industrial Engineering Vol. 24, pp. 1943 - 1950, 2012.

[7] Prateek Mittal and Nikita Borisov. ShadowWalker: peer-to-peer anonymous communication using redundant structured topologies. In Proceedings of the 16th ACM conference on Computer and communications security *(CCS '09). ACM*, New York, NY, USA, pp. 161-172, 2009.

[8] C. S. Chang, T. Ho and M. Effros, Peer-to-peer anonymous networking using coding, 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, pp. 525-532, 2012.

[9] Hao Zhang, Yonggang Wen, Haiyong Xie, and Nenghai Yu, Distributed Hash Table: Theory, Platforms and Applications, SpringerBriefs in Computer Science, Springer, 2013.