An Analysis of Anonymity Technology Usage

Bingdong Li[1]**, Esra Erdin[1]**, Mehmet Hadi Güneş[1], George Bebis[1], and Todd Shipley[2]

[1] University of Nevada - Reno, Reno, NV 89557
{bingdonli, eerdin, mgunes, bebis}@cse.unr.edu
[2] Vere Software, Reno, NV 89519
todd@veresoftware.com

Abstract. Anonymity techniques provide legitimate usage such as privacy and freedom of speech, but are also used by cyber criminals to hide themselves. In this paper, we provide usage and geo-location analysis of major anonymization systems, i.e., anonymous proxy servers, remailers, JAP, I2P and Tor. Among these systems, remailers and JAP seem to have minimal usage. We then provide a detailed analysis of Tor system by analyzing traffic through two relays. Our results indicate certain countries utilize Tor network more than others. We also analyze anonymity systems from service perspective by inspecting sources of spam e-mail and peer-to-peer clients in recent data sets. We found that proxy servers are used more than other anonymity techniques in both. We believe this is due to proxies providing basic anonymity with minimal delay compared to other systems that incur higher delays.

Keywords: Anonymizer, onion routing, Tor

1 Introduction

Anonymizers are services that enable users of the Internet to browse the web anonymously. They allow a user to maintain a level of privacy that prevents the collection of identifying information such as the IP address while surfing on the web. Anonymizers are an offspring of mix networks that use a chain of proxy servers to create hard-to-trace communications [4]. These anonymity services are provided by either commercial companies driven by subscription fees, non-commercial organizations profiting from advertising, or home-brewed services through open source anonymous tools. Community contributed systems include The Onion Router (Tor) [6], the Invisible Internet Project (I2P) [1], and the Java Anon Proxy (JAP) [2].

Anonymity is defined as a state in which an agent is not identifiable within an anonymity set [12,15,17]. The anonymity set is a system of senders, receivers, and servers in the communication network. Anonymity is a combination of both *unidentifiability*, i.e., observers can not identify any individual agent, and *unlinkability*, i.e., observers can not link an agent to a specific message or action.

** Equally contributing authors.

Anonymity has always been a dichotomous issue in both social life and cyber space. Anonymity technologies have been used for criminal purposes as well as legitimate purpose. On one side, anonymous technologies provide legitimate usages such as privacy and freedom of speech, anti-censorship, anonymous tips for law enforcement, and surveys such as evaluation and feedback. On the other side, anonymous technologies provide protection to criminals in facilitating online crimes such as spam, piracy, information and identity theft, cyber-stalking and even organizing terrorism. Additionally, they may be utilized for Internet abuse for bypassing the Internet use policy of an organization, exposing organization to malicious activities, abusing organization resources, and prevent web filters from monitoring.

Anonymizer systems send data packets over randomly chosen relays so that no single system has information about both the sender and the receiver. Since many users use these intermediaries at the same time, the Internet connection of any one single user is hidden among the connections of all other users. Hence, no individual system, internal or external, can determine which connection belongs to which user. Anonymity research remains a very active area where investigators have focused on anonymous communication, traffic analysis, provable shuffles, anonymous publications, private information retrieval, formal methods, communication censorship, and traffics [5, 7, 12].

In this paper, we analyzed usage of popular anonymity systems including anonymity proxy servers, remailers, JAP mix network, I2P and Tor. For this study, we collected the server lists of each technology and looked up the geolocation of servers. During our exploration, we identified 1,441 anonymity proxy servers, 15 remailers, 11 JAP mixers, 483 I2P relays, and 10,510 Tor relays. We observed that U.S. and Germany were among the top 5 server providers for proxy, Tor and I2P systems and additionally France and Russia were among the top 5 for Tor and I2P systems.

We then performed a detailed analysis of Tor system, the most popular anonymity system, by setting up two servers to analyze Tor usage. During the experiment our servers relayed 150GB of traffic. In this experiment, we observed that relays from Germany and U.S. contribute most bandwidth resources to Tor system and that they have the highest number of Tor users.

Finally, we analyzed anonymity systems from service perspective by analyzing spam e-mail and peer-to-peer client sources of recent data sets. In spam data, we observed e-mails sent through commercial anonymizer services such as GoTrusted. Moreover, we found that proxy servers are used more than other anonymity techniques by spammers and peer-to-peer users to hide their IP addresses. We believe this is due to proxies providing basic anonymity with minimal delay compared to other systems that incur higher delays.

In the rest of the paper, we first analyze well known deployed anonymity systems in Section 2. In Section 3, we analyze the usage of Tor anonymity system in depth. In Section 4, we analyze anonymity system usage in different networks. Related work is discussed in Section 5. Finally, we provide our conclusion in Section 6.

2 Analysis of Anonymization Techniques

There are many categories of anonymity systems. From a usability point of view, anonymous communication can be classified in two categories: *high latency systems*, mostly used by email anonymity that provide strong anonymity, and *low latency systems*, mostly used by anonymous web browsing that have better performance. Other categories can be based on trust level, network type, anonymity properties, or adversary capability.

In this section, we review well-known deployed anonymity systems and provide geographic distribution of their servers.

2.1 Proxy Server

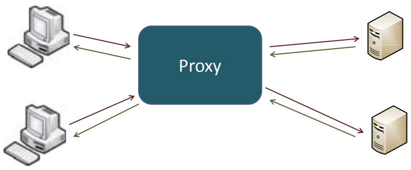


Fig. 1. Proxy Server

A proxy server is the easiest to deploy anonymity system mostly used for low latency browser anonymity [7]. The basic idea behind a proxy server is that a client uses a proxy server to surf the web as in Figure 1. The proxy server performs client requests using the proxy server's identity rather than the client's real identity. Proxy servers relay requests from users to their destinations and deliver responses to the users. Anonymous proxy servers hide the user's IP address and other identifying information to provide basic anonymity. However, these servers are aware of both the source and the destination, and hence can trace user activities. Moreover, they have the weakest security against observers as monitoring in and out traffic of such a proxy server provides a high level of information about its users.

Figure 2 represents the geographic location of 1,441 public proxy servers obtained during Oct 11-17, 2010 from `proxy.org`, `publicproxyservers.com`, `proxy4free.com`, `freeproxy.ru`, and `tech-faq.com`. Note that, the figure is logarithmic scale. Among the available public proxy servers from 88 countries, most were located in the U.S. (i.e., 438) and in China (i.e., 250). Moreover, only 19 countries hosted more than 10 public proxy servers and 28 hosted a single server. These proxies were collected from major announcement lists and are a sample of available public proxy systems. Hence, this is not a complete list of public proxy servers but a representative sample.

In addition to volunteer-based systems, several *commercial anonymizer networks* such as `Anonymizer.com` and `GoTrusted.com` provide anonymous Internet access service to their clients. In these systems, clients pay a subscription fee to be able to relay their traffic through servers operated by the company. Usually, the user is connected to the network through a VPN tunnel and all traffic flows through the tunnel. However, as these companies are in charge of all the communications, they provide a lower degree of protection to their clients.

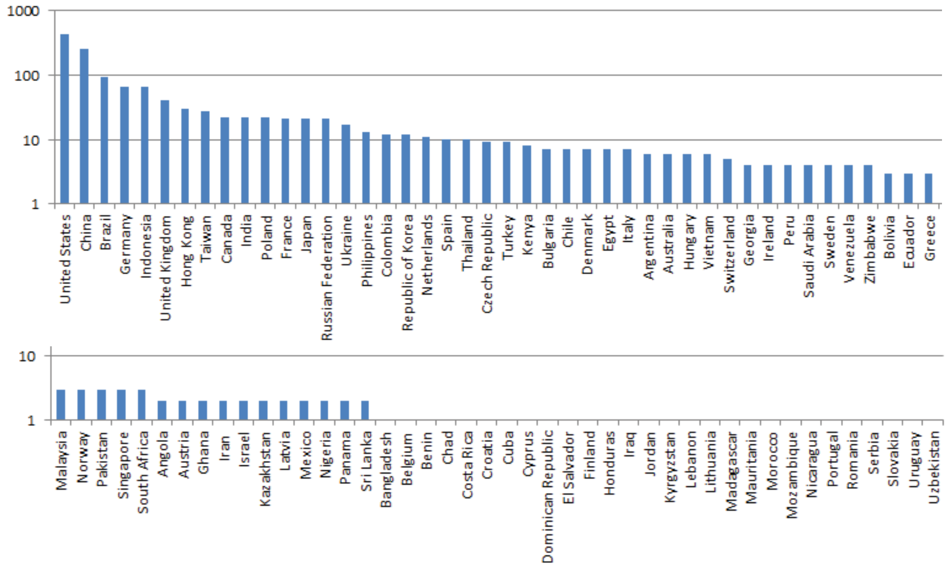


Fig. 2. Geographic Proxy Distribution (log-scale)

As proxy servers provide general web communication, *remailers* enable users to send electronic messages through their servers so that senders can not be traced. Remailers typically remove all identifying information from e-mails before forwarding them to their destination. Known examples of remailers include Cypherpunk, Mixmaster, and nym servers. However, due to heavy use of these servers by spammers in the past, they are not actively deployed any more. During our extensive web/blog search on Oct 2010, we were able to identify only 15 active remailers shown in Figure 3.

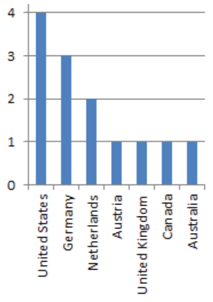


Fig. 3. Remailer Geo-Distributions

2.2 Mix Network

The building block of most of the current high-latency anonymity systems is the mix [4]. The basic building block of these systems, shown in Figure 4, is a set of mix processes where each mix process takes ciphertext messages that are encrypted with the mix process's public key as inputs. Mix process groups messages together as a batch and forwards the encrypted messages to the next mix process at certain flush times along with dummy messages.

Messages reach their destination after being forwarded by a set of mix processes through the network. For example in Figure 4, path $\mathcal{P}$ of a message M consists of 3 mix process Mix-1, Mix-2, and Mix-3. The client builds ciphertext C by encrypting message M along with random text R using each mix's public key K. The ciphertext (e.g., $E_1(A_{Mix-2}, R_1{+}E_2(A_{Mix-3}, R_2{+}E_3(D, R{+}M))))$ specifies the exact path the message will take through the mix network. Each mix node

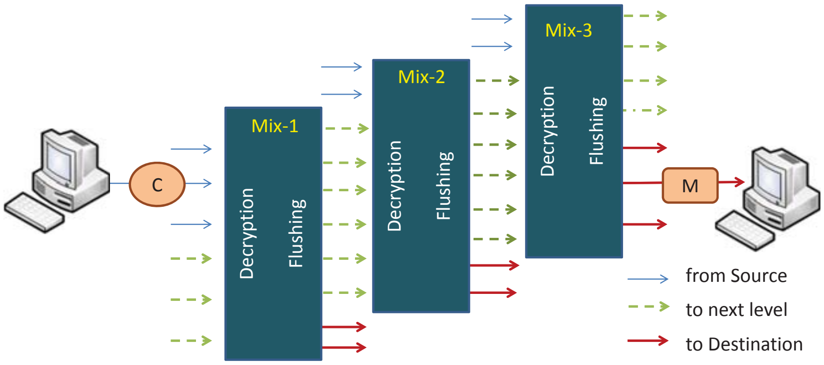


Fig. 4. A Mix Process

(e.g., Mix-1) receives the ciphertext decodes one layer to find next hop destination (e.g. A_{Mix-2}) and forwards payload (e.g., $E_2(A_{Mix-3}, R_2+E_3(D, R+M))$).

Asymmetric encryption and the flushing algorithms are the key for anonymity level and performance of a mix network. As encryption algorithms are provably secure with the current technology, flushing algorithms are an important component that may expose identity of the users. Flushing algorithms buffer incoming messages into a *pool* and forward messages in rounds. At each round, a random subset of the pool messages are mixed with dummy messages and flushed. The random subset can have a constant number or a dynamic number of messages. The duration of each round is decided based on a *threshold*. The threshold can be a number of messages N in the pool, or a timer counter T, or a combination of both.

The *Java Anon Proxy* (JAP) is a mix network that uses servers provided by volunteers, usually institutions that declare conformance to JAP policies, to browse the Internet [2]. JAP cascades encrypted packets through several mixes and keeps the traffic in a constant rate to avoid rate-based traffic analysis. The JAP program displays active mixes and users are able to select JAP cascades from those active mixes. Figure 5 presents the geographic location distribution of 11 JAP servers that were active on 12-19 Oct 2010. Compared to onion routing based systems Tor and I2P, JAP seems to have minimal usage at the time of our analysis.

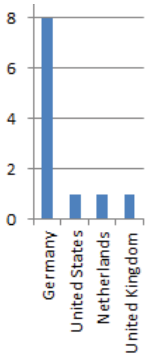


Fig. 5. JAP Geo-Distribution

2.3 Onion Routing

Onion routing is a low latency anonymous communication approach and is currently considered the most prevalent anonymization system design [10]. The basic idea of onion routing is similar to the mix system but performance is improved by using symmetric keys for relaying messages and asymmetric keys to establish circuits in the system.

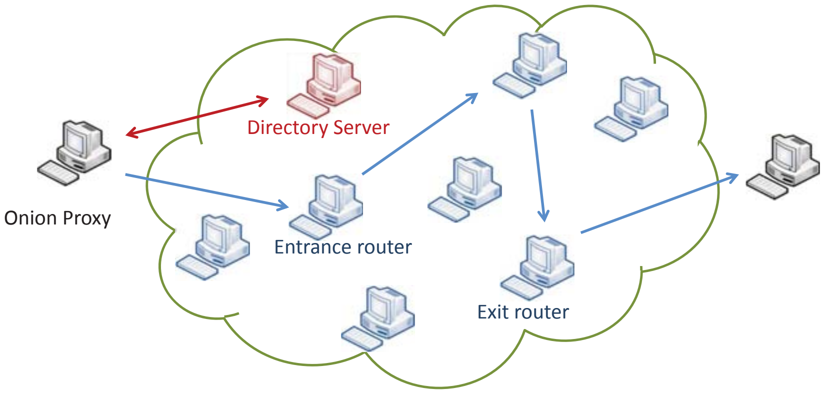


Fig. 6. The Onion Router (Tor) communication

There are different variations of onion routers such as Crowds [18], Tarzan [9], Invisible Internet Project (I2P) [1], and The Onion Router (Tor) [6] based on how the routing servers are organized; how the encryption algorithms are applied; how the tunnels are established; whether the transport-layer protocol uses TCP or UPD; or whether the clients relay traffic to other clients or not.

Tor, shown in Figure 6, is the most popular design as it combines the best parts of previous methods (e.g., the directory discovery of routing servers for clients, telescopic circuit establishment, and hiding locations). Directory servers are responsible for distributing signed information about known routers in the network [7]. Authoritative directory servers, currently 7 systems trusted by Tor developers [11], determine three-hop paths among volunteer servers using secured TCP connections. User messages are then encrypted as in mixes and forwarded through the established circuit to the dedicated exit router, which forwards the message to the final destination and echoes replies back. Entrance and exit nodes are particularly important as they know the source and the destination of the communication, respectively. Hence, authoritative directory servers pick only a subset of existing systems, which seems to be reliable, to become entry nodes and protect client profiling. Moreover, packets originate from the exit system from the destination's perspective and may be questioned regarding user actions. Hence, Tor allows relay systems to not become an exit node.

Figure 7 presents a snapshot of Tor servers based on their geographic location during Oct 20-24, 2010. For this analysis, we monitored the authoritative directory servers to determine the total number and geographical location of Tor servers. During the sampling period, we identified 10,510 unique servers at 95 countries but Tor system has approximately 1,500 active volunteers at a given time. Most of Tor relays are located in few countries. Similar to earlier studies [3,14], Germany and U.S. had highest number of volunteers. Considering continents Europe had the highest number of servers. Interestingly, among Asian countries, Iran was third after technologically advanced countries such as Russia and Taiwan.

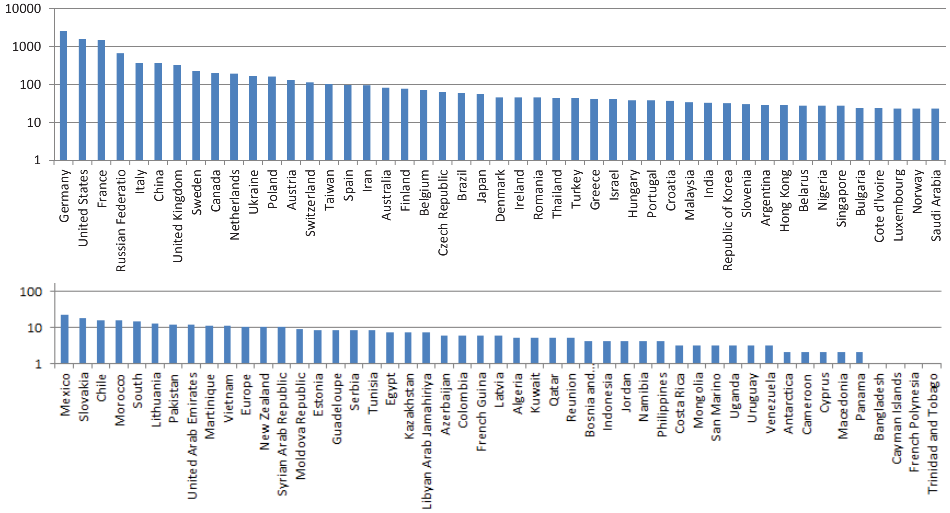


Fig. 7. Geographic Tor Server Distribution (log-scale)

Similar to Tor, the *Invisible Internet Project* (I2P) offers anonymization services that identity-sensitive applications can use. The I2P network is strictly message based, i.e., UDP, but there are libraries that allow reliable streaming communication on top of I2P network. Many applications can interact with I2P including mail, peer-to-peer, and IRC chat. Different from Tor, I2P does not focus on end-to-end delay and is preferred for peer-to-peer applications. To analyze its usage, we collected active I2P relays by joining the system during Oct 11-17, 2010. Figure 8 presents the geographic distribution of 483 servers in 29 countries (origin countries were determined by performing AS look-up of server IP addresses). Even though we had a longer sampling of I2P, we observed fewer servers than Tor system. Moreover, similar to Tor, Germany, U.S. and France had the highest number of volunteers.

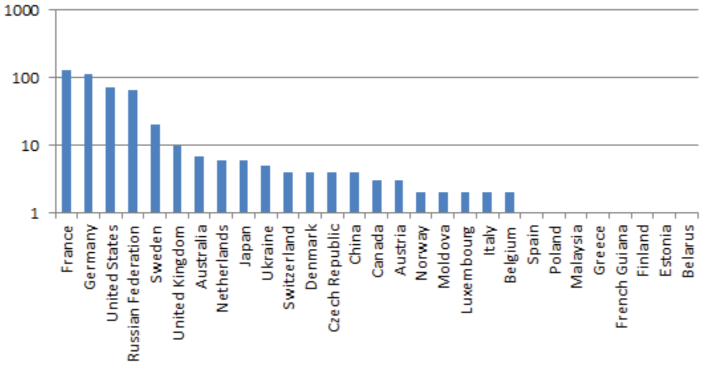


Fig. 8. Geographic I2P Server Distribution (log-scale)

3 Tor Usage Analysis

In this section, we analyze usage of Tor, currently the largest anonymity system. To be able to understand Tor network traffic, we set up two Tor relays using Tor 0.2.2.15- alpha. In order to analyze the traffic passing through our nodes, we used Wireshark to capture packet headers, i.e., IP addresses and port numbers for both source and destination, and payload size. During Oct 20-24, 2010, we had approximately 150 GB of data passing through our relays. According to the authoritative directory servers that provide bandwidth usage of each relay, our nodes were among the most popular relays of Tor in terms of bandwidth utilization.

Moreover, we inspected both incoming and outgoing traffic to observe whether our nodes were entry and exit routers. We observed client IPs when our relays were designated as entry nodes. Looking at IP addresses, we were able to identify the system we were communicating with. If the IP was not among Tor relay nodes, it either belonged to a user or to a server that users were communicating with. In order to distinguish between both, we looked at the payload size as Tor traffic is segmented into cells of 512 bytes. If the payload was 512 bytes that, the packet belonged to a user. Otherwise, the packet belonged to a server users were communicating with.

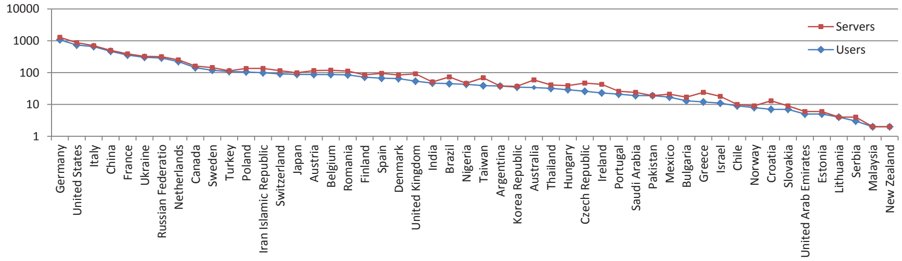


Fig. 9. Tor Usage (log-scale)

As part of our study, we also identified the geographical locations of clients and Tor relays. Table 1 and Figure 9 presents the number of Tor users and the relay servers from these countries. During a day period, when one of our servers was designated as an entry node, we observed **5,932** unique client IPs. According to the usage information we observed, Germany had the highest number of clients using Tor network and hosted most of the relays (similar to what was reported

Country	Germany	U.S.	Italy	China	France	Russia	Netherlands	Canada	Sweden	Turkey
Users	1,076	734	657	469	356	289	223	143	119	108
Servers	205	141	42	29	32	27	29	18	25	6
Usage	5.48	.92	7.28	.36	2.64	1.60	5.01	.17	4.66	1.01

Table 1. Geographical distribution of Tor servers and clients

in [14]). Moreover, we analyzed the usage ratios of observed countries. For this, we obtained the number of Internet users from `http://internetworldstats.com` and estimated the percentage of Tor usage in each country. Interestingly, Italy has the highest ratio of Tor usage relative to its Internet users.

During our data sampling, we also took snapshots of the authoritative directory servers to observe relay bandwidth. On average 1,567 Tor routers were observed to be active. Figure 10 presents the average contribution ratios of different countries in terms of total bandwidth, which was computed as the sum of all bandwidths of relays from a country.

Finally, to model the probability of each router forwarding a particular packet, we analyzed Tor relay usage from our nodes by counting the number of relay IPs. For an hour of traffic, we observed that 2% of relays carry 30% of traffic. Among the 15 most popular routers, 8 were in Germany, 4 in United States, 2 in France and 1 in Sweden. This indicates the disproportion of traffic carried by Tor servers and may weaken user anonymity [8].

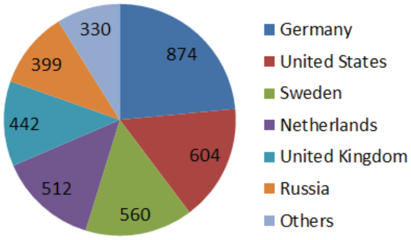


Fig. 10. Tor Bandwidth Distribuiton

4 Service Perspective

In this section, we investigate the usage of anonymity technology from a service perspective. These service applications include a secure web site at a university, spam emails, and peer-to-peer network. In total, 195,919 unique IP addresses were observed and analyzed to understand whether they originated from an anonymity system. For this, we compared the observed IP addresss to the collected IP addresses of anonymity servers in Section 2. Table 2 provides an overview of all the anonymity systems we looked at. The originating countries of these IP addresses were found using AS lookup.

We collected the IP addresses of systems that accessed a *secure web site* from log files of more than 1 year. In this data, we had more than 21K unique IP addresses but there was no IP address from an anonymity server. This is expected because the secure web page requires login information and use of anonymizer would not improve anonymity of the user.

Network	Tor	I2P	JAP	Remailers	Proxies	Commercial
Servers	10,387	483	11	15	1,441	Anonymizer, GoTrusted
Service	General	peer-to-peer	General	E-mail	General	General

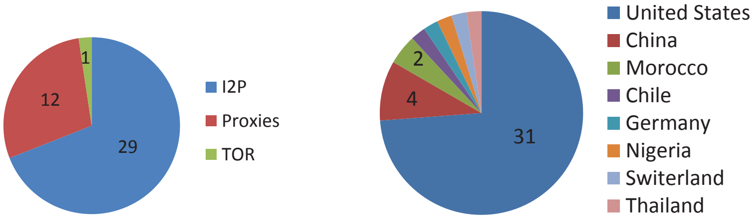
Table 2. Analyzed Anonymity Systems

The following subsections provide our findings about spam e-mails and peer-to-peer traffic.

4.1 Spam mail

Spam email data was collected using two approaches. First, we collected IP addresses of spam emails from Gmail accounts of coworkers and from departmental email servers during Oct 2010. Second, we gathered publicly available spam email IP addresses of recent spammers from the Internet. An important issue was to obtain recent data sets as anonymizer server IP addresses change over the time (except for commercial systems). As explained below, most spam e-mails were sent through relays in China and U.S. which is consistent with [16].

Gmail data set: We collected 4,843 IP addresses of spam e-mails from Gmail accounts of co-workers during Oct 2010. In this data, 42 IP addresses belonged to anonymity servers corresponding to 0.87 % of spam e-mails being sent through an anonymity network. Figure 11 presents the distribution of the anonymizer technology and the server geo-location. In this data set, I2P was utilized as spam relay more than the other sytems.

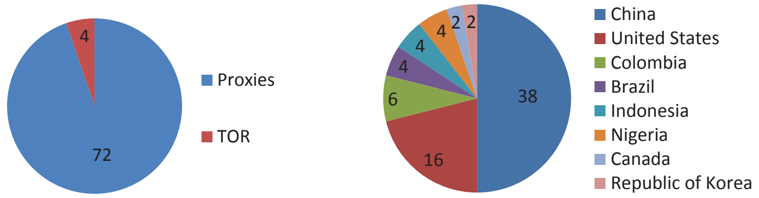


a) Network b) Location

Fig. 11. Gmail Spam

Departmental data set: We collected 11,402 IP addresses of e-mails that were marked as spam by the departmental mail servers during Oct 2010. Among these IP addresses, only 76 were identified to arrive through an anonymity network corresponding to 0.67 % of total departmental spams. Figure 12 presents the distribution of utilized anonymizer technology and the server geo-location for departmental spam that was sent through an anonymity system. Similar to Gmail spam data, China and U.S. were the top two. In this data set, proxies and Tor network were utilized in sending spam e-mails.

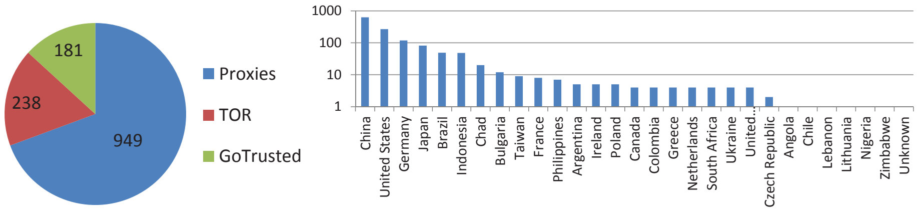
Public data set: We collected 30,959 IP addresses that were recently marked as spam generators by public systems including `projecthoneypot.org`, `ipdeny.com`, `aclweb.org`, `landfall.net`, `spam-ip.com`, `spam-ip-list.blogspot.com`, and `spamlinks.net`. Among these IP addresses, 1,368 belonged to an anonymity



a) Network b) Location

Fig. 12. Departmental Spam

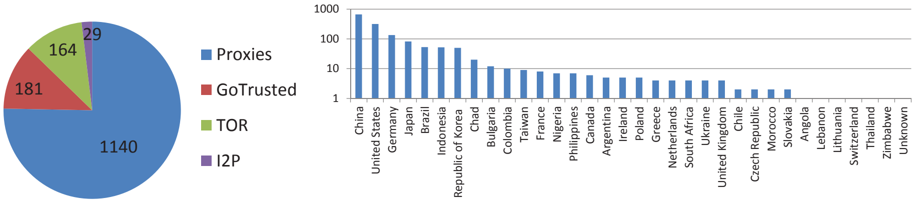
server corresponding to 4.42 % of all spammer IPs. Figure 13 presents the distribution of utilized anonymizer technology and the server geo-location for spammer IP addresses in the data set. Similar to earlier data sets, China and U.S. were the two major relay nodes for spammers among the 31 countries observed and account for 65.4 % of all servers. In this data set, we observed that Proxy and Tor servers were utilized the most. Interestingly, the commercial anonmizer system `goTrusted.com` was utilized by spammers to send e-mails.



a) Network b) Location

Fig. 13. Public Spam Data

All data sets: Figure 14 presents the results of all data combined (i.e., Gmail, department and public spam email data sets). Overall, proxies, `GoTrusted` and Tor were the three major sources utilized by spammers to relay e-mails. Moreover, servers in China, U.S. and Germany were the main relays of spam e-mails.



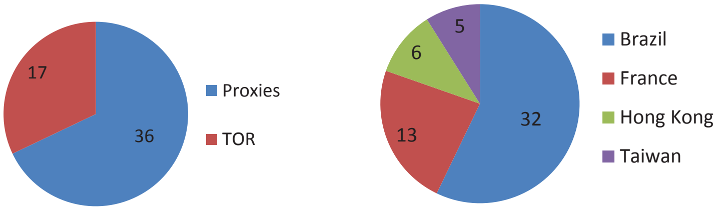
a) Network b) Location

Fig. 14. Combined Spam Data

4.2 Peer-to-peer data

In order to analyze peer-to-peer traffic for anonymizer technology usage, we modified the open source Shareaza client, which joins BitTorrent, eDonkey, Gnutella, and Gnutella2 networks. The code was modified to log connected IP addresses and automatically search 3,600 keywords that were Google trends on Oct 2010 for about 50 countries. Considering copyright and other legal issues, the download feature was disabled so that no files were actually downloaded to our systems. We gathered data from two systems during Oct 10-24, 2010. In total, 114,593 unique IP addresses of peer-to-peer users were observed and analyzed.

Shareaza data set: Among the 114,593 IP addresses observed during our data collection, only 53 belonged to an anonymity system. Compared to the spam e-mail data set, this was very small. We believe that the main reason for this is the delay incurred by the anonymity system. Figure 15 presents the anonymity technology and geo-location distribution of the servers for the identified anonymizer relays. We observed that only Proxies and Tor servers were utilized by peer-to-peer clients to hide their IP addresses. Even though our peer-to-peer clients were in the U.S., only servers in Brazil, France, Hong Kong and Taiwan became relays to connect to our nodes. Among the 114,593 IP sources, United States and China accounted for most of them, but none of those utilized an anonymity network.



a) Network b) Location

Fig. 15. Peer-to-peer data

Finally, within two weeks of data collection, we received a high ratio of bad queries among peer-to-peer client messages. These bad queries may be due to encrypted or compressed messages as reported by Chaabane et al. [3].

5 Related Work

There have been many studies on anonymity and anonymous systems and three studies have analyzed Tor usage as it gained popularity [3, 13, 14].

McCoy et al. looked for answers on how Tor is being used, how it is being mis-used, and who are its users [14]. In their experiments, the authors analyzed application-level protocols that use their nodes as exit node. According to their finding, interactive protocols, such as HTTP, make up 92 % of the connections and 58 % of bandwidth. Similarly, bit-torrent traffic consumes 40% of bandwith even though it accounts for 3.3 % of the connections. The authors also pointed to malicious usage of Tor routers and developed a method to detect malicious logging at exit routers. Moreover, they indicated that Tor has a global user base based on client distribution. Our results in Section 2 also indicate that Tor has the largest volunteer base among anonymity systems.

Moreover, Chaabane et al. performed a study to analyze applications that use Tor [3]. Authors monitored traffic on six servers which were pairwise located in U.S., Europe and Asia to inspect geo-diverse relays. Authors analyzed HTTP and BitTorrent traffic in detail. They pointed out that BitTorent consumes significant resources both in terms of packets and traffic size. Finally, authors pointed that Tor servers are used as 1-hop SOCKS proxies and present a method to detect such misuse.

Loesing et al. provided guidelines for a statistical analysis of Tor data focusing on countries of connecting clients and exiting traffic by port [13]. Pointing to privacy issues the authors derived guidelines for measuring sensitive data in anonymity networks. Moreover, they pointed to interesting cases such as increase in Tor usage by Iranian IP space in June 2009 after the Iranian elections; Tor blocking by China and consequent increase in bridge usage by Chinese IP addresses.

Our study is different from previous studies in that, in addition to Tor network analysis, we presented the analysis of other active anonymizer systems. We pointed out their usage and server geo-location distributions. Furthermore, we analyzed the traffic from different networks including a secure website, spam e-mails and peer-to-peer network. These studies allowed us to measure anonymizer usage in different domains.

6 Conclusion

Anonymity technologies have been utilized for a while. It is important to understand how people are using them, what applications are being used and which anonymity technology is popular. In this paper, we first summarized various anonymity technologies, i.e., proxy servers, mix networks and onion routing, and then focused on widely deployed anonymity systems, i.e., proxy servers, remailers, JAP, Tor, and I2P. For analyzing the current state of anonymizer networks, we joined them and collected information about relay nodes. We observed that similar countries, e.g., U.S., Germany and China, have the highest number of servers in different anonymizer networks.

Moreover, we set up Tor nodes as clients to collect entry and exit traffic information. Our servers relayed 150GB of data over five days. We observed that countries with high number of servers tend to have high number of Tor users. For instance, Germany and U.S. are top both in number of server and number of clients. Furthermore, to understand anonymity technology usage in different domains we analyzed spam emails and peer-to-peer clients. We observed that proxy servers were deployed more than other technologies. We believe that this is due to the higher latency in more secure systems.

Acknowledgments: This work was supported in part by National Institute of Justice.

References

1. I2p anonymous network. www.i2p2.de.
2. O. Berthold, H. Federrath, and S. Kpsell. Web mixes: A system for anonymous and unobservable internet access. In *International Workshop on Designing Privacy Enhancing Technologies*, pages 115–129. Springer-Verlag New York, Inc., 2001.
3. A. Chaabane, P. Manils, and M. Kaafar. Digging into anonymous traffic: A deep analysis of the tor anonymizing network. In *Network and System Security (NSS), 2010 4th International Conference on*, pages 167 –174, 2010.
4. D. L. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Commun. ACM*, 24(2):84–90, 1981.
5. G. Danezis and C. Diaz. A survey of anonymous communication channels, 2008.
6. R. Dingledine, N. Mathewson, and P. Syverson. Tor: the second-generation onion router. In *SSYM'04: Proceedings of the 13th conference on USENIX Security Symposium*, pages 21–21, Berkeley, CA, USA, 2004. USENIX Association.
7. M. Edman and B. Yener. On anonymity in an electronic society: A survey of anonymous communication systems. *ACM Comput. Surv.*, 42(1):1–35, 2009.
8. N. Feamster and R. Dingledine. Location diversity in anonymity networks. In *Proceedings of the 2004 ACM workshop on Privacy in the electronic society*, WPES '04, pages 66–76, New York, NY, USA, 2004. ACM.
9. M. J. Freedman and R. Morris. Tarzan: a peer-to-peer anonymizing network layer. In *CCS '02: Proceedings of the 9th ACM conference on Computer and communications security*, pages 193–206, New York, NY, USA, 2002. ACM.
10. D. Goldschlag, M. Reed, and P. Syverson. Onion routing. *Communications of the ACM*, 42:39–41, 1999.
11. S. Hahn and K. Loesin. Privacy-preserving ways to estimate the number of tor users. Technical report, TOR project, November 2010.
12. D. Kelly. A taxonomy for and analysis of anonymous communications networks. Technical report, Air Force Institute of Technology, March 2009.
13. K. Loesing, S. J. Murdoch, and R. Dingledine. A case study on measuring statistical data in the tor anonymity network. In *Workshop on Ethics in Computer Security Research*, Jan 2010.
14. D. Mccoy, T. Kohno, and D. Sicker. Shining light in dark places: Understanding the tor network. In *In Proceedings of the 8th Privacy Enhancing Technologies Symposium*, 2008.
15. A. Pfitzmann, T. Dresden, and M. Hansen. Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management a consolidated proposal for terminology, 2008.
16. A. Ramachandran and N. Feamster. Understanding the network-level behavior of spammers. In *Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*, SIGCOMM '06, pages 291–302, New York, NY, USA, 2006. ACM.
17. M. K. Reiter and A. D. Rubin. Crowds: anonymity for web transactions. *ACM Trans. Inf. Syst. Secur.*, 1(1):66–92, 1998.
18. M. K. Reiter and A. D. Rubin. Anonymous web transactions with crowds. *Commun. ACM*, 42(2):32–48, 1999.