International Journal of Security and Its Applications
Vol. 10, No. 6 (2016) pp.1-12
http://dx.doi.org/10.14257/ijsia.2016.10.6.01

F-Crowds: An Anonymity Scheme for P2P File-Sharing

Tianbo Lu[1], Xinyuan Zhang[1], Xiaofeng Du[2] and Yang Li[1]

[1]School of Software Engineering, Beijing University of Posts and Telecommunications, 100876, Beijing, China
[2]School of Computer Science, Beijing University of Posts and Telecommunications, 100876, Beijing, China
lutb@bupt.edu.cn, 247958455@qq.com

Abstract

The traditional encryption methods pay more attention to the confidentiality of the message content, but ignore the protection of identity information to the communicating parties. By encrypting, the contents of the communication acquired by attackers became relatively difficult, but they can still found the identity information of the two sides to communicate easily because in the TCP / IP protocol, the sender and receiver of the message are exposed. The anonymity communication technology, is used to solve this problem. The first anonymity communication technology can date back to 1981, Chaum proposed MIX [1] which is a very important technology for reference by a number of researchers of anonymity communication. The rapid development of the Internet has made things convenience to people, but also led to more personal information exposed to the public. As a result, a growing number of areas that required anonymity communication technology to protect the user's privacy. P2P file-sharing is not exception. P2P is undoubtedly the most popular way for file-sharing, but popular P2P file-sharing protocols, such as BitTorrent, can't provide users anonymity communication service. Many people choose to build upon existing anonymous communication tools, such as Tor, to achieve anonymous P2P file transfer, but Tor primarily provides anonymous service for web browsing, instant message or other low latency applications. A large number of P2P file traffic will seriously threaten the service of Tor. Therefore, the anonymity communication service specifically for P2P file sharing system is very important. In this paper, we presents an anonymity communication scheme specifically for P2P file-sharing networks called F-Crowds which based on Crowds [2] to provide an adjustable anonymity service for P2P file-sharing.

***Keywords**: component; Crowds; anonymity system; F-Crowds; P2P; adjustable anonymity*

1. Introduction

1.1. Crowds

Crowds protocol is a peer-to-peer network with a different path selection policy from the Mix-net system. It was proposed by Bell Laboratories in 1998, which aims to protect users' privacy when they are browsing the web. The idea of Crowds is "mixed in the crowd". Users in Crowds can forward messages from other users as well as themselves, we call them Jondos. Each Jondo in Crowds organized by an central server named Blender. When a new member join in Crowds, Blender will notice all Jondos that there is a new Jondo joins us, and it also inform the new members the information of other Jondos in system.

ISSN: 1738-9976 IJSIA

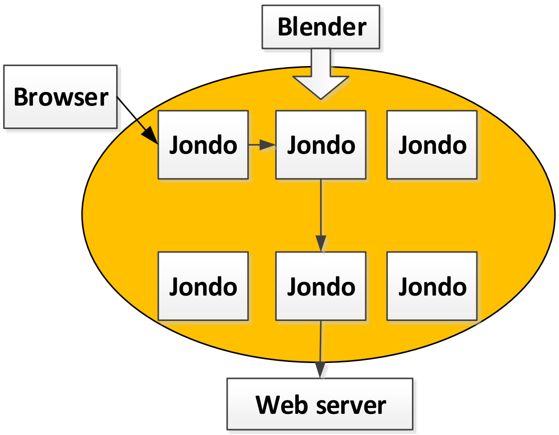

Figure 1. Structure of Crowds

When someone in Crowds visit a website, it can choose to send the request to the web server directly, or forward the request to one Jondo in Crowds. When a jondo receive the request, it does the same things that send the request to the server or forward it to one Jondo in Crowds.

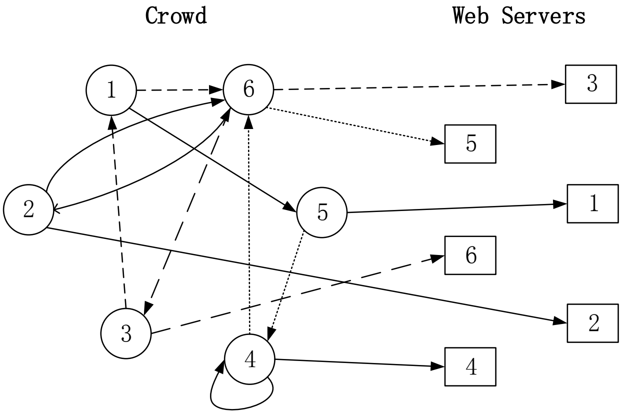

Figure 2. Example of a Path in Crowds System

The request is always sent to jondo with the probability of p_f, as sent to server with 1- p_f. The request forward by a jondo to another jondo until it is sent to server. Some possible path are shown

In "Figure 2". In this figure, the possible path are 1 → 5 → server; 2 → 6 → 2 → server; 3 → 1 → 6 → server; 4 → 4 → server; 5 → 4 → 6 → server; and 6 → 3 → server.

1.2. BitTorrent

BitTorrent is the most widely used P2P file distribution protocol, the BT network has three basic composition: Seed node, Downloader node and Tracker. Seed node is the user who owns the entire file, and Downloader node is the user who wants to download the file. Tracker server maintains a list of Downloader nodes which enable Downloader nodes interconnect between themselves and file can transfer directly between Downloader nodes. That not only reducing pressure on Seed node but also improving the performance of the entire network.

In BitTorrent, the shared file virtualization into equal-sized pieces. When we download a file, we need seed. In fact, the seed is the metafile with an extension name "torrent". The seed contains some information about files to be shared, such as file name, file size, the URL pointing to Tracker server, the index information and the hash verification code for each piece. Seed node will publish the seed to Tracker server, when someone wants to download the shared file, he need a BitTorrent client and seed. BitTorrent client will parse torrent files and get Tracker server address, then send a request to the server. Once received the request, the Tracker server will return the Seed node' address and other downloader nodes' address.

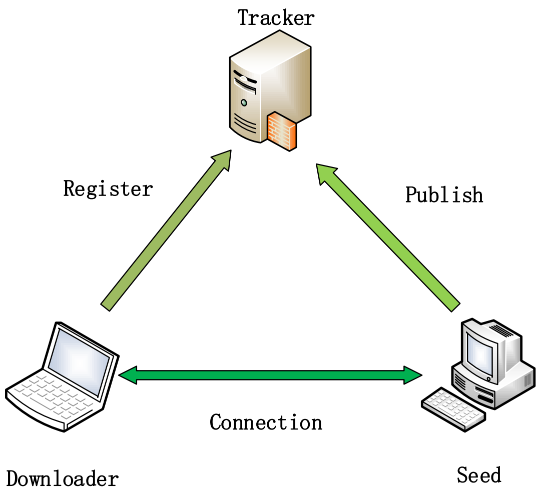

Figure 3. The Structure of BitTorrent

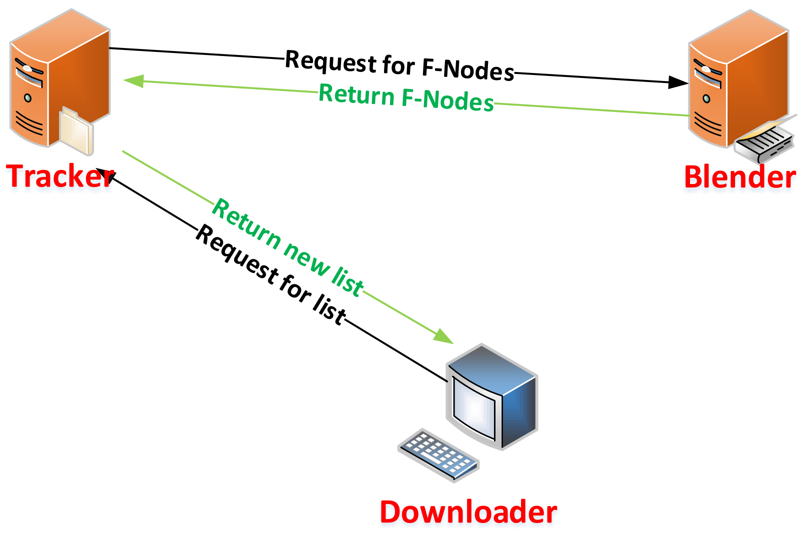

Figure 4. Require F-Node form Blender

Client will download pieces from Seed node or Download nodes, for each piece, the client will calculate the hash value of it, and compare with the hash value in torrent, if they are equal, it indicates that the piece is correct and complete, otherwise, client have to re-download the piece. If some pieces are downloaded, as long as the client does not close, the node will become the new seed of the shared files. BT network using multi-point parallel download technology, so Downloader nodes will provide upload services to other Downloader nodes at the same time. As a result, the download speed increases as the number of users increase.

BitTorrent protocol does not provide a user with anonymity service when he downloads a file. In this paper, we introduce F-Crowds which add a layer on BitTorrent, provide an adjustable anonymity for BitTorrent protocol.

2. F-NODE

F-Crowds are intended to provide an anonymous layer for BtiTorrent, so it can obtain a certain degree of anonymity, thus the peers that are listed by the tracker for a particular torrent achieve plausible deniability.

F-Crowds adopts the main ideas of Crowds, using the same forwarding strategy as Crowds. Furthermore, it allows the trade-off between anonymity and efficiency to be made by adjust the parameter. So, F-Crowds can provide users with the flexibility of anonymity and meet the different needs of different people for anonymity.

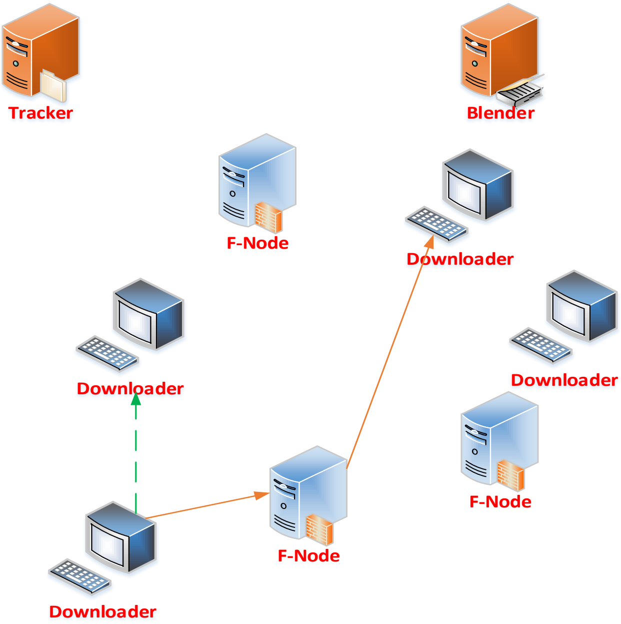

Figure 5. Download Piece

When downloader downloads files with BitTorrent client, the Tracker will return a list containing the addresses of the torrent publisher and other downloaders. And then, the downloader can contact other downloaders and exchange pieces with each other through this list. This way can greatly reduce the burden on the server, and more people download, the faster download speed can be. But in the same time, the downloaders will be exposed, thus leaked the privacy of the downloaders.

As the same with Crowds, F-Crowds also uses probability forwarding to provide anonymity. F-Crowds introduce a kind of node named F-Node, and F-node does not participate in file download, it can only forward other downloaders' requests.

In order to establish an anonymity communication, the downloader send a request to the Tracker for a list of other downloaders. After the Tracker receives the request, it requests some F-nodes from Blender rather than directly response the list to the downloader. After the tracker receives the F-nodes from Blender, it will mix them with other downloaders together into a new list, then it will send the new list to the downloader. So that when someone downloads files through the list, when he requests pieces from a normal node, the node will return the piece to him. When he requests pieces from F-Node, the F-Node simply forwards his request to other nodes. So when a node receives a request for pieces, it can’t determine who's the initiator of the request, because the request may come from a F-Node.

3. Tradeoff between Anonymity and Performance

The existing anonymity schemes in order to pursue anonymity expense to much performance. Sometimes, people may be willing to sacrifice performance in order to get a strong anonymity. But sometimes, people may want to have a good performance. Therefore, F-Crowds provide a mechanism that can make trade-off between anonymity and performance by adjusting the parameters.

We use L represent a set of degrees of anonymity. When the downloader send a request to Tracker, he will send a parameter l_k to the server which on behalf of his desired degree of anonymity.

$$L = \{l_1, l_2, \cdots, l_k, \cdots, l_m\} \quad m \geq 1$$

Where $l_1, l_2, \cdots, l_k, \cdots, l_m$ represent the degree of anonymity. The degree of anonymity gets higher as k grows. In other words, l_1 represents the lowest degree of anonymity, while l_m represents the highest.

We use P represent a set of probabilities of forwarding. In F-Crowds, when a F-Node receive a request, it simply forward it to a normal node with a probability of p_k or forward it to another F-Node with a probability of 1- p_k .

$$P = \{p_1, p_2, \cdots, p_k, \cdots, p_m\} \quad m \geq 1$$

Where $p_1, p_2, \cdots, p_k, \cdots, p_m$ represent the probability of forwarding. The probability of forwarding gets higher as k grows. In other words, p_1 represents the lowest probability while p_m represents the highest probability.

We use Q represent a set of the value for p_k .

$$Q = \{(0, v), (v, 2*v), \cdots, (v*(m-1), v*m)\}$$
$$v = \frac{1}{m}$$

The value of the probability p_k is between the interval(v*(k-1),v*k). For each request from Tracker, Blender randomly get a value q from the interval and let p_k = q.

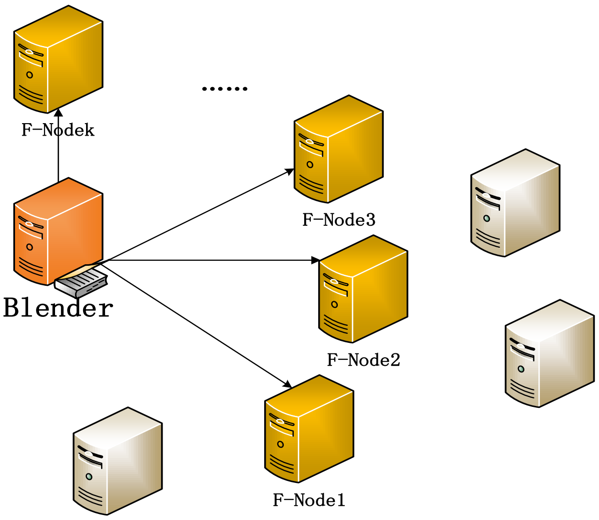

Figure 6. The F-Nodes Participating in the Anonymity Torrent

Let n be the number of peers participating in an anonymous torrent and m be the number of F-Nodes participating in the anonymous torrent. Obviously $m \cap n = \varnothing$. We use N represent a set of the number of F-Nodes.

$$N = \{n_1, n_2, \cdots, n_k, \cdots, n_m\}$$
$$m \geq 1$$

Where $n_1, n_2, \cdots, n_k, \cdots, n_m$ represent the number of F-Nodes participating in an anonymous torrent. The number of F-Nodes gets bigger as k grows. In other words, n_1 represents the smallest number of F-Nodes, while n_m represents the biggest.

The total number of F-Nodes is B. We use F represent a set of the value for n_k.

$$F = \{(0,b),(b,2*b),\cdots,(b*(m-1),m*b)\}$$
$$b = \frac{B}{m}$$

The value of the F-Nodes number n_k is between the interval (b*(k-1),b*k). For each request from Tracker, Blender randomly get a value f from the interval and let n_k = f.

When someone start an anonymous torrent, he sends the request to Tracker with an expected degree of anonymity l_k , than Tracker sends the request to the Blender. According to l_k , Blender first set p_k , n_k with the random value in(v*(k-1), v*k) and (b*(k-1),b*k). Then Blender select n_k F-Nodes from All F-Nodes randomly. Blender return all of the n_k F-Nodes and the forwarding probability p_k to Tracker. Tracker mix the F-Nodes and the normal nodes to a new list and return the new list and the received p_k to the downloader.

Upon receiving the response from Tracker, the downloader start request pieces from the nodes in the list. The request includes the parameter p_k , when a normal node receives the request, it responses to the request directly and ignores p_k, but when the request received by a F-Node, it may forward the request to a normal node with the probability 1- p_k or forward the request to other F-Nodes with the probability p_k.

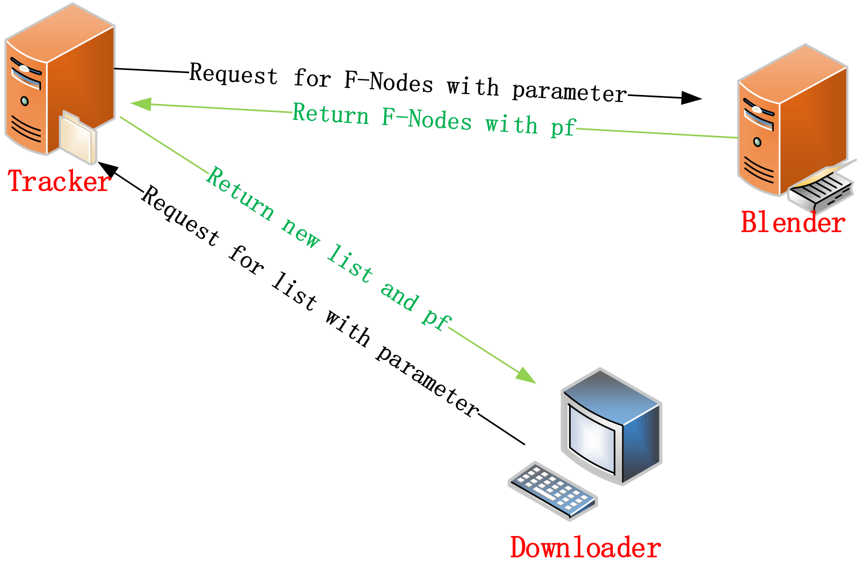

Figure 7. Tradeoff between Anonymity and Performance

F-Crowds makes trade-off between the anonymity and performance by adjust the parameter. When downloader selects a higher degree of anonymity, the F-Node number and forwarding probability also bigger, thereby providing a better degree of anonymity. When it choose a lower degree of anonymity, the F-Node number and forwarding probability become small, thus providing a lower degree of anonymity but a perfect performance.

4. Anonymity Analysis

In a system with N users, the system can achieve the maximum degree of anonymity is that, for the attacker, each member in the system has the same probability that he is the originator of the request. That is, the probability for each member is the same as the request originator. The author proposed measuring the degree of anonymity by entropy in the paper [3].Here, we use the way that paper proposed to measure the degree of anonymity of F-Crowds.

We denote by H(X) the entropy of the system after the attack has taken place:

$$\mathrm{H}(X) = -\sum_{k=1}^{N} p_k \log_2(p_k)$$

We denote by H_M the maximum entropy of the system:

$$H_M = \log_2(N-C)$$

Where C is the number of collaborators.

The information that the attacker can learned is:

$$\mathrm{H}_M - H(X)$$

We than define the degree of anonymity provided by the system as:

$$d = 1 - \frac{H_M - H(X)}{H_M} = \frac{H(X)}{H_M}$$

where $0 \le d \le 1$ and if d = 0 that means the probability of the request initiated by someone is 1. But if d = 1, the attacker can't get any useful information about who is the initiator of the request.

The author of Crowds believes that the probability for the path initiator is the predecessor node of the first attacker is:

$$P_{C+1} = \frac{N - p_k(N-C-1)}{N} = 1 - p_k \frac{N-C-1}{N}$$

The probability of the request is initiated by other nodes is

$$p_i = \frac{1 - p_{C+1}}{N-C-1} = \frac{p_k}{N}, C+2 \le i \le N$$

Thus

$$H(X) = P_{C+1} \log_2(P_{C+1}) + (N-C-1) p_i \log_2 p_i$$

There the degree of anonymity will be

$$d = 1 - \frac{H_M - H(X)}{H_M} = \frac{H(X)}{H_M}$$

$$= \frac{\frac{N - p_k(N-C-1)}{N} \log_2[\frac{N}{N - p_k(N-C-1)}] + p_k \frac{N-C-1}{N} \log_2[\frac{N}{p_k}]}{\log_2(N-C)}$$

According to the above formula, we can obtain the following simulation results.

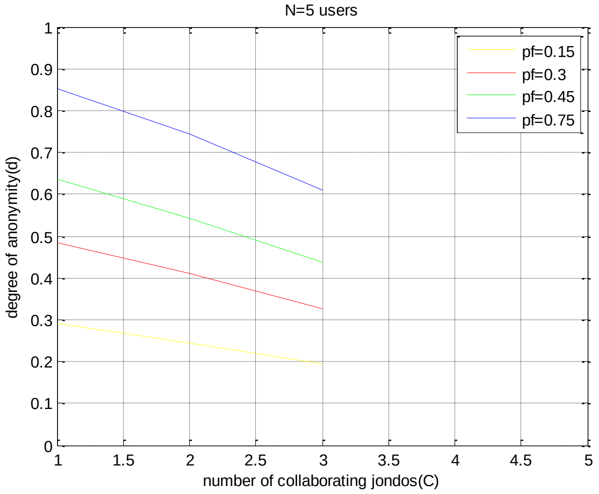

Figure 8. N=5 and C=[1,4]

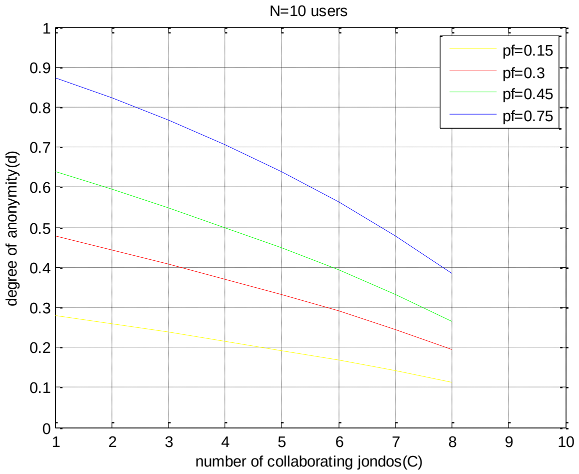

Figure 9. N=10 and C=[1,9]

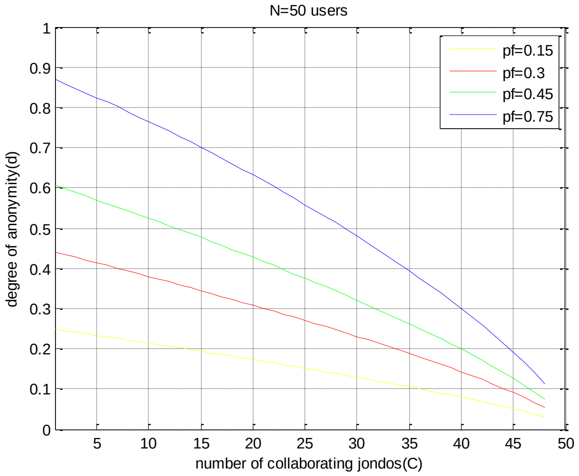

Figure 10. N=50 and C=[1,49]

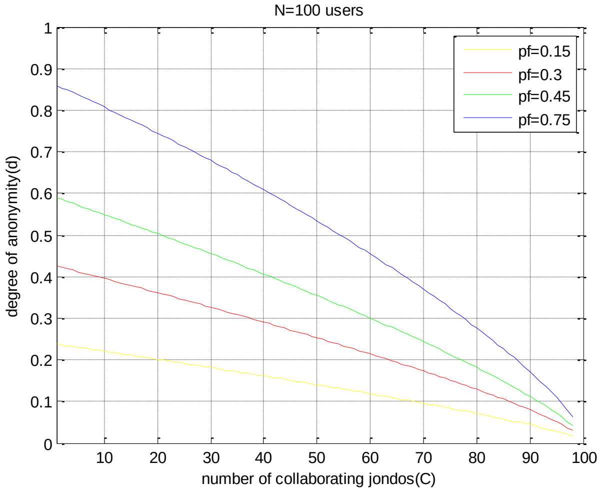

Figure 11. N=100 and C=[1,99]

5. Conclusion

In this article, we have introduced the anonymity Crowds which can provide anonymity service when people surf the Internet. We present F-Crowds, an anonymity communication scheme for the BitTorrent protocol which based on Crowds and we analyze the degree of anonymity of F-Crowds. F-Crowds not only provide BitTorrent the ability to download anonymity but also can tradeoff between the anonymity and performance to meet the different needs of users.

Crowds only provide sender anonymity, in some cases, that can't meet the needs for users. Therefore, extending the Crowds system from sender-anonymity only to both, sender and receiver anonymity becomes very meaningful. In the future, we will continue to study Crowds anonymity communication and committed to provide a anonymity communication which can achieve the sender anonymity and receiver anonymity.

Acknowledgements

This work is supported by the following programs: the National Natural Science Foun dation of China under Grant No.61170273; 2010 Information Security Program of China National Development and Reform Commission with the title "Testing Usability and Sec urity of Network Service Software".

References

[1] D. L. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms", Communications of the ACM, vol. 24, no. 2, **(1981)**, pp. 84-90.

[2] M. K. Reiter and A. D. Rubin, "Crowds: Anonymity for web transactions", ACM Transactions on Information and System Security (TISSEC), vol. 1, no. 1, **(1998)**, pp. 66-92.

[3] C. Díaz, "Towards Measuring Anonymity", Privacy Enhancing TechnologiesSpringer Berlin Heidelberg, **(2003)**, pp. 54-68.

[4] V. Sassone, E. ElSalamouny and S. Hamadou, "Trust in Crowds: probabilistic behaviour in anonymity protocols", Trustworthly Global Computing. Springer Berlin Heidelberg, **(2010)**, pp. 88-102.

[5] K. Bauer, "BitBlender: Light-weight anonymity for BitTorrent", Proceedings of the workshop on Applications of private and anonymous communications, **(2008)**.

[6] Y. Kawabe, "Formalizing and Verifying Anonymity of Crowds-based Communication Protocols with IOA", INTERNET 2012, The Fourth International Conference on Evolving Internet, **(2012)**.

[7] C. Diaz, "Towards measuring anonymity", Privacy Enhancing Technologies, Springer Berlin Heidelberg, **(2003)**.

[8] C. Andersson, R. Lundin and S. Fischer-Hübner, "Privacy Enhanced Wap Browsing With Mcrowds Anonymity Properties And Performance Evaluation Of The Mcrowds System", ISSA, **(2004)**.

[9] M. K. Wright, "The predecessor attack: An analysis of a threat to anonymous communications systems", ACM Transactions on Information and System Security (TISSEC), vol. 7, no. 4, **(2004)**, pp. 489-522.

[10] M. Wright, "Defending anonymous communications against passive logging attacks", Security and Privacy, 2003. Proceedings. 2003 Symposium on. IEEE, **(2003)**.

[11] H. Sui, "The cost of becoming anonymous: on the participant payload in crowds", Information processing letters, vol. 90, no. 2, **(2004)**, pp. 81-86.

[12] K. Kono, Y. Ito and N. Babaguchi, "Anonymous communication system using probabilistic choice of actions and multiple loopbacks", Information Assurance and Security (IAS), 2010 Sixth International Conference on. IEEE, **(2010)**.

[13] Y. Kawabe, "Theorem-proving anonymity of infinite-state systems", Information Processing Letters, vol. 101, no. 1, **(2007)**, pp. 46-51.

[14] I. Hasuo and Y. Kawabe, "Probabilistic anonymity via coalgebraic simulations", Programming Languages and Systems. Springer, **(2007)**, pp. 379-394.

[15] S. Schneider and A. Sidiropoulos, "CSP and Anonymity", Computer Security—ESORICS 96, Springer Berlin Heidelberg, **(1996)**.

[16] W. H. Tang and H. W. Chan, "MIX-Crowds, an Anonymity Scheme for File Retrieval Systems", INFOCOM 2009, IEEE, vol., no., **(2009)** April 19-25, pp. 1170, 1178.

[17] A. Daunou, "Design of a privacy-aware routing protocol for vehicular ad hoc networks", **(2012)**.

Authors

Tian-Bo Lu, was born in Guizhou Province, China, 1977. He is an Associate professor in School of Software Engineering, Beijing University of Posts and Telecommunications, China. His technical interests include information security and computer network.

Xin-Yuan Zhang, was born in Liaoning Province, China, 1990. She is a graduate student in School of Software Engineering, Beijing University of Posts and Telecommunications, China. Her technical interests include information and network security, anonymous communication.

Xiao-feng Du, was born in Shaanxi Province, China, 1973.He is a Lecturer in School of Computer, Beijing University of Posts and Telecommunications, China. His technical interests include information security and computer network.

Yang Li, was born in Hunan Province, China, 1978. He is a PhD and his technical interests include information security, distributed computing and P2P network.