Multimed Tools Appl (2015) 74:6509–6518
DOI 10.1007/s11042-014-2127-7

New detection method and countermeasure of cyber attacks in mix networks

Kwang Cheol Park · Hoon Shin · Won Hyung Park · Jong In Lim

Received: 14 May 2014 / Accepted: 26 May 2014 / Published online: 27 June 2014

Abstract Recently, studies on the Mixed Networking which guarantees the anonymity in Internet environment are actively carried. Since this technology uses the coded communication and its communication paths are changed frequently, it is difficult to detect attacks of the hackers. In this situation, if the cyber-attack occurs between countries, there shall be a high potential for the hackers to use the anonymous network technology in order to hide themselves. Anyway, the anonymous network technology is continuously being updated by the hackers and a new technology is under development. Thus, this paper verified statistically the attacking methods which the hackers shall create by increasing the data transmission rate of TOR through manipulating the speed of the anonymous network, and proposed political countermeasures to detect hacker's attacks effectively which use this technology.

Keywords Mix networks · TOR · Cyber attack · Ext node · Blacklist IP

1 Introduction

The mix network technology has been developed as the demands for protecting personal information and technical security in Internet have been increased. Hackers frequently access the service with other IP instead of user's IP through anonymous IP address, and this technology is used as a means to access the site whose network is blocked as a harmful site

K. C. Park (✉) · H. Shin · J. In Lim
Center for Information Security Technologies, Graduate School of Information Security, Korea University, Anam 5ga, Sungbuk gu, Seoul, Republic of Korea
e-mail: muryo@naver.com

H. Shin
e-mail: kadosu@daum.net

J. In Lim
e-mail: jilim@korea.ac.kr

W. H. Park (✉)
Department of Cyber Security, Far East University, Wangjang-ri, Gamgok-myeon, Eumseong-gun, Chungbuk 369-700, Republic of Korea
e-mail: whpark@kdu.ac.kr

or used in order to access the other party hiding itself. Especially, it is also used for providing the anonymous service which the user is not able to know who provides. In addition, the anonymous network technology could be used by information agency or investigating institutions for detouring the network block of but accessing the harmful site when necessary for information gathering, but also could be used by the vicious hacker for carrying attack hiding himself. Thus, this paper check and understand the anonymous network technology which are mostly used by hackers recently and studies on the technology which improves the data transmission rate using the automatic anonymous network tools, and through which verifies country and node which has the fastest bandwidth through experiments, and prepares a political and technical basis capable of counteracting against cyber-attack which uses the anonymous network. In addition, since there has been no academic paper for the anonymous network either for measuring its speed, we studied the basic technology in this field.

2 Related work

2.1 The anonymous network technology

TOR(The Onion Router) was a project started in 1998 for developed by NRI(Navy Research Institute)[10]. I2P (Invisible Internet Project) was a project started in 2003 for anonymous communication and was developed by modifying the existing Freenet. Unlike TOR, it was not developed in order to use the normal Internet anonymously but was created in order to provide a service such as bulletin board anonymously[12]. JAP(Java Anon Proxy) is a Java Application which supports anonymous use of Internet and its main purpose is to us Internet anonymously. It provides the anonymity for the Web request by MIX method, and the Mix Server decodes the coded input message and passes through each Mix server node by changing the order so that the input message shall be decoded in hierarchy and delivered randomly. Thus, each Mix Server Node shall be able to recognize the information of the immediately previous or next node only. When several central nodes execute JAP by processing the connection request in MIX method, JAP shall access InfoService first to obtain the information of Mix Station. Then, Mix Station shall put all requests from users together and send them to Middle Mix which shall then send them to the Last Mix which then shall send them to Cache-Proxy to access Internet.

Peekabooty is a variance collaborative privacy network which uses P2P method and was developed by Cow hackers group. When it accesses a site which is blocked by a firewall, other computer shall convey the message on behalf. It uses a web-browsing tool which provides the anonymity and carries a collaborative communication in which each client operates as a proxy.

2.2 Characteristics of mix networks

Another different point from the single proxy method for Tor network is that it is difficult for the censorship authority to block IP because servers in Tor network are Peer-to-peer method[13]. For example, if they supported computers assigned to domain of Harvard University by the server of Tor network, the censorship authority in China is not able to block addresses of Harvard University en bloc. Since the Tor network is a method where users support their computers voluntarily for free communication in Internet, it is legal and important institutions, schools and enterprises shall have servers, so it is impossible for the censorship authority to block them completely. Finally, middlemen which just relay the contents in behind the server shall prevent the network from being contaminated by spy servers or hostile servers. Concept of this system is as follows. For example, let's assume that China placed servers equivalent to 10 % of servers used in Toe network or Relay servers. Of course, it's not easy to operate such many servers as number of volunteer increases but it is still nearly impossible to obtain the information China wants even though servers are contaminated by such an invasion. Let's also assume that you use Tor network so that you may receive the information you want by passing through relaying servers by 10 times. The possibility that each server is a spy during passing 10 servers shall be 1/10 respectively. So, when you connect these servers in serial, then the possibility that all servers are spy servers that your information shall be transferred to China shall be only $1/10^{10}$. Even though the frequency of relay is only 5, there shall be no possibility for the censorship authority to win. The anti-IP tracking program such as Tor network is very useful for many righteous people who fight for the freedom in face of danger while hiding their personal information. In the other hand, this useful tool could result what they don't want if it is used for the terrorists or crime group. Such a situation brings us some questions over the developer's philosophy and ethics such as "Is the tool surely neutral? Are there virtues and vice in the tool?" [11][2] (Table 1).

2.3 Transmission types of anonymous network

The cell type of transmission unit is 512 byte cell and the first 3 byte header shall not be coded while the rest 509 byte shall be coded. There are Control Cell and Relay Cell in cell types and total 22 types exist. Below figure shows Control Cell and Relay Cell types of Tor.

For selecting path (length of basic path is 3), the relay node (OR) to be used shall be selected based on bandwidth capacity information using descriptor information for OR which is given from the directory server. Exit OR node shall be selected first, and then Entry OR node and the Intermediate OR node shall be selected lastly. For creating the Circuit (use of Control Cell and Relay Cell), Diffie-Hellman (DH) handshake protocol is used in session Secret Key discussion using TLS/SSLv3 in Link certification and coding. Relay Extend Cell shall be coded using AES-CTR and OP shall discuss the Secret Key with each OR by sending Create Cell and Relay Extend Cell. Below figure shows the Circuit creation of TOR.

TCP stream transmission (using Relay Cell) is that OP transmits Relay Begin Cell to Exit OR, and informs of the connection by sending Relay Connected Cell to OP from Exit OR. Then client shall begin to send data to the server through Circuit, and if client requests OP to disconnect the circuit after finishing data transmission, OP declares the disconnection by sending Relay End Cell to the server.

Table 1 Characteristics of mix network(Tor)

Characteristics of TOR	○ To operate as an infrastructure to support anonymous communication which supports real time interactive anonymous communication independent on application. - Available to utilize in most of network applications through HTTP/SOCKS Proxy - To provide a tool which prevents indirect access information from leaking through DNS inquiry ○ To apply policies of access control and warning node block to prevent unauthorized use. ○ To provide the anonymity of reverse direction/responders/receivers - To provide 'Hidden Service' function which supports anonymous publishing

2.4 Vulnerability of mix network and the transmission speed

The Vulnerability of the anonymous network is that since all clients could obtain all router information of TOR and cells come from several lines are transmitted by round robin method, delay time in other connection shall increase as well if a certain node is over loaded. This means that a vicious user may know all Relay nodes of a certain line by generating a specific traffic and measuring delay time of all nodes of TOR. In addition, since the directory server stores node information of each TOR without checking, vicious information could be sent to the directory server and Diffie-Hellman which is used for distributing Secret Key has the possibility to receive Man-in-the-middle-attack. The reason why the data transmission speed in the anonymous network is slow is that the user may set more quantity than the traffic which it can actually contribute in Congestion Control while the network basically does not have sufficient capacity to control for all users. In addition, the current path selection algorithm has no function to disperse the load. There is a problem in taking any action against High/Variable Latency or Connection Failure, and the user who has a small bandwidth shall have the overload in downloading the directory information. This paper studies the technology which may detect the hacking when the hacker attempts it by improving data transmission speed through manipulating the program of TOR[1].

3 Measuring data transmission speed when hacker manipulates TOR

3.1 An experiment to measure data transmission speed

The Constraints before the experiment to improve data transmission speed is to establish the line using about 1,600 TOR nodes in 50 countries in the world and to access the server through maximum 3 countries because the current line uses the path whose length is 3. In addition, we assume that it is possible to use country information of each TOR router through GeoIP and the data transmission speed shall be higher than the path which is connected with routers located in max. 3 countries when we use the path generated by TOR routers. Then we designate the Web Proxy which is set to use TOR for the experiment of improving the data transmission speed as the environmental variable, and measure the time required for downloading various size of data in servers located in several countries by WEGT program [6], then compare the time to the average time required for downloading the data of 1 KB. Prerequisites for this is that the country information which falls into IP of GeoIP shall be correct and countries which have many routers shall be selected for the stability of TOR lines.

3.2 Test environment of experiment to measure data transmission speed

The test environment is to install Intel Core i7 CPU 860 2.80GHz 3GBRAM PC, Linux Virtual machine (VMware, Fedora 10, Linux 2.6.27.41) and TOR version 0.2.1.22, Polipo (Web Proxy) and update GeoIP file with the latest information. Below (Table 2) shows the number of TOR nodes by each country used in the experiment.

Table 2 Number of TOR node by country [4]

Total	Germany	France	USA	Netherlands
2,963	548	173	835	157

Countries to be used in the experiment shall be those which have big bandwidth routers for the stable connection of TOR lines, and test 0–4 country codes. That is to define in order of Total, Germany(DE), France(FR), USA(US) and Netherlands(NL). The country code combination shall be limited to No use of TOR router and Use of TOR router (Exclude Nodes is optional) and the path length shall be limited to 3(Basic value) and 2[7].

We could obtain the conclusion that for the data transmission speed according to path length of TOR preferred countries; the improvement effect was bigger when designating one country than no country. In addition, we could find that the speed was higher when path length is 3 than 2 but still there is a speed deviation. And, when the preferred country is not designated and the speed was high, the path was formed with nodes of one country or two countries. We could see the speed improvement effect by 24 % higher than basic TOR when use TOR designating Germany which has big bandwidth while the path length of the server in Korea was 3, and by 44 % higher than the basic Tor when use TOR designating Netherlands while the path length is 2. Of course, the speed was different according to each server but we could obtain the conclusion that the speed was higher as the bandwidth is wider than the distance of nodes. Following is the verification of the hypothesis based on the statistics in order to evaluate the result of the experiment. First, a hypothesis shall be established with the method to compare more than 3 average values.

Null Hypothesis (H_0): $u_1=u_2=u_3=u_4=u_5$

Anti-Hypothesis(H_1): not H_0(Average of all groups shall not be same) When assumes that the groups to compare are $A_1,A_2,\cdots,A_k$, and number of data of each group is n_i,

$$y_{ii}, i=1,2,\cdots,k, j=1,2,\cdots,n_i$$

$\overline{\mu}$: Average of Total Groups

$\overline{y_i}$: Average of Group i

when separates deviation of each data as follow,

$$y_{ii}-\overline{\overline{y}} = \left(y_{ii}-\overline{\overline{y_i}}\right) + \left(\overline{\overline{y_i}}-\overline{\overline{y}}\right) \tag{1}$$

The sum of square of deviation of all data could be obtained as follows from (Eq. 1).

$$\sum_{j=1}^{J}\sum_{i=1}^{n_j}\left(Y_{ij}-\overline{Y_j}\right)^2 = \sum_{i=1}^{n_j}\left(Y_{i1}-\overline{Y_1}\right)^2 + \sum_{i=1}^{n_j}\left(Y_{i2}-\overline{Y_2}\right)^2 +..+ \sum_{i=1}^{n_j}\left(Y_{ij}-\overline{Y_j}\right)^2 \tag{2}$$

SST = SSE + SSA

Where SST(Total Sum of Squares) is the sum of squares of deviation between the value of each data and the average of all groups, that is, the sum of total variation obtained not by dividing by groups.

SSE(Error Sum of Squares) is the sum of squares of deviation between groups where each data is belonged, that is, the sum of squares of deviation in the same group while SSA(Among Treatments Sum of Squares) is the sum of squares of deviation between the average of each group and average of all groups, that is, the sum of squares of deviation between each group. Now, we may determine that the difference of the average between groups shall have significance if the mean square between groups (variance between groups) which is given by dividing the sum of squares of each group by Degree of Freedom is sufficiently big comparing to mean square (variance in the group) in the group. In the other hand, we may determine that the difference of the average between groups shall have no significance if a square of average between groups is similar to squares of variation in the group or smaller.

Below (Table 3) shows the variance analysis of data transmission speed materials according to router groups. The variance analysis shows that F ratio of the data is 3.556 which means that the mean square between groups (variance between groups) is 3.556 times of the mean square in the group(variance in the group). However, the provability of 0.02(Significant Provability, P (F>3.556|H_0 is true) observed in our data is too big difference to come by chance if the Null Hypothesis (There is no difference of data transmission speed according to groups) is true because the F value obtained from the data, 3.556 which falls into upper 5 % when the Null Hypothesis is true is far bigger than the base value, 2.946 when Degree of Freedom is 3 and 36 respectively. Thus, we have to reject the Null Hypothesis that there is no difference between groups but accept Anti-Hypothesis and determine that the difference between groups has significance. That is, routers according to data transmission speed show significant difference statistically. Figure 6 showed that the average data transmission speed of DE router group was lowest (Figs. 1, 2, 3, 4 and 5).

That is, since German router has the lowest speed comparing to path length, the speed of its routers is the highest. Finally, there is provability that the hacker shall make a hacking attack by designating German EXT node whose speed is high.

4 Detection method and countermeasure of TOR in cyber attacks

4.1 A countermeasure through utilizing IP blocking list

Above all, as shown on the figure, Exit Node IP bandwidth which uses TOR could be obtained through Vidalia Control Panel → View the Network[9][5]. In addition, since Vidalia itself is formed with Open Source, Exit Node IP bandwidth which varies in real time by modifying program could be easily found. First, the Early Warning System which stores relevant IP shall be established after acquiring Exit Node. The important point in here is the management rating which consists of Green (1st stage), Orange (2nd stage) and Red (3rd stage) and when first Exit Node IP is acquired, IP shall be saved in Green portion, and shall be changed to Orange range when it reached a certain level of marginal value by accessing frequency. In addition, when a specific attack method occurs on the IP in Green and Orange, then it shall be changed to Red (3rd stage) and IP shall be immediately blocked. Below Table 4 shows the list of items needed for Early Warning System.

As we may see from Table 4., it contains detection time, modification time, accident type, number of detection, management rating, country information and IP history and its blocking experience etc. This system enables to manage IP history when a certain IP is found, and to know whether IP in question is accessed through TOR or not, and also to understand the connection relation with other typed accidents easily. The Early Warning System established

Table 3 Variance analysis of data transmission speed materials

Variable	Sum of squares	Degree of freedom	Mean square	F ratio	P-value	F reject value
Treatment (between groups)	0.000954	3	0.000318	3.556873	0.023642	2.866266
Residual (within groups)	0.0003219	36	8.94E-05	–	–	–
Total	0.004174	30	–	–	–	–

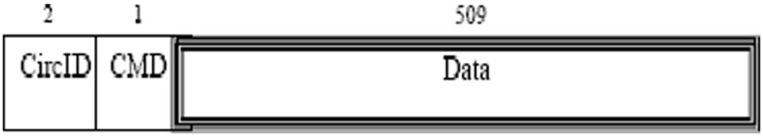

Fig. 1 TOR (Control) cell format

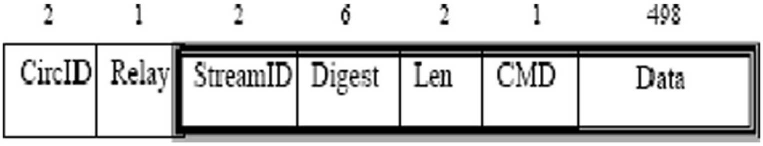

Fig. 2 TOR relay cell format

this way shall make it possible to know whether IP in question uses TOR or not by each IP address and to make a systematic threat management.

4.2 EXT node detection in invasion detection system

Since most of the packet is coded in case of attacks using TOR, the hacker is not able to know the contents of the packet even it acquired the packet through mirroring (tapping) in the network. Therefore, it has to find any way to detect the contents in Front-End of the server or inside the server in the network. In addition, it may know the packet is decoded because Node IP of TOR in the list of Exit Node section is coded. Thus, it may know the bandwidth of Exit Node which is the section conveyed to the user (Fig. 6).

First, when they detect in Front-End, they may understand contents with details shown on header of the protocol. Below figure shows the captured details of the packet which passed Exit Node. Second, a method to monitor the Access Log inside the server in real time shall be considered. In addition, it is impossible to know the

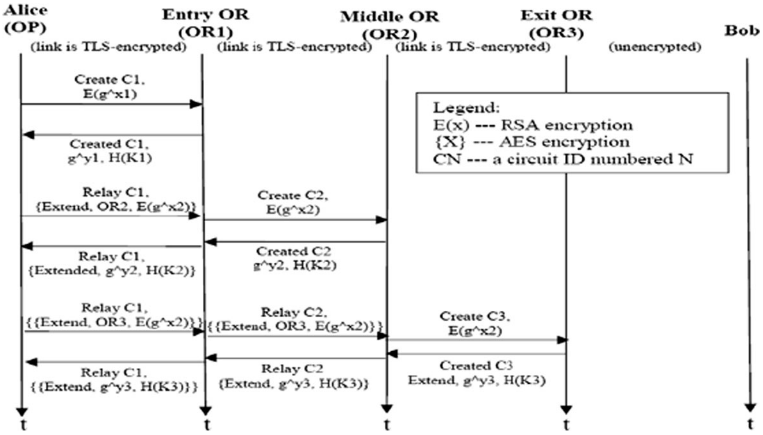

Fig. 3 Circuit creation [8]

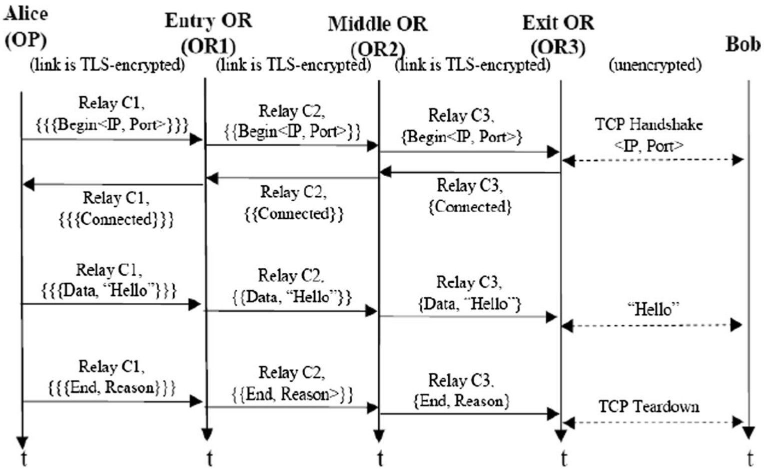

Fig. 4 TCP Stream transmission [8]

contents of the data with only packet itself if the packet is coded in the network such as SSL. When Access Log is monitored in real time inside the server and a vicious typed packet is detected, the IP in question shall be blocked by comparing to IP Blocking List. Of course, it is possible to block in real time through the current security equipment linkage such as Web Firewall and IPS but it is still difficult to detect with the packet itself in case of unknown attack. For IP Blocking List, blocking is available by comparing just IP itself even when a vicious packet is not detected, and a effective management could be carried because the behavior based monitoring to know what the IP does is available. There is a merit to understand the

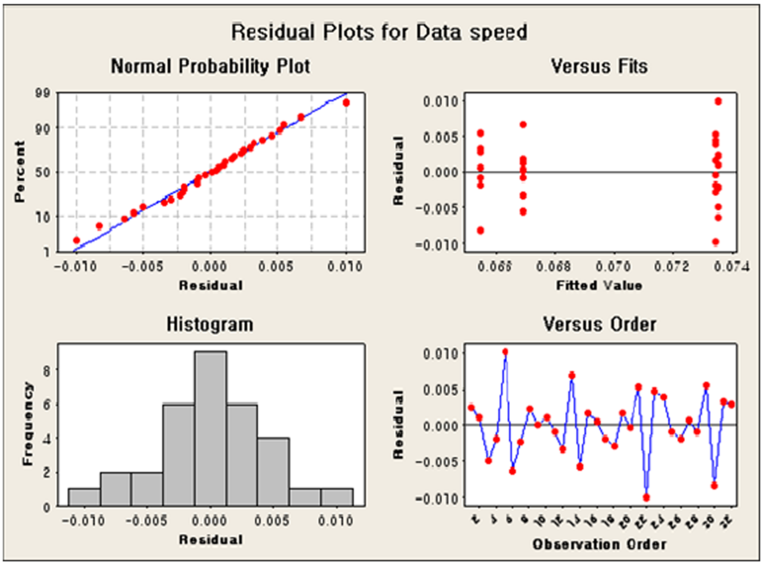

Fig. 5 Data transmission speed

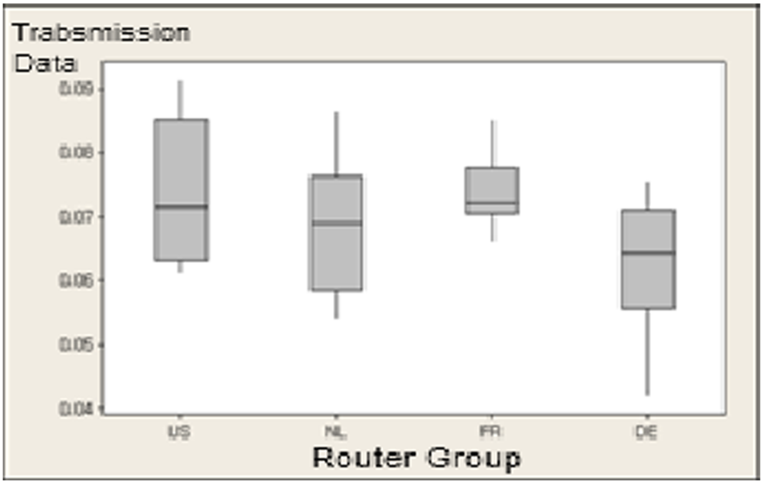

Fig. 6 Data transmission speed vs routers

contents of the packet in detail without any problem even though they are coded if they use these ways.

4.3 A countermeasure of TOR IP through international collaboration

There shall be a high potential that the cyber-crime or large sized cyber-attack in future shall be carried through the anonymous network which ensures the anonymity. However, there exist various limitations in countermeasures against them. When we see the situation of TOR IP composition at the moment, it is used worldwide, especially in developed countries such as Germany, France and USA. Thus, if they collaborate internationally through cooperation between institutions in cyber safety field, significant effects shall be obtained. When the authority which supervises cyber safety of the country understands how many TOR Node IP exist and are used in its own country and determine the inflow path of the IP in question on request from other country, then they could understand at least the country where the attacking packet was come from rather than who is the attacker. Since the cyber-attack shows a tremendous power surpassing time and space, each country shall have to address against hacker's hacking by reinforcing collaboration system between countries.

Table 4 Major items of attacking IP address management system[3]

Item	Function description
Detecting time	The first occurring time and input time
Modification time	Time to reflect varied items of IP information
Accident type	A type of cyber threats
Number of detection	Number of detection of security system and invasions
Management rating	Sorting according to level of attack frequency Green(1) < Orange(2) < Red(3)
Area (country) information	IP management Information
IP history	Whether or not RBL registration in local or foreign countries.
Block or not	Experience of blocking the attacking IP

5 Conclusion

The result obtained from the experiment proposed in this paper says that the hackers shall select the country which has high bandwidth and lots of nodes and are stable when it designates the country. Through this, it is very difficult to take a proper action against the cyber-attack using Tor network. However, if they use the method which this paper proposed, then the cyber-attack could be detected and counteracted in many aspects. In addition, this TOR technology shall be developed continuously and the cyber-attack using this technology are expected to be suddenly increased. This study is not a research to improve the data transmission speed in the position of the hacker, but a research to detect the hacking or hacking attempt to high bandwidth node like Germany in aspect of security. In addition, since it is expected that the cyber-attack using TOR shall be gradually increased in future, developing highly advanced detection technology, identifying the hacker's location through international collaboration and black list policy shall be taken in order to detect and block these attacks.

References

1. Bauer K et al (2007) Low-resource routing attacks against tor. WPES, New York, pp 11–20
2. Invisible Internet Project, http://www.i2p2.de
3. Jeong Wook Kim et al., "The effective monitoring and controlling method through malicious domain-IP traceback, Jouranl of Information and Security, 2009. 3
4. K.W.J. et al., "Egregious use of Tor servers?", RechtenForum, 2007
5. Martin Suess, "Breaking TOR Anonymity", http://www.csnc.ch/misc/files/publications/the_onion_router_v1.1.pdf, 2008
6. Mike Perry, "TorFlow, tor network analysis", In HotPETs 2009, page 14, 2009
7. Steven J. Murdoch, "Tor: anonymous internet communication system", University of Cambridge, Computer Laboratory, 2006
8. Steven J. Murdoch et al., "Low-cost traffic analysis of Tor", IEEE Symposium on security and privacy, 2007
9. Timothy G. et al., "Browser-Based Attacks on Tor", http://web.mit.edu/tabbott/www/papers/tor.pdf, 2007
10. Tor (anonymity network), http://en.wikipedia.org/wiki/Tor_(anonymity_network)
11. Tor Network Status, http://torstatus.blutmagie.de
12. Tor project, http://www.torproject.org
13. TorStatus, http://torstatus.kgprog.com