A Study on Anonymous Communication

Guo Xuanzhen[a], Pan Zulie, and Shen Yi

National University of Defense Technology, Hefei 230037, China

[a]18707017967@163.com

Keywords: Anonymous Communication, P2P, TOR

Abstract: While the popularity of computer technology has brought great convenience, it has also brought great security risks. People pay more and more attention to the security of information transmission, and anonymous communication has also made great progress. Anonymous communication refers to taking certain measures to conceal the communication relationship in the communication stream, making it difficult for eavesdroppers to obtain or infer the relationship and content of both parties in the communication. The purpose of anonymous communication is to conceal the identities or communication relationships between the two communicating parties and protect the personal communication privacy of network users. This paper studies the current anonymous communication technology, summarizes the working principle of anonymous communication systems, compares the advantages and disadvantages of existing anonymous communication systems, and proposes an anonymous communication mechanism based on P2P network and DNS hidden tunnels.

1. Introduction

In this information age, the Internet is spreading to every corner of the world at an unprecedented speed. More and more people use the Internet for communication, online shopping, surfing and other activities. It can be said that the Internet is an integral part of modern life. However, people's demand for Internet security is increasing day by day, and they are more and more worried that the information they spread on the Internet, personal identity information and frequently visited webpage information are stolen by criminals. Information security is not only related to personal interests, but also to social stability, national security, and normal government operations. Therefore, with the development of computer technology, anonymous communication systems came into being.

Anonymous communication system [1] integrates data encryption, proxy nodes, and store-and-forward technology, which effectively solves the online data security problems that people generally care about, effectively hides users' online traces, resists online traffic analysis, and hides communication the identity relationship between the two parties and the information passed on. Therefore, the emergence of anonymous communication has greatly improved the security of network communication.

However, with the increase in the number of users of anonymous communication systems and the increase in transfer traffic, current anonymous communication networks have poor scalability, low operating efficiency, and security vulnerabilities. In order to effectively improve the use efficiency of anonymous communication networks and improve the security of anonymous communication networks, a practical method must be found. Nowadays, P2P distributed thinking is emerging. In the traditional client / server model, the introduction of P2P can effectively solve the bottleneck problem of the traditional client / server model, get rid of restrictions, make the entire anonymous communication network more efficient and convenient, and greatly improve the operating efficiency.

In this paper, we shed light on the importance of the use of anonymous communication. Our main objective is to provide a systematic review of the previously implemented anonymous communication systems that prevail on the Internet.

2. Definition and characteristics of anonymity

2.1 Definition

Anonymity refers to a special state of related communication entities during the network attack. More specifically, the attacking entity cannot be identified by the victim from the Animimity set [2]. That is, when an entity receives a set of information, it is impossible to determine who the sender of the information is, and anonymity is achieved.

The so-called anonymous set refers to a set of all entities that may receive or send certain network information. For example, in a certain network, the sender of anonymous information may be a1, a2, a3, a4, a5, then the set A = {a1, a2, a3, a4, a5} is a sender anonymous set. In general, the larger the anonymity set, the stronger the anonymity of the communicating entity.

2.2 Anonymous classification

Anonymity is divided into three categories: the communication anonymity between the sender and receiver (Unlinkability of sender and receiver), receiver anonymity (Receiver Anonymity), and sender anonymity (Send Anonymity) [3].

(1) Anonymous communication between sender and receiver: Because the identity of the sender and receiver may be identified by a third party during the transmission of information, the anonymity of communication makes it appear to the outside world that the sender and receiver The correlation between them is zero.

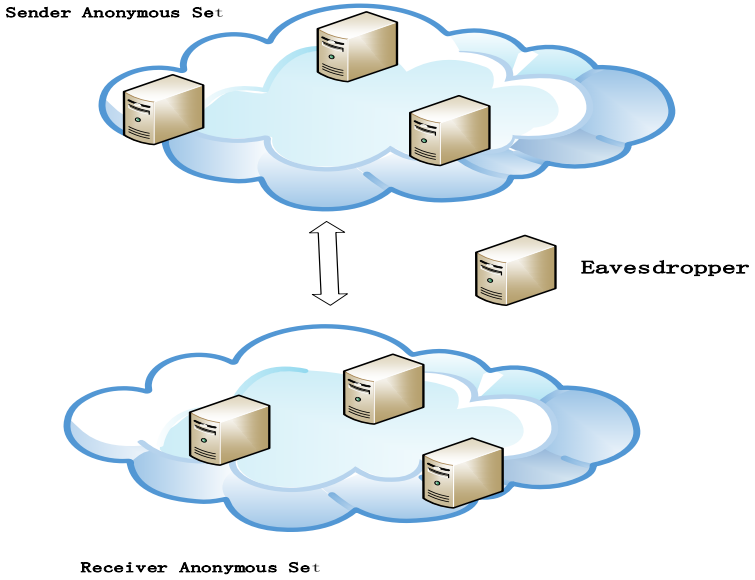

Fig1.sender and receiver anonymous set

(2) Receiver anonymity: In a certain anonymous set S, any member may be the receiver of the information m, so it cannot be determined who the specific receiver is, so it is called receiver anonymity.

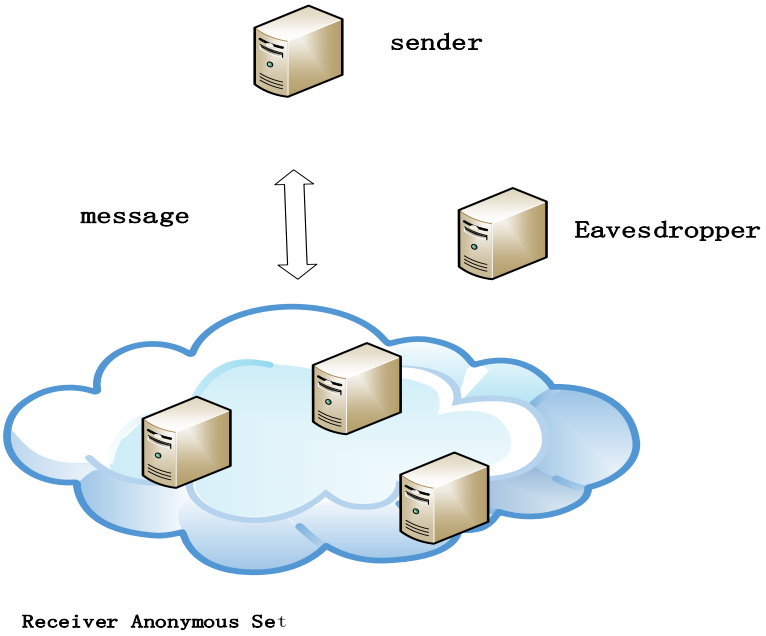

Fig2. receiver anonymous set

(3) Sender anonymity: Suppose a communication entity S sends a message M and is recognized before the information is sent. The probability of being a sender is P. After sending a message, the probability of being identified as the sender has not changed, so it is called the sender anonymity.

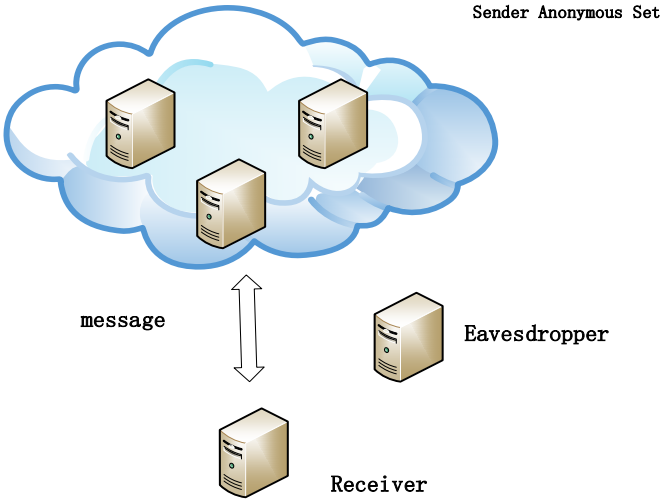

Fig3. sender anonymous set

2.3 Anonymity

Anonymity has two characteristics: undetectable and unassociable [4].

Undetectability (undetctable), that is, it is not possible to observe whether the sender of a certain anonymous set has sent information, and whether the receiver has received the information. The sending behavior of the sender and the receiving behavior of the receiver are completely hidden.

Unlinkable. Even if the observer detects the spread of data in the network, he cannot judge the degree of relevance of the obtained data through known knowledge, and cannot detect the specific identities of both parties in communication through correlation.

3. Anonymous Communication System Overview

3.1 Definition of anonymous communication system

Anonymous Communication does not study how to protect the content of the data, but how to hide the identity information of the communication entity, so that an attacker cannot obtain the user's true identity through wire tapping and traffic analysis data headers, or communicate with the user For tracking. The research purpose of anonymous communication system is to protect the user's private information, such as the IP address and communication relationship of network users, from being detected and discovered by attackers [5].

Anonymous communication does not mean to conceal the content of the communication, but to try to organize the analysis of the content of the communication to achieve the requirements of hiding identity information, hiding the communication address, and hiding the content of the communication.

3.2 The source of anonymous communication

Anonymous communication was first proposed by Chaum in 1981. In his article "Untraceable electronic mail, return addresses, and digital pseudonyms", he described the anonymous communication algorithm based on Mix nodes.

The role of the Mix node is to store and forward data packets. It can change the length and format of the message, making it impossible for third parties to infer the communication relationship. Mix can also encrypt the information, cache it, cache it to a certain amount, and re-send it to the receiver, so that the attacker cannot match the input and output of the Mix node, and thus achieve anonymity.

3.3 The significance of anonymous communication

With the increasing popularity of the Internet, network security is receiving increasing attention, and anonymity is an integral part of network security. Therefore, the anonymous communication system is of great significance to network security, which is reflected in the following points:

(1) The transaction volume of e-commerce has increased year by year, and users have set higher requirements for the concealment of the transaction process and the security of business information transmission. The introduction of anonymous communication systems into e-commerce can promote the development of Internet commerce to a certain extent and protect trade secrets.

(2) Personal privacy is the most basic human right, and it is also the focus of attention of countries around the world. However, with the rapid development of technology, cyber attack methods have emerged endlessly, and personal privacy rights have been greatly threatened. The introduction of anonymous communication mechanisms into the telephone network, e-mail and postal systems can better protect citizens' personal privacy from being violated, and communication information will not be leaked.

(3) Research on anonymous communication is related to national security information. Anonymous communication can greatly improve the national network's ability to resist cyber attacks and the covert communication capabilities of military command systems. It plays a vital role in the protection of key regions, key departments and key business networks.

3.4 Working mechanism of anonymous communication system

There are four types of anonymous communication. The first type is the rerouting mechanism, like the common TOR, and the previous anonymous system are typical representatives. The second type is broadcast multicast technology, which first encrypts the information to be transmitted, and then broadcasts it on the Internet. Only the receiver has the corresponding key to decrypt. Compared with rerouting technology, this method has greatly reduced efficiency and security. The third type is a single-agent anonymous mechanism. The fourth type is an anonymous communication system based on Mix technology.

3.4.1 Re-routing mechanism

The so-called rerouting mechanism is to send messages to the receiver through multiple intermediate routes. It can be regarded as a multi-agent communication network. Data is stored and forwarded through multi-level intermediate agents until it is transmitted to the receiver to achieve the purpose of anonymous communication.

The rerouting mechanism uses encryption technology and connection technology. The client encrypts the information to be sent layer by layer, and adjacent agents can communicate directly. For example, if there are n hosts in the set S = {Si | i = 1,2,3 ...}, a client initiates an anonymous communication request, and three routing hosts are selected as agents through a routing algorithm. Three hosts form a path with the sender and receiver.

3.4.2 Broadcast / Multicast Mechanism

In the broadcast anonymous communication network, all users send fixed-size packets to all members of a broadcast group at the same sending rate. Users without messages can send spam. In this case, the sender's anonymity and the receiver's anonymity can be guaranteed. The receiver cannot know the specific location of the sender, but only knows that the information comes from the upstream node. The sender cannot locate the specific location of the receiver. It only knows that the receiver is a member of the broadcast group.

The point-to-multipoint propagation communication mechanism on the network is called multicast communication. All possible receivers form a multicast routing tree. Each time a sender sends information, it sends information to the entire multicast routing tree. The advantage of this is that first, if the attacker detects that the receiver is in the multicast routing tree, it cannot filter the receiver from other multicast members. Second, it is quite difficult to discover this multicast routing tree from the network. Hordes system uses multicast anonymity technology.

3.4.3 Single agent mechanism

As the name implies, the single-agent mechanism uses only one agent to achieve anonymous communication. It communicates through trusted third parties. All information passed between the sender and receiver is forwarded by the agent. The advantage of this is that it effectively and simply hides the real address and achieves the effect of anonymity.

However, the disadvantages of the single-agent mechanism are: (1) it only passes through a single-level agent, which is easy to be traced by attackers; (2) the identity of the receiver and sender is not confidential to the agent, so the agent must be reliable enough trustworthy.

3.4.4 Mix mechanism

This mechanism is mainly to hide the relationship between output and input and modify the information sent. For example, by delaying and modifying timestamps, the order in which messages are sent is changed. Change the appearance of messages through padding and disguise. This makes the message untraceable.

Mix anonymous communication system mainly has two methods: point-to-point and cascade [6].

4. Common anonymous communication systems

4.1 TOR onion routing

The most widely used and most mature anonymous communication technology today is TOR (Onion Routing). TOR was originally proposed by the US Navy in 1996 and has now evolved to the second generation. The main idea is that the onion agent randomly selects the onion router to form an anonymous communication path through a routing algorithm, and encrypts and encapsulates basic information such as the router's IP address in the data packet, and only allows communication between adjacent routers. Layer decryption to achieve the effect of anonymous communication [7]. The working principle of TOR is shown in the Figure .

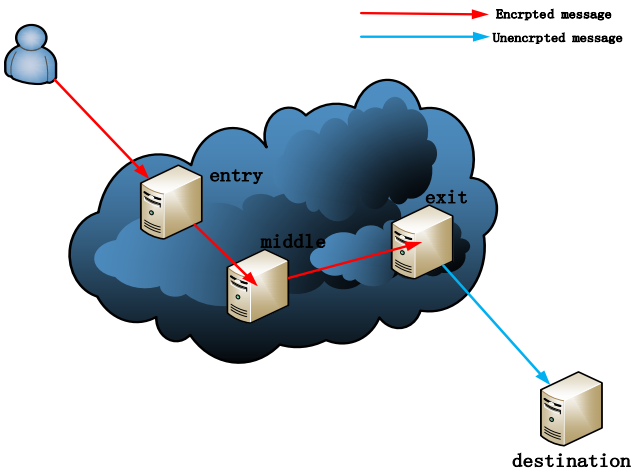

Fig4.working mechanism of TOR

All in all, TOR is a low-latency anonymous communication network with rerouting. It is easy to operate. Users only need to configure the onion proxy (OP) in the host to join the anonymous communication network. TOR has the advantages of congestion control, directory server, and security encryption [8].

4.2 Anonymizer

Anonymizer is actually a WEB agent, it uses a single agent mechanism. It can hide user IP addresses and prevent identity information from being leaked[2].

The advantages of Anonymizer are simple implementation and fast response. The corresponding disadvantage is that the anonymity is low and it is easy for the attacker to trace the source.

4.3 Crowds

The specific anonymous process of Crowds [9] is as follows:

(1) The user sends the message to any relay node in the Crowds network;

(2) The relay node chooses whether to send to the next relay or directly to the receiver;

(3) The next relay performs the second step until the message is sent to the receiver.

Crowds system can realize sender anonymity and communication process anonymity, but it cannot realize receiver anonymity.

4.4 DC-Net

DC-Net's anonymous implementation adopts a broadcast / multicast mechanism. It communicates with the receiver by broadcasting messages from a single sender[10].

The disadvantage of DC-Net is that it consumes a lot of network resources. At the same time, only one user can send messages through broadcast, which greatly reduces the efficiency of the system. In addition, due to broadcasting, each message sent will be widely spread on the network, thereby reducing security.

4.5 Freedom Network

This system, like TOR, uses a rerouting mechanism to provide multiple anonymous services. Different Freedom Network [11] unlike other anonymous communication systems provide low-level services, Anonymous communication of Freedom Network is implemented at the application layer. The advantage of this system is its strong stability, which can resist DOS attacks. The disadvantage is that the dependence on the system is too high.

4.6 DNS Tunnel

DNS is a distributed database that maps domain names to IP and is an important infrastructure of the Internet. Due to DNS packets are generally not parsed and filtered by security software, firewalls, and intrusion detection systems. Therefore, concealed communication through DNS tunnels has

unique advantages. DNS tunnel technology is mainly divided into two categories: IP over DNS and TCP over DNS[12].

IP over DNS is a transmission technology that can encapsulate IP data packets into DNS packets. Common tools are lodine, Iodine can runs on Linux, Mac OS X, FreeBSD, NetBSD, OpenBSD and Windows and needs a TUN/TAP device. The bandwidth is asymmetrical with limited upstream and up to 1 Mbit/s downstream. The tool is remarkable in higher performance, protability and less setup time. iodine uses challenge-response login secured by MD5 hash. It also filters out any packets not coming from the IP used when logging in to assure its security.

TCP over DNS is a transmission technology that can encapsulate TCP packets into DNS packets. TCP over DNS generally uses SSH port redirection technology and SOCKS proxy technology.

dns2tcp[13] This tool is based on the C language. It does not need to install additional plug-ins. It supports TXT and KEY data query. It has strong usability and applicability.

tcp-over-dns[14]. It is not only cross-platform, but also compatible with platforms such as Windows and Linux. tcp-over-dns contains a special dns server and a special dns client. The client and server work in tandem to provide a TCP (and now UDP too) tunnel through the standard DNS protocol.

This is similiar to the defunct NSTX dns tunelling software. The purpose of this software to is succeed where NSTX failed. For me at least, all NSTX tunnels disconnect within tens of seconds in real world situations. tcp-over-dns was written to be quite robust while at the same time providing acceptable bandwidth speeds.

5. P2P Network and Anonymous Communication

A peer-to-peer (P2P) network, also known as a peer-to-peer network, is a distributed application architecture that distributes tasks and workloads between peers. A peer-to-peer computer model forms a network or network application form at the application layer.[15]. It breaks the limitations of the traditional client / server model. Each node is equal in the network, both the client and the server. As shown in Figure 5.

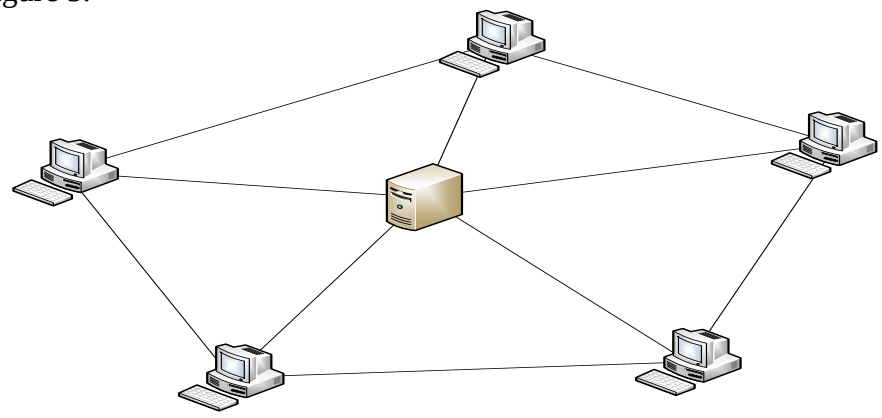

Fig5.P2P network

In the P2P network, the user host is both a provider of resources and a user of services. It can greatly improve the utilization and throughput of computer network resources, which has drawn wide attention from academia and industry. Well-known Internet companies such as Microsoft, Intel, Sony, HP and Sun have carried out corresponding basic research and formulated corresponding industry standards.

5.1 Characteristics of P2P networks

(1) Load balancing: Each node in a P2P network has equal status, which can avoid a large amount of traffic concentrated on one host, which will cause network congestion.

(2) Strong operability: hosts can be added or deleted from the network at any time, host information can be modified, and the network can be expanded. The period will not affect the normal

use of the network.

(3) Robustness: P2P networks have strong robustness. Even if they are attacked by the network or some hosts are unavailable, the network can still survive.

(4) Decentralization: The resources and services of the network are scattered on each node and do not need to be controlled centrally.

(5) Confidentiality: Because there is no need to go through intermediate links, the message is eavesdropped and the probability of being leaked is greatly reduced.

5.2 C/ S network and P2P network comparison

Client / server (C / S mode) is the most traditional and mature control mode of the Internet. This mode has relatively high requirements on the server. A high-performance server needs to process all kinds of information centrally, and respond to various requests from clients. Server-side participation is required for communication between system members. The specific process is: the client initiates a service request to the server, the server responds to the request, processes the data, and feeds the result back to the client.

Most computer system software currently uses a client / server model. It can be said that this model has greatly promoted the development of the Internet industry. It has the advantages of: improving the response speed of the client, more transparent storage of data, efficient and safe data processing.

5.3 P2P network topology

5.3.1 Centralized Topology

The advantages of this topology are convenient and fast maintenance and high resource detection efficiency. Because the central node has a complete directory, the query speed is fast, and the algorithm is found to be flexible and flexible. The biggest problem is that once the central node fails, the entire P2P network will be paralyzed, so the reliability and security are low. In addition, the larger the network size, the higher the cost of running a central directory server.

5.3.2 Fully distributed unstructured topology

The topology is applied in overlapping networks, and the organization method is random graphs. It complies with Power-law rules, has high availability, and has strong fault tolerance in dynamic networks. Can effectively support complex queries, fuzzy queries, multi-keyword queries. But the query content is not complete, the query speed is slow, consumes a lot of network bandwidth, and the problems of poor scalability cannot be ignored.

5.3.3 Fully distributed structured topology

The topology network manages the nodes in the network through a distributed hash table (DHT, full name Distributed Hash Table) [21]. Because each node can be assigned its own hash block, DHT is maintained by all nodes in the network. This structure enables nodes to join and leave the network dynamically and conveniently, with good robustness, scalability, and strong self-organizing ability.

The biggest problem is that DHT maintenance of fully distributed structured topology is more complicated. If nodes in the network frequently swap in and out, causing network fluctuations, it will greatly increase the maintenance cost of DHT.

5.3.4 Semi-distributed topology

This topology fully absorbs the advantages of fully distributed unstructured topology and centralized structure topology. It selects the nodes with excellent processing speed, bandwidth, and storage capacity as forwarding points (hubs), and information on some nodes is stored at each forwarding point. The query algorithm is only performed on each hub, and the query results will be forwarded to the corresponding leaf nodes.

The semi-distributed structure is easy to manage, has superior performance, and good scalability. But because of its heavy reliance on hubs, it is vulnerable to attacks, and its security performance is

worrying.

5.3.5 P2P topology summary

Each of the four P2P topologies has its advantages and disadvantages, and it plays a pivotal role in the entire P2P system. The centralized topology is more suitable for small networks. The fully distributed unstructured topology reflects the idea of P2P decentralization. The fully distributed structured topology introduces a new idea of hash tables. The semi-distributed topology integrates other methods. Innovate. In order to have a more intuitive understanding, the following table compares the comprehensive capabilities of the four topologies.

Table1.comprison of different topology

Comparison standard	Centralized topology	Fully distributed unstructured topology	Fully distributed structured topology	Semi-distributed topology
Scalability	Poor	poor	good	medium
Maintainability	good	good	good	medium
reliability	poor	good	good	medium

5.4 P2P network and anonymous communication

The characteristic of P2P anonymous communication system is that users are both users and proxy forwarding tasks, so all users have the same status in the anonymous communication network.

And the anonymous communication system based on P2P is realized by the rerouting mechanism. Data is forwarded through the storage and forwarding of intermediate proxy nodes and through multi-level agents. Send from sender to receiver.

5.5 P2P Network Application in Practice

Currently P2P technology is widely used in business, office, military, telecommunications and other fields. Roughly includes the following major types:

(1) File download and sharing. Users do not need to download files from the server, but from any host that has downloaded the relevant files.

(2) Cloud computing and storage sharing. Computers that join the P2P network can participate in collaborative computing with free time, and can also use idle resources to help storage.

(3) Instant messaging tools, more famous chat software such as QQ, WeChat, etc. all use peer-to-peer network technology.

(4) Online games based on P2P technology.

6. Anonymous communication technology based on P2P network and DNS tunnel

From the above analysis, it can be known that P2P networks can overcome some of the disadvantages of the ordinary C / S mode, and can achieve the balance and high distribution of nodes. The DNS tunneling technology uses most firewalls and networks to open the DNS service so that DNS packets will not be intercepted. Based on P2P network and DNS tunnel technology, it can effectively enhance the effectiveness and anonymity of anonymous networks.

7. Conclusion

This article studies the current status of anonymous communication networks and summarizes the development trends of anonymous communication networks in recent years. The concept of anonymous communication system based on P2P network and DNS covert channel is put forward, which lays the foundation for future research.

Acknowledgement

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

References

[1] Pfiyxmann,M.Kohntopp,Anonymity,Unobservabilityand Pseudonymity-A Proposal for Terminology[C]. Berkeley CA USA.2000:30-45.

[2] The Anonymizer.http://www.anonymizer.com

[3] Chen Zhili. Research Status and Development Trend of Anonymous Communication Technology [J]. Computer Knowledge and Technology.2007 Vol.1 No.1.48-53 2007.01

[4] MK Reiter,A D Rubin.Crowds:anonymity for web transaction[R].ACM Transactions on Information and System security,1998.60-90

[5] Research on Anonymous Communication Based on TOR [D]. Master's Thesis of Xidian University, 2013

[6] Diffle W. , Hellman, M.E, New Direction in Cryptography[J], IEEE Transactions on Information Theory,Vol.22,No.6,pp:74-84,1977.

[7] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router[J]. Proceedings of the 13th USENIX Security Symposium, pages 303-320, San Diego, CA, Aug. 2004.

[8] HeXv. Research on Anonymous Communication Based on TOR [D]. Xi'an University of Electronic Science and Technology, 2013

[9] MK Miller.et.al. anonymous web tran with crowds[C].COMMUNICATIONS OF THE ACM February 1999/Vol. 42, No. 2

[10] Chaum, D.L.: The Dining Cryptographers Problem: Unconditional sender and Recipient Untraceability. Journal of Cryptology, Vol.1, No.1, (1988), 65–75

[11] Goldberg, I. and A. Shostack, "Freedom Network 1.0 Architecture and Protocols," 1999.

[12] Xv Kun. Research on DNS Hidden Channel Detection Technology [D]. Master's Thesis of Southwest Jiaotong University, 2014

[13] Dns2tcp. https://github.com/alex-sector/dns2tcp

[14] Tcp-over-dns:http://analogbit.com/software/tcp-over-dns/

[15] Luo Jiewen.A review of Peer-to-Peer(P2P). http://www.huihoo.com/p2p/l