International Journal of Security and Its Applications
Vol. 10, No. 5 (2016) pp.63-72
http://dx.doi.org/10.14257/ijsia.2016.10.5.06

An Analysis of Internet Censorship Circumvention Techniques

Tianbo Lu, Jinyang Zhao, Lingling Zhao, Yang Li and WanJiang Han

School of Software Engineering，Beijing University of Posts and Telecommunications, 100876, Beijing, China
lutb@bupt.edu.cn, zhaojinyang@sina.cn, wodepengyouzhao@163.com

Abstract

Since Internet was born, it has deeply influenced almost every aspects of people's life. However, at the mean time when people enjoy the convenience brought by the Internet, there is censorship existed every corner of the internet world. Censorship circumvention technique to protect people's communication against censors. During the past decades, anti-censorship techniques have a broad and extensive development, however not all of the censorship circumvention techniques have effect in circumvent the censors and there is still not a very useful and convenient technique to ensure the anonymity of internet communication. In this paper, we tried to analyze the current situation of censorship and anti-censorship techniques and give a comprehensive view on the censorship circumvention techniques and systems.

Keywords*: Censorship Circumvention, Anonymity, Privacy Protection*

1. Introduction

Along with the development of the Internet and the convenience brought by Internet, the war between censors and anti-censors is never stopped. Once there was a new technique for censoring Internet, soon there would be some ways to invade this censorship, and it is the same in the contrary way. In this paper, we want to give a comprehensive view on the current censorship circumvention techniques and summary on the trend of censorship circumvention techniques.

Censorship is defined as institution, system or practice of reading communication and deleting material considered sensitive or harmful [1]. Sometimes, technological advances not only make it easier for the exchange and spread of information and knowledge, it also may increase the practice and frequency of censorship. Yet the task of maintaining the status-quo through effective censorship policy is undergoing rapid change due to the growth and diversity of different devices and networks including: web traffic, email, P2P file-sharing, video, texting and messaging, VoIP (like Skype), social networks, and so on.

The main purposes of this article is to provide a review of the current anti-censorship technologies, to discuss the most critical design features to enable a successful and effective anti-censorship system, and to discuss the current trends and implications. And we tried to solve the following questions in this article: quantifying the efficacy of current online censorship technologies, metrics to be used, and fundamental limits to existing online (anti-) censorship technologies.

2. Internet Censorship

ISSN: 1738-9976 IJSIA

2.1. Censorship Definitions

Understanding how the censorship mechanism works is important to analyzing and improving censorship circumvention systems and techniques. Many censorship detection systems also use circumvention tools to have a supposedly uncensored access to the Internet. Tariq Elahi *et al*. in [2] give taxonomy of methods, tools and platforms for censorship circumstances. A report about the status of censorship circumstance in China is recently published [3] by Robinson et al.

There is not a well-defined, formalized or consistent definition for internet censorship, which is often treated as filtering, interference, tampering, and surveillance referring to different aspects of Internet censorship. Internet censorship has been provided by Verkamp and Gupta [4], Filasto and Appelbaum [5], in which, they reported the state of art of Internet censorship or provided researchers the detection tools. Many censorship techniques deliberately interfere with access to online resources, from the point of view of internet topology, internet censorship can be classified into client-based, server-based and network-based censorship.

Client-based censorship refers to the blocking of access to online resources, with the help of applications running on the same system of the network application client [6], or a network filter like parental control or company policy control enforcement filters, running as a personal firewall [7]. This kind of censorship can be enforced as a modified version of the network application itself, added with surveillance features like the TOM-Skype [8] [9] and the SinaUC [10].

Server-based censorship takes place on servers, with no disruption of the communication mechanics. The censor may selectively remove, hide or impairs access to specific content in the servers. This form of censorship is specifically hard to be analyzed because its mechanics are internal to the service and not exposed to the users. Tao Zhu et al. in [11] give a qualitative analysis on this kind of censorship and propose a hypothesis on how the mechanisms are actually enacted.

Network censorship takes place where between the client censorship and server censorship, which could provide the censor a wide coverage of network efficiently and allow the control of high number of communications through the management of a relatively few gateways or hubs.

2.2. Censorship Dtection and Monitoring

Censorship monitoring and detection is relatively a new field of research of which the methodology, tools and practices are still in the primary stage. Internet censorship detection can be referred to “the process that, analyzing network data, reveals impairments in the access to content and services caused by a third party (neither the client system nor the server hosting the resource or services) and not justifiable as an outage”. While the Internet censorship monitoring can be referred to “ the automated and continuous process of detecting Internet censorship over time , with the aim of revealing status changes in terms of the affected targets or the adopted censoring techniques” [12]. Herdict [13] is a well-known website for identifying web blockages like the denial of service, censorship and as well as other filtering. It is a crowed-sourced platform allowing users to report the accessibility of URLs form within their browsers. Its operating principle is very simple: leveraging crowdsourcing to collection the targets of internet for users and to have the users to perform the application-level censorship test. CensMon [14] is another tool which is specifically designed for censorship monitoring, which does not rely on users reporting the censored sites.

It is designed to be distributed in nature and to operate automated and continuous, and it solved the problem of selecting sites worth checking by differentiating access to network failure from possible censorship and using multiple input streams. OONI project [15], standing for the Open Observatory of Network Interface, provides complete and wide-ranging tools for censorship detection. It is a free software project, part of the wider Tor project with which it is tightly integrated. The main component is a phython script offering a list of censorship detection tests to be performed using Tor.

In 2015, Giuseppe Aceto *et al.* [12] propose a platform UBICA (User-based Internet Censorship Analysis) to achieve automated monitoring of censorship, which in their paper can provide simple but effective means of revealing censorship over time by adopting an integrated and multi-step analysis. For censorship events are easily spotted and described also in their temporal evolution, the UBICA integrated an algorithm for detecting censorship based on Internet measurements: if the test finds evidence of blocking, additional tests attempt to identify possible mechanisms, including DNS blocking, IP blocking, NO HTTP Reply, RST (TCP-level tampering), Infinite HTTP Redirect, and Block Page. They ran the UBICA for several months on selected targets and found evidence of several censorship techniques, such as DNS tampering and content filtering.

Giuseppe Aceto *et al*. [16] give a survey on Internet censorship detection. They proposed a reference for censoring techniques and a characterization of censoring systems, with definitions of related concepts. Based on the censoring techniques investigated in literature, they propose an analysis and discussion of censorship detection techniques and architectures and we present a chronological synopsis of the literature adopting or introducing them.

2.3. Censorship Techniques

Censorship techniques can be characterized according to different properties: like the trigger that initiate the censoring process, the action itself, and the symptom experienced by the users. According to the actions of the censorship process, censorship techniques can be classified into several categories: BGP tampering, DNS tampering, packet filtering, TCP connection disruption, soft censoring, TLS tampering, keyword blocking, and so on [16].

(1) BGP tampering

BGP (Border Gateway Protocol) is used by routers to coordinate across different administrative boundaries in the Internet. By withdrawing the presence from the BGP network view, a local network can remain unreachable to the outside networks. Guangchao Charles Feng el al. in [17] studied some of the cases of misuse or intentional tampering of BGP and estimated that a significant part of the Internet was potentially subject to prefix hijacking.

BGP tampering, though presents as trigger the destination or source IP addresses, in fact by diverting one direction of traffic to a black hole consequently makes bidirectional exchange impossible: TCP connections are surely affected and only one-way traffic is allowed through [18].

0DNS tampering

The censorship technique of DNS tamper was first analyzed in 2003 [19]. There are different variants involved with DNS tampering according to the presence of surveillance devices and on the path between the client and the recursive resolver and the kind of response that is provided back.

Having the administrative control on the DNS server allows to alter its behavior diverting it from the standard [20]. Different from DNS hijacking—performed directly at the recursive resolver—a more sophisticated technique is the injection of forged packets that imitate the

legitimate response of the queried DNS server but providing fake data. Injection can happen at different locations of the network, not necessarily on the path between the client and the target and requires a surveillance device on the network path between the stub resolver and the recursive resolver or between the latter and the authoritative server that should provide the requested Resource Record [21].

The original design of DNS did not assume a hostile network environment, hence the weakness of this protocol in the face of tampering; to extend it while retaining compatibility with the existing infrastructure the Secure DNS (DNS-Sec) specification has been proposed [22].

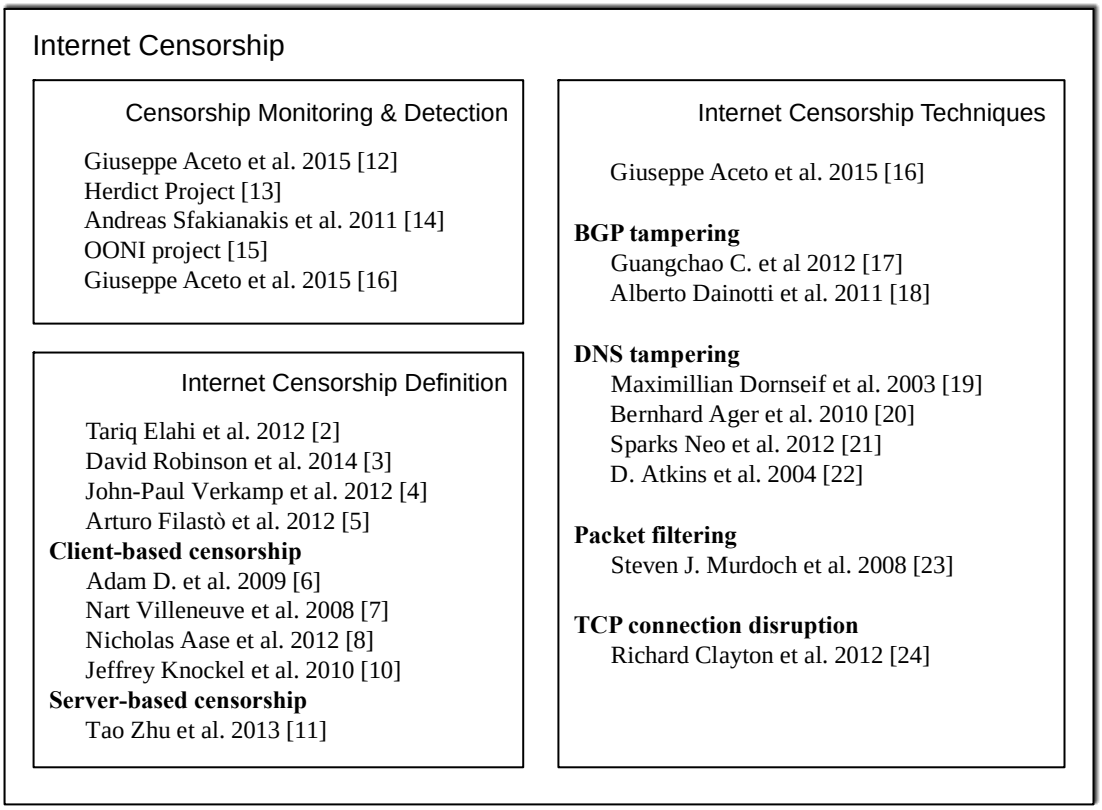

Figure 1. Internet Censorship: Definition, Monitoring and Detection, Techniques

(3) Packet filtering

The technique of packet filtering refers to the censorship techniques whose action is simply discard packets. With this kind of technique the trigger packet is silently dropped causing a symptom of type connection timed out error. This technique requires a surveillance device on the path between the client and the target (as opposed to BGP tampering and DNS hijacking), and the censoring device must be in-line too (as noted, they are possibly the same device). In a stateless censoring system this technique can be used to block IP addresses of targets that are also subject to DNS tampering, so that if the client circumvents censorship in the DNS resolution phase it is caught on the first packet of the TCP handshake [23].

(4) TCP Connection Disruption

An intentional disruption of the communication can happen during either the setup of the TCP connection or during the subsequent packet exchange belonging to the same connection, *i.e*. sharing the same 5-tuple (source IP, destination IP, protocol = TCP, source port, destination port).

The trigger for this technique contains the destination IP address or of the target [24] and possibly the transport level port numbers, to limit censoring to a specific application such as HTTP, HTTPS, SSH; this requires a surveillance device on the path between the client and the target.

3. Censorship Circumvention

3.1. Typical Censorship Circumvention

There are three main steps of censorship and its circumvention: (1) monitoring and surveillance, (2) blocking, filtering, and modifying content, (3) recording events [1]. The second step is the main focus in this article.

A typical censorship circumvention system is composed of many components working together [25]. Conceptually, a censorship network can be viewed as a set of filters or a firewall and possibly coupled with manual processes which restrict users from accessing or publishing certain content. Figure 1 shows an illustration of the typical components compromising an anti-censorship system. The process to circumvent the censors can involve several steps as follows: censored users (1) use circumvention system client software (2) on their computers to connect to circumvention tunnels (4), usually with the help of a tunnel discovery agent (3). Once connected to a circumvention tunnel, a user's network traffic will be encrypted by the tunnels and penetrate the firewall (7) without being detected by the censors (6). On the other side of the firewall, the network traffic will enter a circumvention support network (8) set up and operated by anti-censorship supporters (9). The computers, sometimes called nodes, in the circumvention support network act as proxies to access content from unobstructed Internet (10) and send the information back, not necessarily taking the same route, to the censored user's computer. Initially if a censored user knows nothing about the other side of the firewall, it is necessary to get them boot-strapped by employing out-of-band communication channels (5). Such channels include emails, telephone calls, instant messages, and mailing of CD-ROMS. Sometime users can also take advantage of these channels to locate circumvention tunnels (4), if the client software in use does not have a tunnel discovery agent (3). The key component in the overall system which facilitates covert communication within the censor network is primarily based on the software or tools and the underlying circumvention methodology.

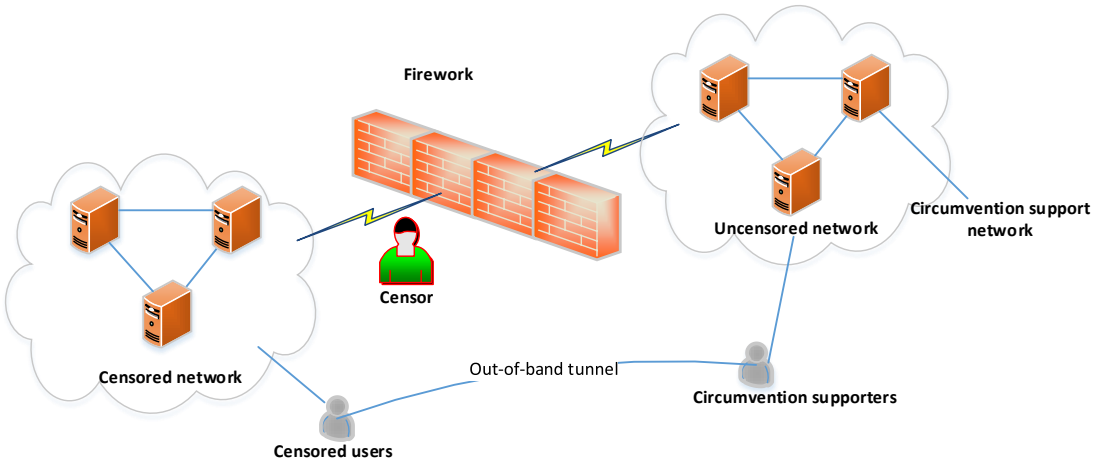

Figure 2. A Typical Censorship Circumvention System

3.2. Taxonomy of Censorship Circumvention Systems

The most common types of online content blocking strategies include IP address blocking, DNS hijacking and content filtering such as keyword or URL blocking [26][27]. IP address blocking operates by restricting users from accessing content by blocking the IP address where the content is hosted. IP blocking has the undesirable effect of over blocking since many sites can be hosted by a single IP address. Blocking the IP address of the site which contains objectionable content will also block all other sites on the same IP address which may not contain objectionable content.

DNS hijacking is finer grained compared to IP blocking, but it still is susceptible to over blocking. DNS hijacking allows operators to blocking access to content by blocking the name of the site instead of the IP address. For example, DNS hijacking would block or redirect users when they are trying to access specific sites. Therefore, if multiple sites are hosted from one IP address only the sites containing the name to be blocked will be restricted. However, in the event some news article needs to be censored on a particular website the contents of the article can't be censored without blocking the entire site. To address this limitation advances in content filtering such as keyword or URL filtering have been implemented to enable a higher degree of accuracy and granularity. So, for online censorship, operational cost grows following the increase of accuracy or granularity.

During the past decades, several anti-censorship techniques have been developed, which are mainly based on the following circumvention methods: HTTP proxy, multicast and broadcast, multilayer encryption, IP tunneling, re-routing based, and distributed hosting.

(1) HTTP proxy

HTTP proxy sends HTTP requests through an intermediate proxy server. A client connecting through a HTTP proxy sends exactly the same HTTP request to the proxy as it would send to the destination server unproxied. The HTTP proxy parses the HTTP request, sends its own HTTP request to the ultimate destination server, and then return the response back to the proxy client [1]. Circumvention systems using HTTP proxy include Freenet[28], Triangle Boy[29], Anonymizer[35], and so on. This kind of anonymous system is very simple and easy to use. For example, the Anonymizer provides protection of users' privacy when surfing the Internet, by just setting up a third-party website (http://www.anonymizer.com) to act as a middleman between the user and the site to be

visited. However, at the same time it can provide little anonymity and is very vulnerable to attacks. Once the agent based on is compromised, all the information would be exposed, including the address information.

(2) Multicast and broadcast

Censorship circumvention systems based on multicast and broadcast achieve anonymity through one-to-many communications among hosts. For example, P5, which is designed for providing scalable anonymity, is an anonymous protocol based on broadcast. P5 [36] creates a broadcast hierarchy, in which different levels of hierarchy provide different levels of anonymity at the cost of communication bandwidth and reliability. In P5, all messages to a given receiver come from one single upstream node, thus the receiver doesn't know the original message sender, and the sender also doesn't know where in the broad cast the receiver is or which host or address the receiver is using. With use of multicast or broadcast technology, senders send messages, which look the same, to a group of recipients. The more the number of group members is, the less probability that an attacker could guess who the real receiver is. Compared to those based on HTTP proxy, this kind of system may obtain more anonymity. Systems using CGI proxy include Circumventor [30], Psiphon [31], P5[36] and so on.

(3) IP tunneling

Some of the most common tools used for IP tunneling include virtual private networks or VPNs. VPNs give the user client a connection that originates from the VPN host rather than from the location of the client. Thus a client connecting to a VPN in a non-filtered country from a filtered country has access as if he is located in the non-filtered country. Systems like Firephoenix [32], and Gtunnel [15] using IP tunneling to perform circumvention.

(4) Re-routing

Re-routing based censorship circumvention system mainly get anonymity through one or several intermediate nodes called Mix. A Mix is a message pool to store messages from the former nodes and then send the messages in a confusing order. In this way, attackers couldn't detect the corresponding relations between senders and receivers. For example, Tor network can provide users with low-latency anonymous communication, by building circuits with publicly listed relays to anonymously reach their destinations. Later, Tor envisions the possibility of unlisted entry points to Tor network, called bridges, since the publicly listed relays are very easy to be blocked. However, bridges can still be found by powerful censors by observing the communications between bridges and user nodes. Re-routing systems route data through a series of proxy servers, encrypting the data again at each proxy, so that a given proxy knows at most either where the traffic came from or where it is gonging to, but not the both. TOR [33], Crowds [37] and JAP[34] use re-routing method.

(5) Distributed hosting

A distributed hosting system mirrors content across a range of participating servers that serve the content out to clients upon request. In this kind of circumvention systems, there are also a group of senders and receivers cooperating with each other in forwarding messages. The intermediate node work not to store messages, but to immediately make choice of sending the message to a next intermediate node randomly or the real receiver with use of some certain probability. The primary advantage of a distributed hosting system is that it provides access to the requested data even when the original server cannot, for instance if the original server has been overwhelmed by traffic or even taken down by a denial of service attack. Coral and Mix-Crowds [38] use the method of distributed hosting.

4. Conclusion

In this paper, we first analyzed the censorship definition, censorship monitoring and detection, and censorship techniques, on which the censorship circumvention systems and techniques' strategies focus. Through the analysis, we show that current censorship techniques have evolved very sophisticated and powerful, but still invadable.

In the remain part of this article, we analyze and summarize the basic architecture of a typical censorship circumvention system and give a comprehensive view on the current existing censorship circumvention systems and technique and a taxonomy of the basis and strategies the censorship circumvention systems are based on. However, through the analysis, we find that some censorship circumvention systems are just theoretical feasible, and some are even narrowly applicable. There is still not a very valid techniques or censorship circumvention system to invade the censor network effectively.

Acknowledgements

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

References

[1] C S. Leberknight, M Chiang, H Vincent Poor and F Wong, "A Taxonomy of Internet Censorship and Anti-Censorship," Princeton University, Department of Electrical Engineering, **(2010)**.
[2] T. Elahi and I. Goldberg, "CORDON: A Taxonomy of Internet Censorship Resistance Strategies", Technical report, Technical Report CACR 2012-33, University of Waterloo, **(2012)**.
[3] D Robinson, H Yu and A. An, "Collateral Freedom: A Snapshot of Chinese Users Circumventing Censorship", Technical report, The Open Internet Tools Project, **(2014)**.
[4] J-P Verkamp and M Gupta, "Inferring mechanics of web censorship around the world", Free and Open Communications on the Internet, Springer, USENIX Association, Bellevue, WA, USA, **(2012)**.
[5] A Filastò and J Appelbaum, "OONI: open observatory of network interference", Proceedings 2nd USENIX Workshop on Free and Open Communications on the Internet, **(2012)**.
[6] L Campher and C Bezuidenhout, "An evaluation of existing measures aimed at restricting the use of the Internet as an avenue to initiate sexual activities with adolescents", Child Abuse Res. South Africa, vol. 11, **(2010)**.
[7] A D. Thierer, "Parental controls & online child protection: a survey of tools & methods", Soc. Sci. Res. Network, **(2009)**.
[8] N Villeneuve, "Breaching Trust: An Analysis of Surveillance and Security Practices on China's TOM-Skype Platform", Technical report, Information Warfare Monitor/ONI Asia, **(2008)**.
[9] N Aase, J R. Crandall, A Diaz, J Knockel and J Ocana Molinero, "On measuring censors' resources and motivations", FOCI'12: Second USENIX Workshop on Free and Open Communications on the Internet, **(2012)**.
[10] J Knockel, J R. Crandall and J Saia, "Three researchers, five conjectures: an empirical analysis of TOM-Skype censorship and surveillance", Proceedings 1st USENIX Workshop on Free and Open Communications on the Internet (FOCI 2011), San Francisco, CA, USA, **(2011)**.
[11] T Zhu, D Phipps, A Pridgen, J R. Crandall and Dan S. Wallach, "The velocity of censorship: high-fidelity detection of microblog post deletions", March 2013, **(2013)**.
[12] G Aceto, A Botta, A Pescapè, Nick Feamster, Tahir Ahmad and Saad Qaisar, "Monitoring Internet Censorship with UBICA", Seventh International Workshop on Traffic Monitoring and Analysis (TMA'15) Barcelona, Spain, **(2015)**.
[13] Herdict Project, http://www.herdict.org
[14] A. Sfakianakis, E. Athanasopoulos and S. Ioannidis, "Censmon: a web censorship monitor", USENIX FOCI, **(2011)**.
[15] F Arturo and J Appelbaum, "Ooni: Open observatory of network interference", USENIX FOCI, **(2012)**.

[16] G. Aceto and A. Pescapé, "Internet Censorship detection: A survey", Computer Networks, vol. 83, no. 4 **(2015)**, pp.381–421
[17] G C. Feng and S Z. Guo, "Tracing the route of China's Internet censorship: an empirical study", Telemat. Inf. (0), **(2012)**.
[18] A Dainotti, C Squarcella, E Aben, K C. Claffy, M Chiesa, M Russo and A Pescapè, "Analysis of countrywide Internet outages caused by censorship", Proceedings of the 2011 ACM SIGCOMM Internet Measurement Conference, ACM, **(2011)**, pp. 1–18.
[19] D, Maximillian, "Government mandated blocking of foreign web content", arXiv preprint cs/0404005, **(2004)**.
[20] B Ager, W Mühlbauer, G Smaragdakis and S Uhlig, "Comparing DNS resolvers in the wild", Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, ACM, **(2010)**, pp. 15–21.
[21] S Neo and S Tank, "Dozer. The collateral damage of internet censorship by dns injection", SIGCOMM Computer Communication Review, vol.42, no.3, **(2012)**, pp.21-27.
[22] D. Atkins and R. Austein, "Threat Analysis of the Domain Name System (DNS)", RFC 3833 (Informational), **(2004)**.
[23] S J. Murdoch and R Anderson, "Tools and technology of Internet filtering, in: Ronald J. Deibert, John Palfrey, Rafal Rohozinski, Jonathan Zittrain (Eds.), Access Denied", The Practice and Policy of Global Internet Filtering, **(2008)**, pp. 57–72.
[24] R. Clayton, S J Murdoch and RN Watson, "Ignoring the great firewall of China", George Danezis, Philippe Golle (Eds.), Privacy Enhancing Technologies, Lecture Notes in Computer Science, vol. 4258, Springer Berlin Heidelberg, Berlin, Heidelberg, **(2006)**, pp. 20–35.
[25] "Defeat Internet Censorship: Overview of Advanced Technologies and Products", Global Internet Freedom Consortium, **(2007)**.
[26] Roberts, "2007 Circumvention Landscape Report: Methods, Uses, and Tools", The Berkman Center for Internet & Society at Harvard University, **(2009)**.
[27] R. Clayton, S. J. Murdoch, and R. N. M. Watson, "Ignoring the great firewall of china", I/S: A Journal of Law and Policy for the Information Society, vol. 3, no. 2, **(2007)**, pp.70–77.
[28] I Clark, O. Sandberg, B. Wiley and T.W. Hong, "Freenet: A Distributed Anonymous Information Storage and Retrieval System", Cornell University, **(1999)**.
[29] A Guun, "Triangle Boy: A New Angle on Free Speech", **(2001)**.
[30] J Karlin, D Ellard, A W. Jackson and C E. Jones, "Decoy Routing: Toward Unblockable Internet Communication", Raytheon BBN Technologies, **(2011)**.
[31] J Jia and P Smith, "Psiphon: Analysis and Estimation", **(2004)**.
[32] Z Jianfeng and N Sun, "Fire phoenix cluster operating system kernel and its evaluation", Cluster Computing, 2005. IEEE International. IEEE, **(2005)**.
[33] R. Dingledine, N. Mathewson and P.Syverson, "Tor: The Second-Generation Onion Router", Proceedings of the 13th USENIX Security Symposium. San Diego, California, **(2004)**, pp. 303-320.
[34] O. Berthold, H. Federrath and S. Köpsell, Web MIXes, "A System for Anonymous and Unobservable Internet Access", In Proceedings of International Workshop on Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Unobservability, LNCS, vol.2009, **(2000)**, pp.115-129.
[35] J. Boyan, "The Anonymizer: Protecting User Privacy on the Web", Computer-Mediated Communication Magazine, vol.4, no.9, **(1997)**.
[36] R. Sherwood, B. Bhattacharjee and A. Stinivasan, "P5: A Protocol for Scalable Anonymous Communication", Proceedings of the 2002 IEEE Symposium on Security and Privacy, IEEE Computer Society Press, **(2002)**, pp.58-70.
[37] M.K. Reiter and A.D. Rubin, "Crowds: anonymity for web transactions", ACM Transactions on Information and System Security, vol.1, no. 1, **(1998)**, pp. 62-92.
[38] W.H. Tang and H.W. Chan, "MIX-Crowds: An Anonymity Scheme for File Retrieval Systems", INFOCOM 2009, IEEE, **(2009)**, pp. 1170-1178.

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.

Jin-Yang Zhao was born in Hebei Province, China, 1991. He is a graduate student in School of Software Engineering, Beijing University of Posts and Telecommunications, China. His technical interests include information and network security, anonymous communication.

Ling-Ling Zhao is a graduate student in School of Software Engineering, Beijing University of Posts and Telecommunications, China. Her technical interests include Cyber-Physical System and P2P 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.

Wan-Jiang HAN was born in HeiLongJiang province, China, 1967. She received her Bachelor Degree in Computer Science from Hei Long Jiang University in 1989 and her Master Degree in Automation from Harbin Institute of Technology in 1992.She is an assistant professor in School Of Software Engineering, Beijing University of Posts and Telecommunication, China. Her technical interests include software project management and software process improvement