ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)

https://doi.org/10.17559/TV-20200313124146
Original scientific paper

# An Anonymous System Based on Random Virtual Proxy Mutation

Rongbo ZHANG, Jibin NIU, Xin LI*, Shanzhi CHEN

**Abstract:** Anonymous systems are usually used to protect users' privacy in network communication. However, even in the low-latency Tor system, it is accompanied by network communication performance degradation, which makes users have to give up using the anonymity system in many applications. Therefore, we propose a novel anonymity system with rotated multi-path accompanying virtual proxy mutation for data transmission. Unlike onion routing, in our system the randomly generated virtual proxies take over the address isolation executing directly on the network layer and expand the anonymity space to all terminals in the network. With the optimal algorithm of selecting the path, the network communication performance improved significantly also. The verification experiments show that the anonymity system terminal sends and receives data at 500 kbps, and only a slight delay jitter occurs at the receiving end, and the other network performance is not significantly reduced.

**Keywords:** anonymous system; moving target defense; onion routing; software-defined networking

## 1 INTRODUCTION

With the development and widespread use of network technology, people rely more and more on network communication in their daily activities, which makes network attackers to possibly obtain a lot of valuable information by eavesdropping and analyzing the communication of users. Therefore, how to protect users' communication privacy from malicious adversaries becomes a strong demand. To effectively obtain the user's communication information, its basis is to obtain and accurately identify the user's communication information package. In the current packet switching network based on IP protocol, it is easy for attackers to accurately determine the two sides of the communication through the IP address contained in the communication packet head, and carry out the next step of violations based on this. At the same time, IP packets are forwarded along the route independently in the network, and the route selection algorithm is usually fixed (such as the shortest path first), which allows the attacker to leisurely select a weak node in the route to launch an attack. Therefore, the security of classical network communication is in a worrying state [1]. So, improving the privacy of network users is an important goal of network design.

Anonymous System has been proposed to protect the privacy of network users. In most cases, attackers often can recognize participants of network communication easily, which will seriously weaken the network users' privacy. So, Anonymous System does not care about the confidentiality of content, because the confidentiality of content is always the task of encryption. Anonymity System usually uses a proxy relaying mechanism to achieve anonymity, and the proxy relay is usually divided into two types: trusted and semi-trusted. In most Anonymous systems, information on packets' identification modifying and forwarding is placed on the proxies, so users communicate based on indirect style. Through address isolation of proxy relay, communication identity information can be effectively hidden to protect information from eavesdroppers [2]. At present, there are a large number of Anonymity systems implementations based on the application layer [3], such as mixes [4], onion routing [5], and crowds [6]. However, this kind of indirect communication mechanism will lead to additional costs (such as delay, payload). Therefore, the high overhead of Anonymity systems will affect communication performance and limit its use.

So far, many proxy-based Anonymous systems have been proposed, but this does not mean that we have many options to use. There are only a few systems that have been truly implemented and applied. There are many reasons for this, the number of users is an important factor and due to it often equivalent to the anonymous set which decided the effectiveness of the system. Especially in some systems, they require users to act as relay proxies simultaneously. In this way, with the increase in the number of users, the security of the system in terms of anonymity will become stronger and stronger. Unfortunately, a new system is often unattractive in its early stages. When the user group reaches a large scale, the centralized control will become a new bottleneck of the system in the classic distributed control model of network architecture. At the same time, communication complexity often leads to the poor performance of these anonymous communication systems [7]. These make some users have to give up using the anonymous system.

Moving Target Defense (MTD) is one promising security technology. In recent years, MTD has been widely applied in the software field. At the same time, to meet the needs of network security, the researches of MTD are also constantly advancing, and a few schemes have been proposed and explored in some literature [8]. Typical methods of MTD on network security include port mutation, TCP fingerprint change, IP address mutation, and so on. However, in the traditional network, the application of MTD is relatively scarce. We consider that there are two major reasons for this problem. One is complex network management, and the second is the rigidity of network configuration.

In classic networking architecture, MTD bogged down for lack of programmability and flexibility, which can be addressed effectively in Software-defined networking (SDN) [9]. SDN separates the control plane and data plane of network devices, the network devices become simple packet forwarding devices, and a logical central controller dictates overall behaviors of the network [10]. Different from the legacy network, the manager will be more flexible and agile at management in SDN. With SDN, we can introduce an innovation to the network agilely, and the complex networking applications will be implemented easily [11]. Meanwhile, the centralized global network state view can facilitate adding new features to the network in the form of networking protocols.

In this paper, adopting SDN architecture, we presented a novel anonymous system based on random virtual proxy mutation (RVPM) in SDNs. RVPM can construct an anonymous communication connection between the communication pairs, which has the following features. First of all, in the process of a session, routing is dynamic. Second, the three IP address translations are processed respectively at three proxies along the routing path. As a result, the identity of packets will change continually. Thirdly, relay proxies are randomly selected, and the proxies are virtual rather than real. The virtual relay proxy which consists of the assigned IP address and some resources on the specified switch is not a real terminal in the networks. Therefore, the virtual relay proxy can mutate with minimal operational overhead. Thus, RVPM can effectively hide the association between the IP address contained in the delivery packet and the communication terminal in the networks. The main contributions of the RVPM are as follows:

1) Decoupling the association between system anonymity and the number of users. Users of the anonymous system do not need to play the role of the proxies for obfuscating the attackers in RVPM. Through the application on the controller, the system can select the switch as the virtual proxy image in a larger range; the forwarding switches on the routing path are all in the optional set to enlarge the anonymity of the system.

2) Three-phase IP address translation architecture is adopted for anonymous communication. The virtual relay proxy needs to replace the identity of the packet and send it. The packet does not contain source route information like onion routing, so virtual relay proxy does not need to decrypt the packet and reassemble it to send the packet, and the terminal host does not need to encrypt the routing information in the packet header layer by layer.

3) The packet forwarding operation can be completed in the underlying data plane without the participation of the terminal host and application layer. Compared with onion routing, the communication efficiency is higher and the delay is less.

In RVPM, we introduced a virtual proxy mechanism with the dynamic variation of routing path which will cause more control information to be downloaded from the controller to the switches in the data plane. Meanwhile, the three-phase IP address translation also needs to install extra actions in the response switches. All operations will generate overhead on the control of the network. At the same time, higher anonymity means configuring network devices at a high frequency which brings the degradation of communication performance. Therefore, an effective algorithm is needed to find a proper trade-off point among the cost, communication performance, and anonymity.

## 2 RELATED WORK

In recent years, the privacy of users in network communication gets more and more concern. As one of the main ways is to protect the privacy of network users, anonymous communication systems have been widely concerned and studied. According to the size of the communication delay, the anonymous communication systems can be broadly divided into high latency anonymity systems (such as crowds [6], mixminion [4], etc.) and low latency anonymity systems (such as onion routing [5], Tor [12], P2P [13], etc.).

The high-latency anonymous communication systems are mainly used for applications that require strong anonymity but are not sensitive to delays, such as e-mail. The main implementation of the high latency system is Mix network. However, for instant messaging, web browsing, and other network applications, the users may not tolerate an excessive communication delay and reduce their experience of users. Therefore, low-latency anonymous networking needs to be deployed to alleviate this problem. Tor is a low-latency anonymous system which has been widely deployed and used. As a second-generation onion routing network, Tor is also one of the most widely studied anonymous systems.

The main principle of Onion routing to achieve anonymity is to realize the isolation of network addresses in packet delivery through relay proxy forwarding. Based on an overlay network, Tor's network communication performance is greatly reduced [14, 15]. This performance loss makes users have to abandon the anonymous system in some applications, thus giving up privacy protection. Existing research shows that Tor's defense against some network attacks is not efficient, such as traffic analysis [16], replay attacks [17], and so on. As the current research on attacks against anonymity systems is mainly focused on the Tor systems, such as traffic analysis attacks [18, 19], this makes the security of Tor itself a new research issue. The performance problems of Tor also lead to relevant research which is devoted to improving the communication performance and security of Tor, such as [20, 21].

SDN is an innovative network architecture that has emerged in recent years [22]. Its programmability makes network management and configuration agile. These characteristics make it possible to implement some anonymity methods that are hard to achieve in classic networks. Network security research based on SDN is also one of the hot topics in current network research [23].

Many achievements of research and exploration based on the advantages of SDN have been applied to promote the development of network security [24]. For example, OF-RHM uses IP mutation between end hosts to prevent malicious scanning and worm propagation [25]. However, the research on privacy protection based on SDN is still in its infancy, and there are few achievements. For example, based on SDN architecture, Tingwei Zhu et al propose a new anonymous communication system MIC for a data center [26]. In [27], the authors introduced address mapping (IP and MAC) into the SDN-based anonymity systems. As far as we know, there is no research on the anonymous system based on the isolation of network addresses through the virtual proxy dynamically deployed in the simple multi-routing path at the network layer and enhancing the anonymity through mutation.

## 3 PRELIMINARIES

To help the reader understanding our method better, in this section we reviewed some basic styles of the anonymous systems that have been deployed in practice. Then we proposed our anonymous communication system based on SDN.

Tor using onion routing is the most famous anonymous communication system. The communication of the Tor depends on the packets forwarding on Onion Routers which constitute an overlay network. Each Onion Router is an internet terminal running applications that provide onion routing services. These terminals constitute an Onion router set for the selection of users. So, Tor is essentially an overlay network, and the links connecting Onion Router are some common internet paths. They are connected through TCP as a channel for information forwarding. To implement anonymous communication with the help of Tor, a user must install the onion routing service proxy application. These proxies listen to the local TCP connection of Tor all the time and redirect the packets that are not for it [12].

In the Tor system, users need to select some Onion routers from the list of all current active routers to construct the circuit to the destination, and subsequently the data packets are forwarded to the destination through these proxies. At the source node, the address of each forwarding proxy in the path is encapsulated in the head of each packet in turn, and the routing information is encrypted layer by layer. When a packet arrives at a proxy, the proxy decrypts one layer to obtain the address of the next proxy of the routing path, and encapsulates the remaining data in a new packet with the local IP address and the next proxy address, then forwards it to the next proxy in the route. This kind of operation makes each proxy in the route only know who sends packets to it, and whom it sends packets to, and other participants of the routing path are unknown to it. A typical Tor circuit consists of three independent proxies. Therefore, this multi-layer encryption of routing information and forwarding mechanism obscure the real source and destination IP address of the transmitting packets.

Network address translator (NAT) is widely used as a public address multiplexing technology and private network access technology. Adopting dynamic address allocation, NAT can make it impossible to know the destination to which it is connected, thereby playing a role of concealing the real participants of the communication to a certain extent.

For the Tor system, its packets delivery works at the application layer over an overlay network, and the communication link depends on the number and location of proxies in the network. The actual packet communication link length is much longer than the optimal route on the network layer, which will enlarge the propagation delay. At the same time, the anonymity space of Tor is related to the number of routing proxies active in the network. If the number of proxies is insufficient, the anonymity is poor. Under NAT, there is not much influence on the routing of packets in the network, except that the NAT server should be located at the gateway of its service network, that is, all service packets must pass through the NAT server. However, its anonymity is limited by the fixed location deployment of NAT and the shared address space it can provide.

SDN has become a paradigm widely used in the network field, and more and more enterprises and organizations adopt it in their networks. Therefore, a novel anonymous communication system based on SDN has become a problem worthy of study. In SDN architecture, the intelligence of the network is extracted from the switches in the data plane to a logically centralized controller. Therefore, the switches of the data plane are only responsible for forwarding packets according to instructions from the controller. This feature provides network developers with flexible programmability, agile management, and the global state view of the network. Therefore, based on the advantages of SDN, we proposed a novel anonymous system RVPM which has some features as follows:

1) Proxy's function is implemented on a centralized application on the controller, it is separated from the network elements and abstracted to the virtual proxy. The proxy produces some flow entries correspondingly, then these entries are downloaded into the switches for modifying and forwarding the packets;

2) A communication circuit consists of three proxies, and each proxy only knows the address of the proxy which the packets come from, and is responsible for forwarding the packets to the next proxy;

3) Do not need to be decrypted at getting the address of the next proxy, the address isolation is realized by NAT;

4) The anonymous communication links are implemented at the network layer.

All in all, based on function and execution separation, RVPM is capable of relieving the defects of the communication performance of Tor and the limited anonymous set of NAT system. The resulting system enables users to get better anonymous services.

## 4 PROXY MODEL AND ARCHITECTURE

In this section, we illustrate the virtual proxy model used in the anonymous system in detail. Then, an overview of the system architecture is given.

### 4.1 Virtual Proxy Model

In onion routing, the proxy application software running on the terminal receives the anonymity service data packet, uncovers the encrypted routing information in the packet to obtain the address of the next station, encapsulates the rest of the packet with the source address of itself and the destination address of the next station, then sends it. Therefore, the proxy realizes the address isolation of the data packet. NAT can realize the isolation of packet source address at the gateway also. In the SDN network, the same function can be achieved by configuring the flow table entries in the flow table of the forwarding device in the data plane, changing the source and destination address of the matched packets in the switch, and forwarding the modified packets to the destination.

To better describe the implementation of address isolation in SDN, there is a simple example of the following: As shown in Fig. 1, terminal D is a proxy in onion routing. When receiving anonymity service data packet $P_{BD}$ from port 2 of switch S3, D decrypts the routing information part in the packet head, obtains the address of terminal F(10.0.0.6) which is the next station, and then encapsulates the remaining part of the packet in the packet

$P_{DF}$ and sends it to switch S3. S3 receives the packet from port 2, matches the flow table entries in flow tables and sends the packet to port 3. The packet format of packets $P_{BD}$ and $P_{DF}$ is shown at the top of Fig. 2. In terminal D, packet $P_{BD}$ is converted into packet $P_{DF}$ to realize address isolation of data packet delivery.

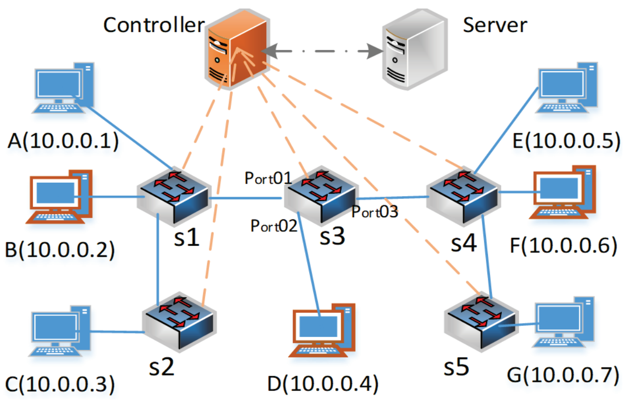
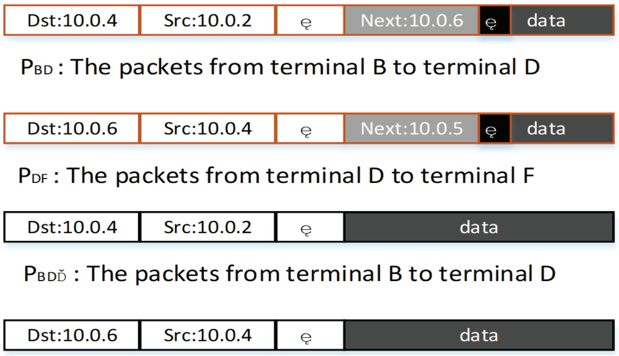


**Figure 1** An example of anonymous communication

To better describe the implementation of address isolation in SDN, there is a simple example of the following: As shown in Fig. 1, terminal D is a proxy in onion routing. When receiving anonymity service data packet $P_{BD}$ from port 2 of switch S3, D decrypts the routing information part in the packet head, obtains the address of terminal F(10.0.0.6) which is the next station, and then encapsulates the remaining part of the packet in the packet $P_{DF}$ and sends it to switch S3. S3 receives the packet from port 2, matches the flow table entries in flow tables and sends the packet to port 3. The packet format of packets $P_{BD}$ and $P_{DF}$ is shown at the top of Fig. 2. In terminal D, packet $P_{BD}$ is converted into packet $P_{DF}$ to realize address isolation of data packet delivery.



| | | | | | |
|---|---|---|---|---|---|
| Dst:10.0.4 | Src:10.0.2 | ę | Next:10.0.6 | ę | data |

$P_{BD}$ : The packets from terminal B to terminal D

| | | | | | |
|---|---|---|---|---|---|
| Dst:10.0.6 | Src:10.0.4 | ę | Next:10.0.5 | ę | data |

$P_{DF}$ : The packets from terminal D to terminal F

| | | | |
|---|---|---|---|
| Dst:10.0.4 | Src:10.0.2 | ę | data |

$P_{BD}$ŏ : The packets from terminal B to terminal D

| | | | |
|---|---|---|---|
| Dst:10.0.6 | Src:10.0.4 | ę | data |

**Figure 2** Packet changing in the address isolation

In SDN, an anonymity application service can configure the flow tables of S3 through the controller. As shown at the bottom of Fig. 2, based on installing or updating the flow table entries, $P_{BD}'$ can be directly modified to $P_{DF}'$ when S3 receives the $P_{BD}'$. Next, the $P_{DF}'$ will be sent out from port 3. The address isolation is achieved also. Specifically, the place where the packet address is changed and forwarded occurs in S3, and its function is controlled by anonymity application service, which does not require the participation of real terminal D, nor the installation of corresponding proxy software into terminal D. So, the implementation of the proxy function is called the virtual proxy model. Because no real terminal is needed to participate in address isolation, every terminal in the network can be mapped as a virtual proxy.

## 4.2 System Components

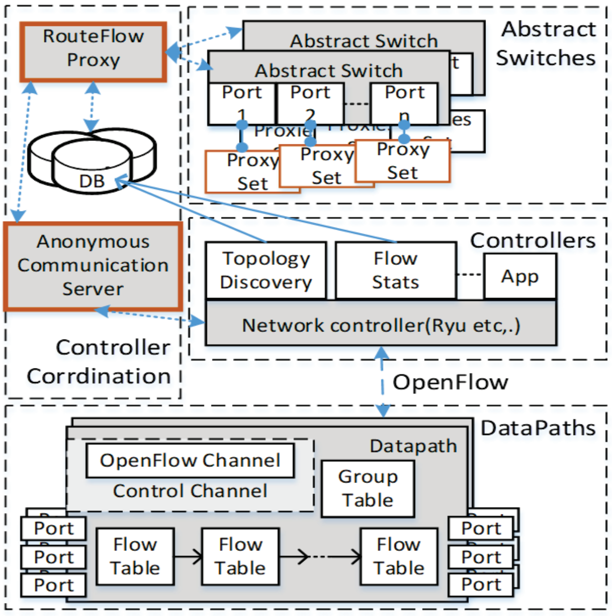


**Figure 3** Architecture of the RVMP system

There are a variety of split architectures we could use to implement the RVPM system, and the architecture design of RFCP [28] is one of the best. Therefore, we use it in the implementation of RVPM. RVPM system is an evolution from onion routing designs and is implemented in a logical centralized way based on the virtual proxy to better anonymity. It is flexible enough to accommodate different numbers of users and simplifies the deployment of ISPs. Due to no need for installing any proxy software into new users, RVPM facilitates the support of a large number of users. The system implementation is based on SDN controller, and mainly consists of the following two components (see Fig. 3):

1) RouteFlowProxy: simple "proxy" application, located on the SDN controller (such as POX, Ryu) which needs to support anonymous communication service, topology discovery, and monitoring network states.

2) Anonymous Communication Server (ACS): standalone application responsible for the core logic of the RVPM system (e.g., receiving service requests, triggering routing change events and proxy mutations, etc.). Anonymous communication server is implemented as an interactive module, using the knowledge information base to provide services for users.

## 5 OPERATION AND SCHEDULING

In the Tor system, to provide better anonymity, the circuit (anonymous communication path) can change in a session of communication. That is, the user can reselect the circuit during transmission. Typically, a circuit consists of three proxies. If a session's packets are represented as a flow, and a group of adjacent packets belonging to the same path in a session is represented as a sub-flow. That is a flow consists of several sub-flows, and the transmission route paths of any two sub-flows can be different.

Therefore, a communication session is divided into a sequence of sub-flow transmission.

In the RVPM system, to enhance anonymity, not only the routing path of flow but also the proxies in a sub-flow transmission will change continuously and randomly. Like the Tor system, three virtual proxies are introduced into a routing path to achieve three times of address isolation. The address isolation operation causes the identity (source address and destination address) of the packet to be replaced three times along the routing path during delivery. The state of virtual proxies is maintained in RoutFlow Proxy, while an anonymous communication server (ACS) is responsible for routing path variation and virtual proxy mutation. In the case of virtual proxy mutation during a sub-flow transmission, the transfer path of packets does not change. However, in the case of route changing, the virtual proxy must be updated at the same time. Based on the agility management and global network view of the SDN controller, the RVPM system realizes the effective combination of low-frequency routing change and high-speed virtual proxy change, so as to effectively restrain the burst of control resources requirement caused by the route mutation and improve the scalability of the system.

### 5.1 Three-Phase Address Isolation

In this section, we will explore three-phase address isolation with mutation of proxy and route. Its Implementation is based on OpenFlow, a widely used South interface protocol in SDN. Therefore, the RVPM can be easily deployed in the network. As shown in Fig. 3, the anonymous communication service (ACS) is responsible for accepting the anonymity service request of users in the RVPM system, and the service is based on the prior agreement of both parties (service provider and users). When a packet of a new flow of the anonymous communication accesses the switch at the network edge, the switch will send a request of flow service to the controller while there is no matching flow entry in flow tables of the switch. The controller receives the request, then the flow monitoring application will be triggered, and a user anonymous communication service request will be triggered as well. After receiving the request, the ACS will take over the transport service of the flow from the controller. The ACS obtains the current network status based on the data in the DB, calculates the routing path set for the flow, selects the current transmission path, and selects three suitable virtual agents from the Route-Flow proxy based on the current path as well.

In the RVPM system, each path of one flow has three virtual proxies which are responsible for three-phase network address isolation. We denote them virtual source proxy (vsP), virtual intermediate proxy (viP) and virtual destination proxy (vdP) respectively. The routing paths from the sender, vsP, viP, and vdP to the receptor must overlap in turn. We use $P_{SR}$, $P_{vsR}$, $P_{viR}$, and $P_{vdR}$ to represent the paths from the sender, vsP, viP, and vdP to the receptor respectively, which are calculated by the route path generation strategy. The paths satisfy the Eq. (1):

$$\left(P_{vdR} \cap PS_R\right) \subset \left(P_{viR} \cap PS_R\right) \subset \left(P_{vsR} \cap PS_R\right) \subset \left(PS_R\right) \qquad (1)$$

As shown in Fig. 4, the sender is the initiator of anonymous communication and the receptor is the destination of the communication. After the ACS has deployed a circuit (install the flow entries into the corresponding flow table in the data plane through the controller along the currently selected path), the communication is initiated. In communication, the identity of data packets sent by the sender is replaced by vsP at the first time, and for the receptor, the vdP is always used to be a substitute of the sender. viP is between vsP and vdP. viP receives packets from vsP, replaces the identity of them and sends them to vdP.

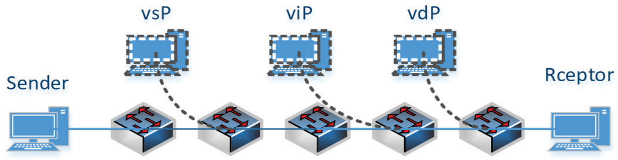


**Figure 4** An example of an anonymous communication

Unlike the Tor system, because vsP is specified by the system, the sender does not know the IP address of vsP, and the destination for the sender's communication packet is still the receptor, so the address isolation of vsP is more similar to NAT. To ensure the stability of communication, vdP cannot be changed in a session, which ensures that the receptor seems to consistent communication with vdP. In the whole communication process, vsP and viP can change continuously according to anonymous communication needs. The high-frequency hopping of VP will increase the unpredictability of communication, but at the same time, it should be noted that the hopping will bring additional control cost.

### 5.2 Control Strategy

There are two kinds of variety in the RVPM system: route varying and proxy mutation. When the route varies, the packets' path of flow changes, which may cause the change of proxies. When the route remains unchanged, the proxy can mutate independently in a sub-flow transmission. In the initial phase, the sender's flow needs anonymous communication service, the request is sent from the access switch to the ACS, and if the ACS accepts the request, an anonymous communication channel is constructed for the flow. The specific process is as follows:

1) When the ACS accepted the sender's request, it obtains the address of the sender and the receptor. Then, based on the network topology, the BFS algorithm is used to calculate and obtain multiple simple paths between the sender and receptor, and these paths are saved in a path set;

2) During the period of communication, the ACS changes the transmission path according to the network status, path set, and path usage status. The specific path update selection algorithm is detailed in Section 5;

3) After selecting the path, the ACS needs to determine three OpenFlow switches along the path according to Formula 1. The three proxies vsP, viP, and vdP are selected to satisfy the maximum anonymity;

4) The flow entries for setting the routing path and the flow entries for the three proxies to complete the address isolation are generated by ACS, then these entries are sent

to the corresponding switches along the path through the controller;

5) The sender transmits the packet to vsP. vsP receives the packet from the sender, changes the IP address in the packet header, and then forwards it to viP. Similarly, viP receives packets from vsP, then changes the IP address in the packet header and sends it to vdP. Finally, vdP receives the packet from viP, changes the IP address in the packet header, and sends it to the receptor;

6) During the period of communication, when the flow transmission attains a certain time, the routing path and the proxies will be changed by ACS. Similarly, during the period of sub-flow transmission, when reaching a certain time, the vsP and viP mutation will be triggered. The new virtual proxy is selected according to Formula 1. After a proxy changes, it needs to issue the update instructions to the corresponding switches to complete the address isolation and ensure the communication is not interrupted.

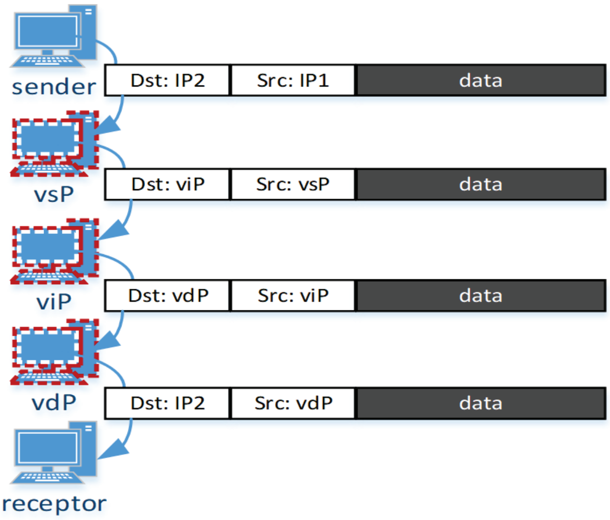


**Figure 5** A series of IP address translations for the packets

In the RVPM system, the packets traverse the network through a series of IP address translations. As shown in Fig. 5, packets from the sender (IP address is IP1) to the receiver (IP address is IP2) will experience IP replacement three times at three proxies respectively.

To issue update instructions of the RVPM system to the switches in the data plane, corresponding flow entries should be generated and downloaded from the controller to the flow table of the corresponding switch. In general, a flow entry consists of three components: match fields, counters, and a set of actions or operations. The main operations include forwarding, deleting, modifying, etc. In the RVPM system, it is necessary to modify and forward packets when address isolation is needed.

### 5.3 Communication Protocol

The sender's data flow arrives at the access switch on the network edge which will trigger the request of path service, and the anonymous request to the ACS will be triggered by the monitoring on the controller (step 1) (as shown in Fig. 6). If the ACS accepts the request, it produces and downloads the relevant update instructions to the switches along the path based on the controller (such as Ryu [29]), which includes the selection of the three switches that perform the proxy function (step 2). The functions of proxies are performed on the switches to complete the source/destination IP address conversion and packet forwarding (steps 3, 4, 5, 6, 7, 8, 9). Other switches except for proxy on the path are configured as forwarding; they are only responsible for forwarding packets according to the destination IP address. From now on, the system can match and forward (or with modify) the packets of the flow through the switches according to the flow entries in the flow table. At this time, the packets from the sender to the receptor will realize three times of address isolation to complete anonymous communication.

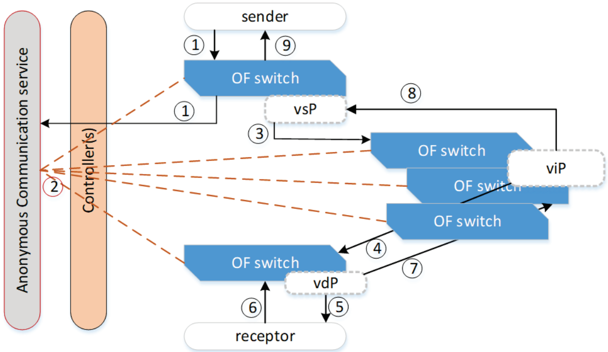


**Figure 6** Communication via anonymity service

## 6 ROUTING ALGORITHM

In the RVPM system, the ACS needs to keep some resources on the RouteFlow Proxy to maintain the communication itself. So, a large number of services means a large amount of resource overhead. Therefore, the routing algorithm adopted by the RVPM must be simple and effective.

Anonymous communication is essentially a kind of flow delivery based on special routing service. A flow is divided into many sub-flows according to a series of time slots. A time slot is one duration of a sub-flow transmission. Before a sub-flow transmitting, the ACS must choose one path for the next sub-flow and adjust the proxies based on the path.

The number of virtual proxies used in the system is different at a different time and the number of active virtual proxies determines the strength of the anonymity of communication directly. Due to selecting the proxies based on the routing path, the path selecting algorithm of the system is very important. At the same time, the anonymity of communication will be enhanced by vsP and viP constant hopping in the process of sub-flow transmission. When vsP or viP changes, the newly selected proxies should meet the Eq. (1).

Because the number of simple paths between the sender and the receptor is far less than the number of virtual proxies, and the number of switches in one path responsible for executing the proxy function is limited. Therefore, it is necessary to carefully select the path set used in the delivery and optimize its scheduling, so as to enhance its defense ability against attackers.

The data flow in the network can be divided into elephant flow and mice flow according to its size. For the mice flow, its duration is short, so the attack time window left to the attacker is also small. Therefore, here we focus on the anonymous transmission of elephant flow. To simplify the discussion and implementation, we will use the method of static time slot division to decompose an elephant flow into some small sub-flows, where the sub-flows are independent of the number of packets and only related to the length of timeslots. In other words, a sub-flow is a set of packets over a continuous period along a path.

We refer to the duration of a sub-flow as $\Delta t$, and its size is determined by the ACS. At the same time, we refer to the duration of flow as $T$. Therefore, $T$ consists of a series of $\Delta t$, i.e. $T = n \times \Delta t$. The start of sub-flow means that a new routing path should be selected and configured. So, ACS selects a new path from the path pool and issues update instructions into the data plane to complete path replacement. The simple path calculation method (such as BFS, Suurball algorithm etc.) between two points in the graph is used in the system to obtain the initial path set. The vdP is determined according to the path set, and the path $P_{vdR}$ generated from the vdP to the receptor according to the routing policy meets Eq. (2).

$$\left(P_{vdR} \cap P_{SRi}\right) \neq \Phi,\ P_{SRi} \in P \text{ and } i \in \left(1, 2, ..., n\right) \tag{2}$$

where P is the paths in the path pool which consists of all candidate paths, $P = \{P_{SR1}, P_{SR2}, \dots, P_{SRn}\}$.

In the initial phase of anonymous communication, the path pool consists of K paths. Each path has an equal initial contention window size. We refer to it as CW. In our algorithm, we set the initial value CW = 1. Therefore, each path has the same opportunity to be selected as a communication channel at the beginning of the flow transfer.

We formulate the path selection problem based on time slot $\Delta t$ as a problem that every path in the path pool strives to become the communication channel. To ensure fairness, we set each path participating in the competition as independent of each other, and each path adopts the same random competition mode, so as to achieve the randomness and fairness of selection. The idea of fair competition mainly comes from the BEB algorithm (the BEB algorithm is detailed in IEEE 802.11dcf standard). Specifically, after a path selection, CW of all paths needs to be adjusted. There are two types of adjustments. The winner's CW will be adjusted to the minimum, while the loser's CW will be doubled if it does not attain the maximum.

To further improve the fairness and randomness, collective constraint and weighted combination is a common method [30]. In the algorithm, we introduce a parameter $D$ to describe the history of path selection and refer to $D$ as a duty cycle. $D$ will participate in CW tuning to reduce the duty cycle difference for each path. The CW adjustment method with $D$ is different from the traditional BEB algorithm, which only considers adjusting the size of the contention window according to the competition results. Parameter $D$ represents the past state of the path that has transferred the sub-flows. At the same time, the threshold of $D$ is introduced: $D_h$ and $D_l$, where $D_h$ is the upper limit value and $D_l$, is the lower limit value. The duty cycle is given by Eq. (3).

$$D = \frac{P_c}{P_{sum}} \tag{3}$$

where $P_c$ represents the number of sub-flows that the path has undertaken and $P_{sum}$ represents the number of all sub-flows at the calculating time.

**Algorithm 1** Routing algorithm

```
Determine the path pool for anonymous communication
Determine the vdP of the session.
Determine the number K of the paths in the pool.
Determine the slot time Δt of the sub-flow.
for a path p in the pool do
    CW[p] = 1
end for
do
if a new Δt beginning then
        $CWsum = \sum_{p \in pool} CW[p]$;
        r = RAND(1..CWsum);
        inti;
        fori=0 to K do
            CWsum-= CW[i];
            if r >CWsumthenbreak;
end for
        $Ps$=i; //$Ps$ is the selected path in this Δt
        fori=0 to K do
            if(i == $Ps$)then
                CW = max (CWmin, [CWpre*(1−D)]);
end if
            CW = min (CWmax,[2*CWpre*(1−D)/kD]);
end for
        $T = T - \Delta t$;
end if
while ($T > 0$)
```

At the beginning of an anonymous communication. the $P_c$'s value of every path in the pool is 0. As a path has succeeded in gaining the delivery task, its $P_c$ will be increased 1. So, the $P_c$ represents the number of successes in contention of the path. After a sub-flow started delivery, the value of the $P_{sum}$ will be increased 1.

We consider, for example, a flow sends according to the time slot sequence as follows:

$\Delta t_1, \Delta t_2, \dots, \Delta t_m, \dots, \Delta t_n$

When selecting a path for $\Delta t_m$+1, it is assumed that the $D$ value of two paths is the same. One path $X$ is transferred at $\Delta t_1$, the other path $Y$ is transferred at $\Delta t_m$. Considering fairness and randomness, path $X$ should get more opportunities. Therefore, we should increase the weight of factor time. Here a first-order linear digital filter is used as follows:

$$D\left(\Delta t_m + 1\right) = \left(1 - \alpha\right) \cdot D\left(\Delta t_m + 1\right) + \alpha \cdot D\left(\Delta t_m\right) \tag{4}$$

where the factor $\alpha$ represents the effect of the past state. The value of $\alpha$ is between 0 and 1. In addition, the higher the value of $\alpha$, the more important the past state is, and a can get a proper value through big data analysis method similar to [31].

After a sub-flow starts to transport, the $CW$ of each path in the pool will be adjusted as follows:

If the path gets the delivery tasks, then

$$CW = \max\left(CW_{\min}, \left[CW_{\text{pre}} \cdot (1-D)\right]\right) \tag{5}$$

If the path loses the delivery tasks, then

$$CW = \min\left(CW_{\max}, \left[2 \cdot CW_{\text{pre}} \cdot (1-D) / K \cdot D\right]\right) \tag{6}$$

In short, the routing algorithm has two characteristics. First, each path in the set can be fair. Secondly, the correlation among the selected paths in the sequence is weak. The routing algorithm is given by algorithm 1.

## 7 SIMULATION AND EVALUATION

In this section, we introduce the experiment configuration firstly. Then, we describe the communication performance and anonymity of the RVPM system. Finally, we discuss the overhead of the system and the comparison between RVPM and Tor.

### 7.1 Experiment Configuration

To verify the feasibility of the RVPM system, we set up a simple simulation system. The experiment environment consists of three computers and a server. Two computers act as the terminals needed communication, and ACS and the SDN controller (Ryu) run on the other computer. The simulation network adopts the Internet 2 OS3E topology with 34 nodes [32], which are deployed in Mininet implemented on the server. The components of the simulation system are shown in Fig. 7.

One of the two communication terminal accesses the network by Sunnyvale nodes of the network and the other is accessed by Houston nodes. The ACS and the controller are accessed by Chicago nodes.

For evaluating the communication performance, the link delay of the network is required. With the latitude and longitude information of the nodes in the network, we can get the distance between nodes by using haverne formula [33]. Then, the propagation delay of every link is given by the Eq. (7).

$$\text{Delay} = \text{distance}(\text{m}) / 2 \times 10^8 \, (\text{m/s}) \tag{7}$$

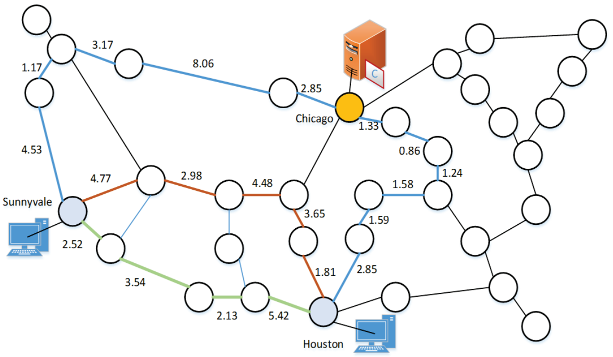


**Figure 7** Simulation network topology and components

The terminals deploy software to send UDP packets and receive the packets. In the experiment, we choose three paths to constitute the set of candidate paths, which are identified by green, red and blue respectively in Fig. 7. In the verification experiment, the terminal accessed by the network node Sunnyvale sends data to the receiving end terminal which is linked to the network node Houston at the rate of 500 kbps for 3 minutes each time.

### 7.2 Communication Performance and Anonymity

Communication performance. The RVPM system adopts a more complex routing paradigm than the common routing method. So, the communication performance in the RVPM will degrade at parameters such as delay, packet loss ratio, band width, etc. For evaluating this phenomenon, in the experiment, the path varying interval is set to 5 s, 10 s, and 20 s respectively, and the proxy mutation interval is set to 1 s, 2 s, and 3 s respectively.

Therefore, the packets are transmitted alternately along three paths to realize path various. At the same time, the source address and destination address of the packets are replaced during the transmission process. The receptor receives the data packets with the source addresses all of the selected VDP IP addresses.

The experimental data are shown in Fig. 8. During the communication process, the packets received by the receptor have delay jitter. We find that the frequency of the delay jitter and the path varying frequency are the same, which is due to the differences in the propagation delay of the three paths (green: 13.61 ms, red: 17.72 ms, blue: 29.23). The virtual proxy mutation has no significant impact on the performance of data communication.

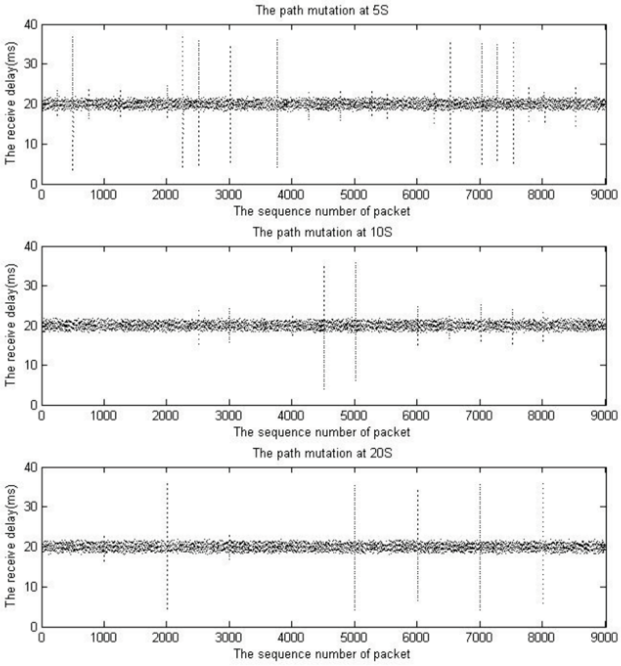


**Figure 8** The receive packet interval in the experiment

Due to no background traffic introduced in the experiment, the routing paradigm does not affect the bandwidth. At the same time, the pack loss rate has little change, the conservative strategy of varying the path and proxy is the other cause. So, the experiment is only used to verify the feasibility of RVPM system communication. In actual session communication, severe delay jitter will cause received packet to be out of order, resulting in packet loss. In severe cases, the session may be interrupted.

Therefore, when selecting the path set, pay attention to selecting the path with similar delays. When scheduling, avoid directly switching between two paths with large propagation delay differences to reduce the magnitude of delay jitter. That is also our further work.

Anonymity. In a proxy-based anonymous communication system, the number of active proxies and the number of proxies enabled in a session determine the anonymity of the system communication. In general, the presence of a large number of active agents in the network means users have more choices when communicating. For attackers, if they want to implement a successful work, they have to scan and monitor more terminals, which will increase the attack resource cost and exploring the time window, and reduce the success rate of the attack. Therefore, more active proxies in the anonymity system mean that the system can provide better anonymity services.

In the RVPM system, all active computers or terminals in the network theoretically become candidate activity proxies, thereby maximizing the anonymity space. In a specific application, in order to reduce the transmission delay in a session, the candidate path set is tailored to some simple path. At the same time, the proxy selection needs to meet the Eq. (1), and the space of the actual active proxy candidate will be reduced. Considering that the anonymity space in the RVPM system is theoretically all the active users in the network, and this reduction will bring more network transmission performance improvement, so this choice is worthwhile. At the same time, RVPM will also have a stronger ability to prevent Man-In-The-Middle attacks that may exist in the network. This ability comes from the fact that the session changes the communication path at a certain time interval, and that proxy is changed regularly in the same path during communication, which makes the attacker have to monitor more terminals under the same scan window. Therefore, the RVPM system has the ability to get the maximum anonymity, but in most cases, we need to select an optimal trade-off point between the communication performance and the anonymity.

## 7.3 Discussion

Overhead. In the Tor system, the users can freely choose proxies in the active proxy list, and the selected proxies are unknown to a third party. In the RVPM system, since the controller needs to configure the flow entries of flow tables in the corresponding switch to implement address isolation, so the selection of proxy happens in the centralized anonymity service application. At the same time, during the session, the anonymity service application also needs to maintain the path set, path usage, and proxy usage information for the next path or proxy selection, which will bring the extra cost to the server.

As shown in Fig. 9, unlike common communication, during the communication period, it is needed to perform path and proxy mutation, so the number of control packets continuously increases. The flow entries stored in the data plane have some fluctuation, but the total storage requirement is not large.

Due to the additional control overhead (the delivery of flow entries) and the degradation of communication performance (delay jitter, etc.) caused by paths and proxy varying, additional algorithms are needed to optimize these schedules. At the same time, the reliability and scalability of the RVPM system need to be studied in the next step.

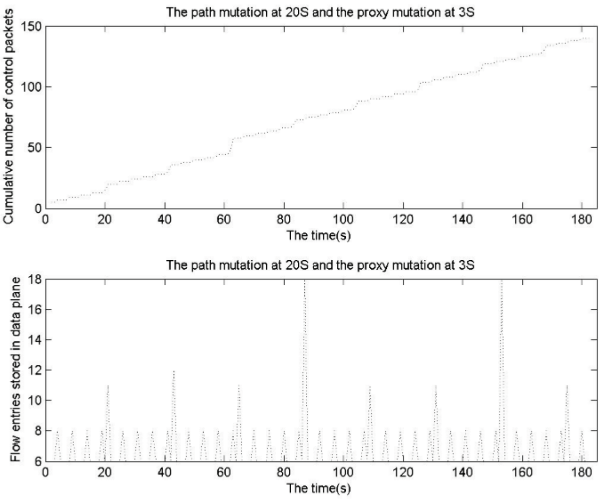


**Figure 9** The overhead of communication in the RVMP system

The comparison between RVPM and Tor. Tab. 1 shows some differences between the RVPM system and the Tor system. Tor achieves anonymous communication through address isolation. The anonymous communication achieved anonymity by performing address isolation three times on the transmission channel of a data packet. However, this system needs enough proxies in the network to form sufficiently anonymity space, which is difficult to guarantee in the initial deployment stage of the system. Insufficient anonymity space will lead to the attractiveness of the system and the loss of users, which will make the system fall into trouble. The RVPM system can bypass the trouble by adopting virtual proxy.

**Table 1** The differences between the RVPM system and the Tor system

| | RVPM | Tor |
|---|---|---|
| Anonymity space | All active terminals | The agents with proxy software |
| Deployment | Centralization | Distribution |
| Routing policy | Simple path | Path above transport layer |
| Address isolation | Located in switches | Located in agent |
| Forwarding model | Based on destination address | Based on source routing (between the source and destination) , and based on destination address (between agents) |

In the onion routing, the user is not only the user of the system but also the service provider of the system. The problem of traffic cost and resource occupation will make some users unwilling to provide more services, which will cause the network service quality to decline. At the same time, onion routing is an application layer network above the transport layer, the path for the actual transmission of data packets through the relay is much higher than the optimal path of the packets, which results in a waste of network bandwidth. Long delivery path, application layer forwarding, packets' encryption/decryption will bring greater packet delay. At the same time, onion routing requires each packet to carry routing information which results in bandwidth waste.

In the RVPM system, applying the simple routing path to forward the packets on the network layer and directly replacing IP addresses of the packets on the switches along the path will relieve the waste of bandwidth and reduce the delay of the packets as well.

## 8 CONCLUSION

In this paper, we present RVPM, a multi-path oriented virtual proxy mutation anonymous communication system. Different from the classical onion routing based on the overlay network, based on the SDN architecture, the system realizes the strategy of combining multi-path time-sharing transmission and virtual proxy mutation, which makes all active terminals in the network become candidate proxies, thereby the anonymity space of anonymity networking is maximized in theory, and at the same time, it creates a premise for the reasonable choice of multi-path.

Since there is no need for the selected agent to participate in data forwarding, users do not need to download software to act as a proxy to forward data which leads to resource cost. The multi-path rotation and virtual route mutation during communication enhance the system's resistance to traffic analysis attacks and replay attacks. Because the address isolation occurs directly on the switches in the data plane, the data delivery path is simplified, and the occupation of network resources is eased.

Although path hopping and proxy hopping will bring a certain amount of control overhead and service cost on the controller, the verification experiment shows that it has little impact on the overall performance of network communication, comparing with the anonymity and security obtained, it is worthy. Due to the lack of corresponding background traffic, our experiments are still in the simulation stage. Therefore, in the next step, we will introduce more efficient path scheduling algorithms and virtual proxy hopping based on network state awareness to improve the applicability of RVPM.

### Acknowledgment

This work was supported by research on the multi-cloud DCI oriented network metering and optimization (No.201911223, Capitalonline Data Service Co., Ltd) and the Fundamental Research Funds for the Central Universities (2014PTB-00-02)

## 9 REFERENCES

[1] Hoang, N. P. & Davar, P. (2014). Anonymous communication and its importance in social networking. *Advanced Communication Technology (ICACT), 2014 16th International Conference on*. IEEE. https://doi.org/10.1109/ICACT.2014.6778917

[2] Chakravarty, S., et al. (2015). Detection and analysis of eavesdropping in anonymous communication networks. *International Journal of Information Security*, *14*(3), 205-220. https://doi.org/10.1007/s10207-014-0256-7

[3] Ren, J., et al. (2016) Anonymous communication in overlay networks. *Security and Communication Networks*, *9*(3), 229-240. https://doi.org/10.1002/sec.539

[4] Danezis, G., Dingeldine, R., & Mathewson, N. (2003). Mixminion: design of a type III anonymous remailer protocol. *Proc. IEEE Symposium on Security and Privacy*, Berkeley, USA, 2003, 2-13. https://doi.org/10.1109/SECPRI.2003.1199323

[5] Goldschlag, D., Reed, M., & Syverson, P. (1999). Onion routing. *Communications of the ACM*, *42*(2), 39-41. https://doi.org/10.1145/293411.293443

[6] Michael, K. R. & Aviel, D. R. (1999). Anonymous web transactions with crowds. *Commun. ACM*, *42*(2), 32-38. https://doi.org/10.1145/293411.293778

[7] Rolf, J., WendoSab'ee, L. V. M. de V., & Johan, P. (2014). The fifteen year struggle of decentralizing privacy enhancing technology. *CoRR*, abs/1404.4818.

[8] Carvalho, M. & Ford, R. (2014). Moving-target defenses for computer networks. *IEEE Security & Privacy*, *12*(2), 73-76. https://doi.org/10.1109/MSP.2014.30

[9] Wolfgang, B. & Michael, M. (2014). Software-Defined Networking Using OpenFlow: Protocols,Applications and Architectural Design Choices. *Future Internet*, (6), 302-336. https://doi.org/10.3390/fi6020302

[10] Hyojoon, K. & Nick, F. G. (2013). Institute of Technology. Improving Network Management with Software Defined Networking. *IEEE Communications Magazine*, (4).

[11] Nick, M., et al. (2008). OpenFlow: enabling innovation in campus networks. *ACM SIGCOMM CCR*, *38*(2), 69-74. https://doi.org/10.1145/1355734.1355746

[12] Dingledine, R., Mathewson, N., & Syverson, P. (2004). Tor: The secondgeneration onion router. *In Proceedings of the 13th conference on USENIX Security Symposium* (13), 21-21. https://doi.org/10.21236/ADA465464

[13] Xu, C., Hua, Z,, & Qiaoyan, W. (2015). Analysis on anonymity of P2P anonymous communication system. *2015 4th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering. Atlantis Press*. https://doi.org/10.2991/icmmcce-15.2015.572

[14] Titz, O. (2018). Why TCP over TCP is a bad idea. http://sites.inka.de/bigred/devel/tcp-tcp.html. Accessed 16 Nov 2018

[15] Tschorsch, F. & Scheurmann, B. (2012). How (not) to build a transport layer for anonymity overlays. *In: Proceedings of the ACM Sigmetrics/Performance Workshop on Privacy and Anonymity for the Digital Economy*. ACM, New York, June 2012.

[16] Le Blond, S. et al. (2013). Towards efficient traffic-analysis resistant anonymity networks. *SIGCOMM*, New York, NY, USA, 2013, 303-314. https://doi.org/10.1145/2486001.2486002

[17] Pries, R., Yu, W., Fu, X., & Zhao, W. (2008). A new replay attack against anonymous communication networks. *IEEE Int. Conf. Commun. (ICC)*, 1578-1582. https://doi.org/10.1109/ICC.2008.305

[18] Meier, R., Gugelmann, D., & Vanbever, L. (2017). iTAP: In-network traffic analysis prevention using software-defined networks. *SOSR*, New York, NY, USA, 102-114. https://doi.org/10.1145/3050220.3050232

[19] Le Blond, S., Choffnes, D., Caldwell, W., Druschel, P., & Merritt, N. (2015). Herd: A scalable, traffic analysis resistant anonymity network for voip systems. *SIGCOMM*, New York, NY, USA, 639-652. https://doi.org/10.1145/2785956.2787491

[20] Hsiao, H.-C. et al. (2012). Lap: Lightweight anonymity and privacy," *in Proc. SP*, Washington, DC, USA, 506-520. https://doi.org/10.1109/SP.2012.37

[21] Snader, R. & Borisov, N. (2015). Improving security and performance in the Tor network through tunable path selection. Trans. *Depend. Secure Comput*. (85), 728-741. https://doi.org/10.1109/TDSC.2010.40

[22] White paper, (2013). Software-Defined Networking: The New Norm for Networks, Open Networking Foundation, April 13, 2012. Retrieved August 22, 2013.

[23] Danda, B. R. & Swetha, R. R. (2016). Software Defined Networking Architecture, Security and Energy Efficiency: A Survey," *IEEE Communications Surveys & Tutorials*. https://doi.org/10.1109/COMST.2016.2618874

[24] Syed, T. A., Vijay, S., Adam, R., & Sanjay, J. (2015). A Survey of Securing Networks Using Software Defined Networking. *IEEE Transactions on Reliability*, *64*(3), 1086-1097. https://doi.org/10.1109/TR.2015.2421391

[25] Jafarian, J. H., Al-Shaer, E., & Duan, Q. (2012). Openflow Random Host Mutation: Transparent Moving Target Defense Using Software Defined Networking. *In Proceedings of the First Workshop on Hot Topics in Software Defined Networks*, 127-132. https://doi.org/10.1145/2342441.2342467

[26] Tingwei, Z., Dan, F., Fang, W., Yu, H., Qingyu, S., Jiahao, L., et al. (2017). Efficient Anonymous Communication in SDN-Based Data Center Network. *IEEE/ACM Transactions on Networking*, *25*(6), 3767-3780. https://doi.org/10.1109/TNET.2017.2751616

[27] Taiyu, W., Hongyan, C., Yuepeng, S., Wenqi, L., & Tao, Yu. (2018). Anonymous network communication based on SDN. *2018 4th International Conference on Universal Village (UV).* https://doi.org/10.1109/UV.2018.8642139

[28] Rothenberg, C. E., Nascimento, M. R., Salvador, M. R., Corrêa, C. N. A., Cunha de Lucena, S., & Raszuk, R. (2012). Revisiting Routing Control Platforms with the Eyes and Muscles of Software-defined Networking. *In Proc. HotSDN*. https://doi.org/10.1145/2342441.2342445

[29] Ryu SDN Framework Community, Ryu SDN Framework. [Online]. Available: http: //osrg.github.io/ryu/

[30] Zhang, D., Sui, J., & Gong, Y. (2017). Large scale software test data generation based on collective constraint and weighted combination method. *Tehnicki vjesnik*, *24*(4), 1041-1050. https://doi.org/10.17559/TV-20170319045945

[31] Zhang, D. (2017). High-speed train control system big data analysis based on the fuzzy rdf model and uncertain reasoning. *International Journal of Computers Communications & Control*, *12*(4), 577-591. https://doi.org/10.15837/ijccc.2017.4.2914

[32] Internet2 open science, scholarship and services exchange. http://www.internet2.edu/network/ose/.

[33] Robusto, C. C. (1957). The cosine-haversine formula. *The American Mathematical Monthly*, *64*(1), 38-40. https://doi.org/10.2307/2309088

**Contact information:**

**Rongbo ZHANG,**
State Key Laboratory of Networking and Switching Technology,
University of Posts and Telecommunications,
No. 10 Xitucheng Road, Beijing, China
E-mail: zhangrongbo@bupt.edu.cn

**Jibin NIU,**
Capitalonline Data Service Co., Ltd,
No. 10 Xitucheng Road, Beijing, China
E-mail: jibin.niu@capitalonline.net

**Xin LI,**
(Corresponding author)
State Key Laboratory of Networking and Switching Technology,
University of Posts and Telecommunications,
No. 10 Xitucheng Road, Beijing, China
E-mail: cplalx@163.edu.cn

**Shanzhi CHEN,**
State Key Laboratory of Wireless Mobile Communications,
China Academy of Telecommunications Technology (CATT),
No. 40 Xueyuan Road, Beijing, China
E-mail: chensz@datanggroup.cn