The Dark side of I2P, A Forensic Analysis Case Study

Behnam Bazli, Maxim Wilson, William Hurst

Behnam Bazli is a Lecture in the School of Computing and Digital Technologies, Staffordshire University, UK. His research interests include, Network Security, P2P Networks, Digital Forensics and Internet of Tings (IoT). Email: B.Bazli@staffs.ac.uk

Maxim Wilson is a graduate and research in the School of Computing and Digital Technologies, Staffordshire University, UK. His research interests include P2P networks, Digital Forensics and Security Programing. Email: M.Wilson@out-of-hours.it

William Hurst is a senior lecture in the Department of Computer Science in Liverpool John Moores University, UK. His research interests include Big Data Analysis, Data Visualisation, Machine Learning, Critical Infrastructure Protection and Computer Animation. Email: W.Hurst@ljmu.ac.uk

Abstract

File sharing applications, which operate as a form of Peer-to-Peer (P2P) network, are popular amongst users and developers due to their heterogeneity, decentralised approach and rudimentary deployment features. However, they are also used for illegal online activities and often are infested with malicious content such as viruses and contraband material. This brings new challenges to forensic investigations in detecting, retrieving and examining the P2P applications. Within the domain of P2P applications, the Invisible Internet Project (IP2) is used to allow applications to communicate anonymously. As such, this work discusses its use by network node operators and known attacks against privacy or availability of I2P routers. Specifically. We investigate the characteristics of I2P networks in order to outline the security flaws and the issues in detecting artefacts within the I2P. Furthermore, we present a discussion on new methods to detect the presence of I2P using forensic tools and reconstruct specificI2P activities using artefacts left over by network software.

Keywords; P2P Networks, I2P Artefacts, Security, Forensics Analysis

1. Introduction

Self-organising overlay networks, which are distributed on IP networks, are called P2P networks. They are scalable platforms and popular for file sharing due to the level on content-distribution platforms available [1] 2P file-sharing networks reflect the Internet of Things (IoT) paradigm, with autonomous networked devices within distributed and decentralised systems. The topology of P2P networks is advantageous to a traditional client-server approach, in that they are self-scalable [2] and the amount of data available is proportional to the number of participants. Generally, P2P networks are managed by protocols implemented at the application level on top of the User Datagram Protocol (UDP) or Transmission Control Protocol (TCP). Furthermore, P2P overlays provide support for scalability within dynamic and decentralised systems. The nodes within a P2P system act in a self-managing manner in contrast to the client-server model. Such overlay networks go beyond the services offered by conventional client-server systems [3]. P2P systems are pervasive and largely used for file sharing and data communication.

While the rapid growth and ubiquitous use of file sharing applications is generally positive for users, it results in many challenges in forensic investigations: especially as the IoT introduces further smart devices, which contribute to the volume of network traffic [2]. This is further exacerbated by the continuous dynamic changes of membership, the geopolitical stance on copyright materials and the legal and ethical issues in dealing with file sharing applications. However, the most challenging issue is overcoming the misuses provoked by capabilities of P2P network, specifically within the I2P environment. Many law enforcement agencies struggle to keep up with the new tools and techniques, which are misused by P2P/I2P users, who contribute to and enable illegal activities online.

For that reason, in this work, we examine the characteristics of I2P networks and related applications that attract illegal activities and pose a problem for the forensic analyst. Furthermore, we propose a alternative approaches for the identification and reconstruction of a suspect's activity on an I2P network and the analysis of the remaining artefacts. To accomplish this, a combination of custom made and industry approved forensic tools are employed.

This paper is organised as follows. Section 2 introduces the I2P network, its current developments and the challenges it presents in forensic investigations. Section 3 outlines solution design and description of forensic procedures in I2P investigations. Section 4 includes related works and section five provides discussion and future direction of the research.

A. Network Overlay

Network overlay provides support for scalability within a dynamic and decentralised system with self-managing nodes [4]. This means they can take advantage of the available resources, content and traffic stability independent of central servers. Nodes have dual client and server roles and can both initiate and listen for incoming connections.

Overlay networks operate on top of another network (referred to as the "base") and deliver additional functionality, which is not offered by the base network. Since overlay networks avoid direct interaction with underlying hardware, they can be rolled out to interested users without a costly upgrade of the infrastructure or interruptions to the base network services.

There are currently three mature overlay networks, which use the Internet as a base layer and improve on it by adding privacy features namely; Tor, Freenet and I2P. These networks are open-source and were released to public more than 10 years ago and still in active development. This work, however, focuses on investigating the I2P network due to its potential for future growth and public perception of it being the most secure of the three solutions.

B. P2P Networks

A network overlay is a solution which addresses the scalability issues within distributed systems [5]. It is a virtual network of nodes and logical links built on top of the existing network infrastructure. It can, therefore, be used to deliver additional services and functionalities which are not offered by the base network. Since overlay network avoids direct interaction with the underlying infrastructure, it can be deployed without costly upgrades or interruptions to the base network services. Furthermore, it does not require modification of existing software or protocols in order for new nodes to join the overlay network.

A P2P overlay provides support for scalability within a dynamic and decentralised system with self-managing nodes. This means all nodes contribute to and benefit from shared pool of network resources and content, without being reliant on any central server. Yet, P2P networks with diverse properties classified based on different methods, such as, performance metrics, topology, protocol and structure [6]. P2P overlays are popular amongst users for file sharing and communication such as Skype[1], BitTorrent[2] and Freenet[3]. Each class of the system has its own advantages and disadvantages, but the focus in this paper is on P2P overlays that offer some degree of anonymity to their users. Anonymity is the main attribute that provides user privacy. This feature is for illicit activities which are the result of sharing copyrighted materials, illegal transactions and general cybercrimes.

C. I2P Network

The I2P is an adaptation from Kademlia [7], which was originally developed to go a step further than just anonymity and enables users to be within invisible spaces called a 'Darknet'. I2P provides a P2P communication channel, along with various protocols and encryption standards to maintain user anonymity. The end-to-end communication between two users is not globally advertised. Furthermore, it is fully encrypted. I2P improves on standard TCP/IP communication model by ensuring that IP (Internet Protocol) packets exchanged between participating hosts always contain encrypted data. Instead of relying on IP addresses to uniquely identify hosts and route traffics, I2P introduces its own identifiers and routing logic at higher layer of the protocol stack. As long as Layer 4 network connectivity exists between hosts, I2P is able to operate in complete isolation from the rest of the public Internet infrastructure. These enhancements aim to improve user anonymity by reducing the risk of malicious third parties such as a compromised service provider, intercepting or altering the network traffic. These security and privacy enhancements are particularly valued by I2P users living in countries with restrictive policies governing Internet use. However, the same features may pose a

[1] Skype - https://www.skype.com/en/

[2] BitTorrent - www.bittorrent.com/

[3] Freenet - freenetproject.org

problem for law enforcement organisations, therefore making I2P an attractive solution for cybercriminals to safely operate their business. Within this section, we discuss the characteristics of the I2P applications.

a. I2P Routers

I2P nodes communicate using P2P tunnel, which is facilitated by I2P routers. The I2P nodes and routers communicate using a 'garlic cloves-type' tunnelling infrastructure. The end-to-end communication uses a form of Public Key Infrastructure (PKI). However, the inbound tunnels are separated from the outbound tunnels. The sender has no information or knowledge about the outbound routes. As such, the transmitted messages are relayed using the I2P router many times, hiding the user identity completely. The communication between peers has no entry or exit point, preventing any vulnerability existed within similar systems such as Tor [8].

b. I2PSnark

I2PSnark is the default torrent client for the I2P network and is distributed as part of I2P router software [9]. As a native I2P application, I2PSnark does not process standard IP addresses and is, therefore, unable to communicate over normal Internet. This limitation is intentional and ensures that no personally identifying P2P traffic can leak outside of encrypted I2P tunnel. Security and ease of use ensure continued popularity of I2PSnark with I2P network users. I2PSnark user base is larger than that of all other I2P torrent clients combined, with I2PSnark being responsible for one-third of total I2P network traffic.

P2P (BitTorrent) clients operating over normal Internet provide any member of the torrent swarm with information about all other peers. A forensic examiner can therefore obtain the torrent creator's location and IP address, which would then be followed through courts and the Internet Service Providers to determine the identity of illegal file-sharer. Applying this method to I2PSnark yields only a list of peers' I2P network identifiers, which have no forensic value.

c. Domain Name Resolution

I2P is fault tolerant and is designed to withstand accidental or deliberate failures of Internet public services. An example of such a public service is the Domain Name System (DNS), responsible for translating human-readable domain names into their matching IP addresses. The Internet's public DNS infrastructure is strictly hierarchical, with top level domains (e.g. ".com") being maintained by a neutral, non-commercial organisation Internet Corporation for Assigned Names and Numbers (ICANN). Lower level (e.g. ".co.uk") domains are controlled by hosting companies, smaller business organisations or individuals. This hierarchy is demonstrate in Figure 1 as follows.

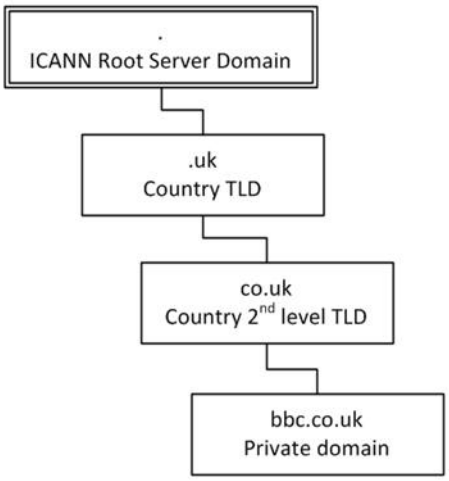

Figure 1 - Internet has strict registrar hierarchy, which is absent in I2P

An Internet-connected host can resolve these domain names by making iterative queries to every DNS server in the hierarchical chain. This method ensures that final response received is authentic, accurate and up to date. However, it is inefficient and scales poorly to multiple hosts. Resolution is delayed by awaiting for replies from several servers. In addition, several hosts querying the same domain name are performing the same task repetitively. Common alternative involves offloading this task to single DNS server, which will perform iterative queries once requested by the host. This setup is common in both home and enterprise environments, where internal hosts send all their queries to caching DNS server maintained either by their Internet Service Provider or corporate.

Public DNS infrastructure therefore is built on trust, with each host assuming that the servers within higher level will keep providing accurate data. Its hierarchical structure is susceptible to both technical failure and takeover of domain names by malicious third parties. Such an infrastructure is not suitable for I2P, which, emphasises on reliability and user anonymity. For these reasons, I2P network implements its own system for resolving short, human-readable domain names.

As part of the developed system, every node in I2P network is expected to keep a local "addressbook". The addressbook is a file, which stores the associations between an I2P domain name and I2P network identifier (as a replacement for IP address). The concept is similar to use of hosts file nodes of the early Internet before the introduction of DNS [10].

To reduce the need for manual editing of host files, I2P implements a '*name record update*' mechanism known as '*subscriptions*'. An I2P node can specify several other nodes on the I2P network to be 'subscription' sources that will be regularly polled for their copies of the addressbook. Any entries of the domain names which are not present in the subscribed node addressbook are merged within current copy.

The aim of the described I2P system is to keep as much information required for resolving domain names locally on the I2P node. Therefore, a forensic analyst who manages to obtain name resolution query logs from I2P users' Internet Service Provider, local DNS caching server or DNS resolver cache, is unlikely to find any entries related to the I2P.

d. Darknets

I2P uses its own domain name service, which enables the existence of 'eepsites' also known as 'darknets'. They are considered as hidden websites that can only be accessed by users connected to I2P overlay network [11]. Eepsites are hosted directly on I2P network nodes and are accessed by domain names ending with '.i2p' top-level domain, as shown in Figure 2.

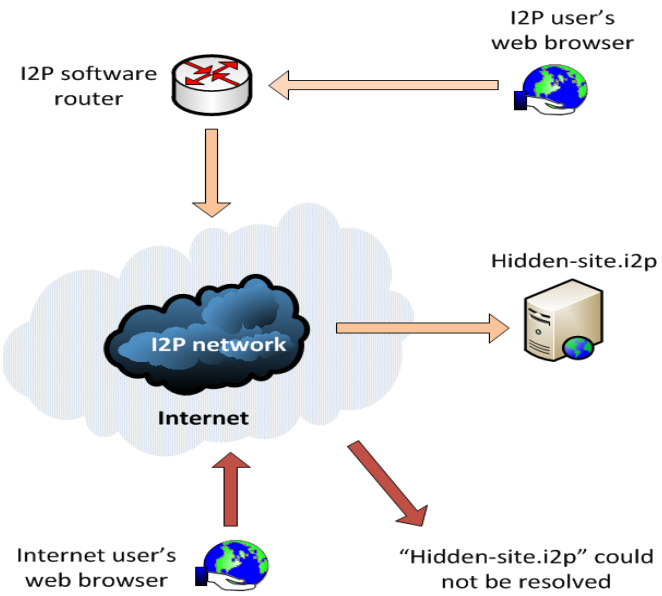

Figure 2 - eepsites cannot be accessed from Normal Internet

Within an investigation of a normal Internet website, the domain registration records, and copy of DNS zone files, can provide several key pieces and of forensically valuable information. These include contact details of the registrar, personal details of the domain owner and mail exchange records providing the IP address of the host. As such, the identity of the website owner can be identified, the offending domain established and the web host server seized for further investigation by the forensic analyst.

Such an approach is not effective against I2P eepsites. Normal Internet registrars are part of DNS hierarchy and therefore encouraged to cooperate with law enforcements under the Internet Corporation for Assigned Names and Numbers (ICANN)[4] scheme. The process of eepsite domain name registration ensures that no personal information or IP addresses are stored by registrar. The I2P domain name registrars are anonymous; they have no governing body and do not face any consequences for ignoring rules, regulations and requests from law enforcement agencies.

The lack of access to hidden eepsites makes them invisible to search engines, such as Google Search cache, and Wayback Machines. Forensic analysts frequently rely on these to prove the content of the

[4] www.icann.org

suspect website at certain point in time. Eepsites, therefore, are less consistent as evidence compared to normal websites because, there is no backup copy for eepsite can be located if it is shut down by its owner.

e. Discovering an I2P Installation

The possible misuses of I2P network is less known among law enforcement agencies and forensic analysts. This may lead to I2P installation on seized machine not being discovered and identified as a source of potentially valuable forensic artefacts.

The industry approved software such as EnCase, Autopsy and FTK have no analysis or detection functionalities for I2P artefacts. As such, they do not provide any evidence of I2P artefact presence in addition, its activity on the system under investigation.

I2P can be installed in one of the two modes on a windows machine; either as an application or as a system service. The system service installations of I2P have more value to forensic analyst. This is because users who require permanent connection to I2P for hosting eepsites or sharing illegal content prefer the I2P installation as service. However, I2P installed as system service is more of a challenge to discover due to lack of entries in Start Menu, Desktop and Most Recently Used (MRU) software list.

2. Forensic Analysis of I2P

To continue the discussion, in this section, we present different methods of forensic investigation into I2P artefacts. We exploit the characteristics and vulnerabilities of I2P activities to propose an inclusive and effective method that can be used within forensic investigation. Such techniques are integrated within existing tools making investigations efficient.

I2P router software focuses strongly on security of network traffic rather than the data stored locally on participating I2P nodes. As a result, the local data is stored without encryption and can be of use to forensic analyst investigating a seized machine. Furthermore, we highlight how the functionalities and flaws within the network should be considered in forensic investigations.

A. Investigation of I2P Installers

I2P installers for Windows family of operating systems contain several layers. The outer layer is a self-extracting archive as 7-Zip format, used to distribute the installer components in a single file and reduce the file size. The inner layer is a 'PACK' file generated by IzPack installer for applications written in Java. Although there is no official unpacker existing for the files generated by IzPack, the structure of the package file slightly resembles that of a forensic image and can therefore be reversed. The IzPack package file contains a general file header, which is followed by files belonging to individual components of I2P router software. Individual component files within the package are designated by header and footer signatures, which also list the component file name, type and intended installation path.

These component files can be extracted with a single script written in a programing language, such as Python, which is compatible with most forensic tools, and then used to either construct a hash set library or manual comparison by forensic analyst.

B. Detection via Known Hash set Library

The individual I2P components extracted from installer files can be used to produce hash set libraries. These libraries can imported into approved forensic software that is currently unable to identify the presence if I2P within evidence. EnCase suite by Guidance Software is one example of forensic tool which is approved for generating legally valid forensic reports, but cannot detect I2P in its default configuration. EnCase, however, supports use of hash libraries containing MD5 and SHA1 hashes of known software. Forensic labs to either filter out known good software or detect known bad content such as illegal images or software with dubious uses use these hash libraries.

EnCase can therefore be equipped for detection of I2P by importing a legacy hash library containing MD5 hashes of I2P components. Some components of I2P are more suitable for this detection that others due to their unchanging attributes. For example, I2P application itself is not a good candidate for hash library, as the hash changes with frequent release of I2P. However, the digital certificates of

I2P developer eepsite are good candidates, as these are present in every I2P node installation and remain unaffected across multiple version releases for years.

C. Comparison of Addressbooks

One of the components of the I2P, which can be extracted from I2P installer, is the copy of the default addressbook. Every new I2P node is provided with the same copy of this addressbook during installation, so that it can access a basic minimal set of trustworthy eepsites. The I2P node is then expected to expand on this minimal addressbook by importing information from its own set of subscription sources and manual addition of eepsite domain name entries.

Forensic analyst can use this minimal default addressbook as a reference to be compared against addressbook found on the seized machine. Entries which are not found in the default addressbook have either been imported via subscription updates or added manually by I2P node owner. Eepsite entries originating from subscription can be further eliminated from this list via inspection of subscription update log , such as records time, source and domain name of imported entry. Through this process of elimination, the addressbook from seized computer can be reduced to set of domain records which are highly likely to have been added by I2P node owner manually for his personal eepsite browsing. This information can be especially useful if the suspect has taken anti-forensic steps to eliminate browsing history and artefacts from his local machine.

D. Takeover of Existing Registrars

Registrars in I2P network do not have to pass through any kind of accreditation or approval process. This makes it possible for any interested party to operate their own I2P domain name registrar node. Although it is possible for law enforcement or forensic analyst to setup their own new registrars on I2P network, some of existing known good registrars may become vulnerable to takeover. The primary candidate for takeover would be registrar known as ‘NO.i2p’. NO is a small registrar compared to developer operated ones like ‘Stats.i2p’, but still occupies a special position in I2P name resolution system.

NO does not share the permissive policy operated by 'rogue' registrars such as INR. Instead, NO shares the same version of registration policy as developer-owned registrars, disallowing illegal or questionable content. Since there are no policy disagreements, NO is one of the few registrars which are considered "trusted" by I2P project team. As a trusted registrar, it is one of only four service choices shown to each I2P user who tries to access website not known to local addressbook.

As of early 2016, the NO registrar appears to have been abandoned by its owner. New domain requests can still be submitted by users through NO's site, but are not reviewed by its operator. The database of the existing domains have not been purged or screened for content violations, since NO still keeps name entries for resources which violate its own terms of registration, such as I2P mirror of Silk Road Reloaded website.

The lack of maintenance together with the trusted status should make NO an attractive target for both law enforcement and malicious parties. Due to the lack of maintenance, NO is currently not receiving security updates for its I2P router software or the web server. An attacker with ability of compromising NO can gain control over a critical part of I2P infrastructure.

The alternative option would be to wait until the failure of NO or its I2P server and resort to social engineering. Most of I2P registrars including NO are running name resolution software known as Py-I2PHosts, which is available for download from its developer eepsite on I2P network. It is therefore possible to recreate NO after its failure, but on different B32 and eepsite addresses. The recreated registrar can be then advertised on I2P community resources as resumed after hardware failure.

Success of this method is possible due to the decentralised structure of the I2P users where, there is no control over the membership. Any user can setup a high-value network service, without any resources. Registrar 'RUS.i2p' was known for hosting I2P knowledgebase and eepsite entries for users located in ex-CIS countries. After several extended outages and restorations of services, this registrar succumbed to a server hardware failure and is no longer available. Another of I2P registrars 'NIC.i2p' lost ownership of its original eepsite domain name and can be reached only through its full network address. Several I2P operators found this incident suspicious and questioned the operator's ability to

run a critical network service. Despite this, the registrar remains operational at the time of this report and is included on unofficial "known good" registrar lists circulating in I2P user community.

E. Mirroring of eepsites

The non-hierarchical model of I2P name resolution system makes it possible for the forensic analyst to create his own mirror of suspect website and register it under the same domain name. Domain names in I2P are unique per registrar only and may be registered to more than one host at the same time as shown in Figure 2.

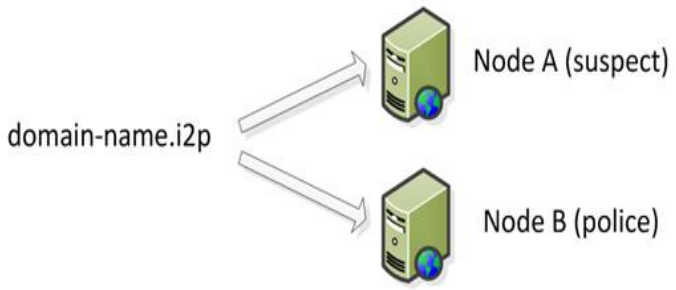

Figure 3 - I2P Short Domain Names Are Unique per Registrar

Due to complex propagation of I2P name updates, it is possible for the existing domain name to remain available on different registrars. For example, domain names registered via I2P known as INR do not always propagate to other registrars due to INR's untrusted status.

This method on penetration should also be considered in relation to one peculiarity of I2P's naming system - persistence of name records. Once the I2P node addressbook is stored, it never expires. The registrar from which this eepsite information was originally retrieved may have since updated the information or purged the domain from its database entirely. However, none of these events will affect an existing addressbook entry. The owner, staff and regular visitors of mirrored I2P site will therefore remain unaffected by intelligence gathering carried out on the false mirror site.

The persistent addressbook entries work in favour of a forensic analyst or law enforcement agency. Eepsite owner and regular visitors are more likely to be security-conscious and very familiar with the "look and feel" of the compromised eepsite. This knowledge increases the risk of one of the visitors

detecting inconsistencies in the false mirror site and alarming other users. In comparison, new or occasional visitors are less likely to be alarmed, since they do not have a reference to compare the mirror eepsite with.

The resulting benefit is that false mirror eepsite can remain undetected for a long period of time, constantly gathering information about activity of new visitors. The longer false eepsite stays operational, the higher is the chance of it trapping one of the regular visitors. This may occur through migration to new device (e.g. to secure virtual machine or a machine with full-disk encryption) without adequate preparation or release of incompatible update to I2P, therefore requiring user to repopulate his addressbook entries.

F. Locating I2P Node by Network Performance

The use of denial-of-service attacks against I2P network has been proposed by Christopher Kack [12]. In attack known as “darkloris”, the malicious I2P nodes keep cyclically opening a large number of connections to service provided by target I2P node. These connections are initiated with the sole purpose of consuming the resources of target I2P node, but are never properly used or terminated by the malicious nodes.

Kack successfully demonstrated the effectiveness of this attack against Jetty web server used by default in new installations of I2P router software. Despite Kack running his attack from a single malicious I2P node, the target node could not handle any incoming connections to its web server, which resulted in all new eepsite visitors receiving an error page as shown in Figure 3.

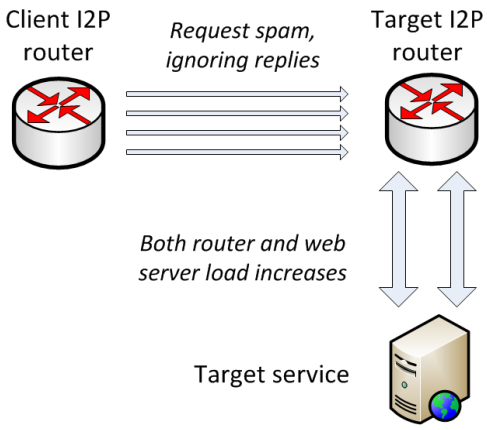

Figure 4 - Kack's Generates Unused Connections to Target I2P Node

The original version of denial-of-service attack used by Kack was mitigated with introduction of request rate limiter in I2P router software. However, this approach is inadequate and does not protect I2P nodes against other varieties of denial-of-service attacks. Instead of web server domain, the attack may instead focus on saturating the I2P encrypted tunnel limit, bandwidth or other resources of I2P node. The request rate limiter can be bypassed due to its reliance on I2P network identifier of the attacking node, which is not permanently assigned and can be changed by attacker once his node becomes blacklisted.

After the initial denial-of-service attack on I2P service node, the expected response of its operator was to increase the ratio of system resources available on I2P. This includes increasing the total bandwidth permitted, tunnel limit and memory size within I2P router configuration. However, this allows I2P router to accommodate an even larger denial of service attack, capable of having an impact on the I2P router. Network equipment used by the host is the first candidate for failure. The P2P concept of I2P means that an active I2P router will be constantly receiving a large number of inbound TCP and UDP packets from a similarly large pool of unique remote IP addresses. The Self-hosted I2P nodes running behind NAT may stop responding altogether, both on I2P network and the Internet. If the I2P router is monitored, the change in network performance or availability can be linked to denial of service attack.

An investigation of I2P network found that approximately half of total I2P network nodes do not stay connected for longer than a week [13]. This behaviour suggests presence of large number of nodes

that are running over residential DSL or mobile broadband connections, which would be unable to properly handle large number of P2P packets and therefore vulnerable to this kind of attack.

The initial sample of IP addresses that, need monitoring are obtained by parsing I2P's floodfill database, a copy of which is kept by every I2P floodfill router. Adrian Crenshaw's research into identification of remote eepsites has produced a set of Python scripts for extracting this information [14].

Ordinary nodes, which are not floodfill routers can be removed from candidate list, therefore reducing, total list of candidates to a manageable number (several hundred). This is due to the way I2P determines promotion of ordinary routers to floodfill nodes. As of version 0.9.23, any I2P router that allows sufficient bandwidth to be shared by network is allowed to switch into floodfill node. A node, which managed to remain available after the first denial-of-service attack, is therefore highly likely to be a floodfill node and present on the extracted list.

If required, candidate list can be reduced further by continuous monitoring of floodfill nodes' availability and geolocation by IP. Any node that becomes unreachable while the targeted I2P service is still online is no longer a candidate. Geographic filtering becomes possible if some information about targeted node operator is known from other sources: social engineering, timestamps or accidental posting of information. All floodfill router records in NetDB have their IP addresses stored and therefore can be remapped [9].

The remaining suspect node IP addresses can then be monitored for signs of change in network performance such as dropped packets and increased round-trip times. The attacking nodes on I2P network can be commanded to cyclically connect and disconnect from suspect eepsite or other resource in order to produce a more visible pattern of changes over longer period.

3. Related Works

Freenet [3] is an unstructured P2P system that has been designed to exchange information between users. It enables the publishing and retrieving of contents in an anonymous way where the source and

destination of the information is withheld from third parties and system servers. Peers in the network participate in queries, data storage and retrieval of data items.

Freenet does not assign responsibility for documents to specific nodes instead; lookups are carried out by searching for cached copies. Freenet aims to provide a flat Internet topology. In other words, an IP address next door can be communicated with; the same way one would communicate with another IP on the other side of the planet, without being discovered. It was first used by a large community of online users to distribute copyrighted materials on the internet without being discovered. Clarke [7] claims that this was not the purpose of the project. They discuss that the Internet is the biggest bastion of freedom of speech, since governments try to impose censorship on the flow of information in the press, multicasting and printed materials. Freenet nodes are encrypted and routed through other nodes to make it extremely difficult to determine its originator as well as content [7]. A request for key is passed along peers using flooding algorithm, which returns the corresponding data. These keys are location-independent. If a node received a request and knows the location of the file, it forwards it to the destination, which holds the information. If the node does not know the destination address, it forwards it to a node, which might hold the information or is likely to know the whereabouts of the resource.

To make the routing more efficient and smart, Freenet uses historical information and statistics from previous routing experiences to make a decision-based estimate of the time it might take to reach the destination. Caching based on specialisation of the nodes accumulated cache of the information that it then resulted Freenet not to cope with overwhelming requests and collapsed in July 2003. It was then that the designer addressed the load balancing issues by ensuring the uniform load distribution and constrains queries to maintain the defined quota. Considering this approach has addressed the problem and works effectively, but it may lead to functionality issues by limiting incoming requests to retrieve resources. This means that individual nodes behaving other than anticipated may affect load balancing and increase request failure rate. Therefore, the challenge in terms of scalability and performance still persists within the Freenet structure. Like any other P2P system, nodes in Freenet can have a dual role and are not distinguishable by name. This component of the system improves the anonymity.

However, an adversary can easily identify the traffic load and distinguish server nodes using a packet analyser. Having said that, Freenet remains one the important systems in providing user anonymity.

The Onion Router (Tor) is a distributed overlay network to anonymise TCP-based applications such as instant messaging, web applications and secure shell [15]. Each node in tor chooses a path, build a circuit with its neighbours known as successor and predecessor. The traffic is relayed through fixed-sized circuits and unwrapped by symmetric key at each node similar to layers of an onion. Tor uses the incremental relay of messages provides complete anonymity. The use of encryption at each layer provides data integrity. However, to avoid the alteration by nodes, Tor encrypts the messages before they leave the source node. However, some weaknesses have been found within Tor [8], [16]. Adversaries can attacks the Tor nodes on the exit or entry points.

Similar to I2P, Tor is vulnerable against CPU-consumed denial of service attacks. However, Tor provides low latency and high bandwidth which makes it attractive for users who share instant messages and large size files. The issues found in tor can be used to de-anonymise the users or decrypt the transmitted messages. However, this is beyond the scope of this paper. Nonetheless, like any other anonymity service online, Tor remains a challenge in any forensic investigation.

4. Conclusion and Future Works

In this paper, we discussed specific security issues and vulnerabilities of I2P networks. In addition, the functionality and its capabilities were outlined. As such, an argument was put forward, that while I2P provides an opportunity to maintain confidentiality and user privacy, it is generally exploited and utilised for illegal activities. While the anonymity systems maintain user privacy, promotes free speech and facilitates the free flow of information, for the level of illegal and criminal activities within I2P networks maintain a cause of concern. Due to the technological, geopolitical and legal challenges, accessing information on such activities is an issue for the forensic analyst and law enforcement agencies. As Such, in this paper, we provided different techniques, based on the vulnerabilities and flaws existed in I2P network [17], to forensically identify and retrieve I2P artefact. Our analysis and experiments show that such solutions can be integrated within the industry approved

forensic tools to promote better practice in I2P investigations within law enforcement and to enhance the continuity of the evidence.

For future work, we will investigate the security flows of I2P in additional detail in order to provide a more complete understanding of the network as a whole. This will contribute to effective and efficient investigation of I2P activities and provide a comprehensive approach in the forensic analysis of the malicious artefacts.

REFERENCES

[1] Valdez, Jason, et al. "An expanding reference library for Peer-to-Peer content." eCrime Researchers Summit (eCrime), 2011. IEEE, 2011.

[2] Braun, Patrik J., Péter Ekler, and Frank Fitzek. "Network coding enhanced browser based Peer-to-Peer streaming." Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on. IEEE, 2016.

[3] Y. Wei, C. Wang, Y. Chu and R. Chang, R, "A Secure and Stable Multicast Overlay Network with Load Balancing for Scalable IPTV Services," Int. J. Digit. Multimedia. Broadcast. pp. 1–12.B. 2012

[4] Nadeem, Muhammad Aamir, and Taimur Karamat. "A survey of cloud network overlay protocols." Digital Information and Communication Technology and its Applications (DICTAP), 2016 Sixth International Conference on. IEEE, 2016.

[5] Jayashree, G., and V. Perumal. "Enhancing similarity based query searching performance using self-organized semantic overlay networks." Computer Communication and Systems, 2014 International Conference on. IEEE, 2014.

[6] M. Jawad, P. Serrano-alvarado, and P. Valduriez, "Supporting Data Privacy in P2P Systems", Table of Contents," pp. 1–51, 2013.

[7] I. Clarke, "Freenet: A Distributed Anonymous Information Storage and Retrieval System" 1999, [available at http://freenetproject.org/freenet.pdf]

[8] D. McCoy, K. Bauer, D. Grunwald, T. Kohno and D. Sicker. Shining Light in Dark Places: Understanding the Tor Network. In Proc. of Privacy Enhancing Technologies Symposium (PETS), Leuven, Belgium, 2008.

[9] C. Timpanaro, I. Chrisment and O. Festor, "A bird's eye view on the I2P anonymous file-sharing environment", online at: https://hal.inria.fr/hal-00744919/PDF, 2012.

[10] Hesselman, Cristian, and Giovane CM Moura. "Increasing DNS Security and Stability through a Control." IEEE Communications Magazine: 2-8.

[11] Coudriau, Marc, Abdelkader Lahmadi, and Jerome Francois. "Topological Analysis and Visualisation of Network Monitoring Data: Darknet case study." IEEE International Workshop on Information Forensics and Security. IEEE, 2016.

[12] C. Kack, "Layer 7 DOS against I2P darknet", 2012 [available from http://blog.kejsarmakten.se/all/projects/2012/09/11/dark-loris.html]

[13] P. Liu, L.Wang, Q.Tan, Q.Li, X.Wang, J.Shi, “Empirical Measurement and Analysis of I2P Routers”, 2014 [Available from: https://pdfs.semanticscholar.org/3e5f/2b136df32beef1281b6b2f206093806c57f6.pdf]

[14] A. Crenshaw, “Darknets and hidden servers: Identifying the true IP/network identity of I2P service hosts”, 2011 [Available from: http://www.irongeek.com/downloads/Identifying%20the%20true%20IP%20of%20I2P%20service%20hosts.pdf]

[15] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation Onion router. In Proc. 13th USENIX Security Symposium, 2004.

[16] M.Ehlert, “I2P usability vs. Tor usability: a bandwidth and latency comparison” [Online] Available from: http://userpage.fu-berlin.de/semu/docs/2011_seminar_ehlert_i2p.pdf

[17] Zantout, Bassam, and Ramzi Haraty. "I2P data communication system." Proceedings of ICN. 2011.