TRAFFIC ANALYSIS AND RELAY FINDING IN TOR SURVEY

KEYUR RATHOD	HEPI SUTHAR
M. Tech Student	Asst. Professor
Marwadi University, Rajkot, India	Marwadi University, Rajkot, India
keyurrathod951@gmail.com	hepisuthar@gmail.com

Abstract

The main motto of this research is if any adversary or criminal uses the Tor browser for doing an illicit activity which causes govt. and people, as most of the techies and Tor users know that finding illicit communication in a Tor network is quite difficult because there is no direct or indirect mechanism to capture the encrypted traffic and decode it, and most importantly it's hard to find the relay location. So, we need a mechanism to trace them and identify the real criminals, this research is helpful in achieving this goal. The researcher creates a malicious payload and uses it to penetrate the Tor network with the help of traffic analysis. Here researcher also analyzes the tor middle relay in order to understand the network profound and generate statistics which creates a certain results; here the main focus is not only Tor. To check the workability of the payload, the researcher performs an experiment on the normal network in order to check the feasibility of payload and its working towards Tor. This research is helpful for those who want to study and get deep knowledge about Tor and also for those who are deucedly to penetrate Tor.

Keywords: Traffic analysis, relay finding, Tor Traffic analysis and network attacks, relay attack, Tor network attacks, finding relay location.

1 INTRODUCTION

The word cyber comes from the word cybernetics; in late 1940s cybernetics is used to describe the communication between two machines or people, now a day's a word cyber or cyberspace relates to the internet or computers and communication between them using a medium. Cyber security is an essential part of current technologies because as fast as technologies and digital security evolve day by day the hacking world grew faster so, safeguarding the network and vital information of organization and users is crucial. Cyber security involves cyber-attacks and cyber defenses where Cyber security or computer network security mainly deals with unauthorized access, protection against data breaches and digital damage to the network or computer-related systems and finds how to secure and protect them. Cyber security is more likely a process and practice of finding the vulnerability and fixing those vulnerabilities of a network or relates to electronic devices. The current era is of technology and communication where Cyber security is involved and covers every network, computer or electronic device which communicates using the internet. The vital part of safeguarding information which relies upon the internet like smart devices and smart technologies that can be affected by breach or security needs Cyber security protection.

The onion routing also known as Tor is a well-known low latency based anonymity network that is very popular among users of Dark & Deep Web, cyber criminal's favorite toy for doing any cyber-criminal activities. Tor is developed by U.S. Naval Research Lab in the 1990s for providing security and protecting their data privacy online. Their mathematician Paul Syverson and scientist Michael G. Reed and David

Goldschlag introduce Onion Routing protocol to provide strong protection against network surveillance. Tor is used for licit and illicit communication and utilization of Tor becomes very popular day by day as information and data are playing a vital role in cyberspace and mostly the data are unindexed on the internet so to access crucial data cyberpunk uses Tor, mostly data used for illicit communication like accessing govt. confidential data, unauthorized news leaks of sensitive information (ex. WikiLeaks), buying, selling, smuggling drugs and weapons, stolen credit card numbers, money laundering, bank and credit card fraud, Gain access to censored information, distribution of illegal sexual content, exchange of counterfeit currency, etc. every coin has two sides Tor have issues regarding research i.e. Tor uses onion routing technique so it's difficult to trace illicit activities online, Remembering the onion address is quite headache for Tor users and the big issues is Tor slower than normal network and uses high bandwidth for network usage and network is anonymous so difficult to identify criminal activities.

To achieve the goal of the research, the researcher divided the process flow into two phases one is a simulation-bed environment and the other is the emulation-bed environment. In a simulated-bed environment researcher establishes Tor non-exit relay and analyzes the Tor network and according to the statistic of results researcher creates a payload to penetrate the normal network and exploit router so, by penetrating it each router pingback routing table to the adversary. Here penetrating the half of the network router adversary (researcher) get full idea of the how many router in a network and list of ip-address of each router. In an emulated-bed setup, the researcher works with the actual Tor network and uses a payload to penetrate the Tor network and trace the location of the relays and also statistics and analysis of middle relay will helpful for penetration.

2 OBJECTIVES

The main objectives for this research are –

- To know about number of relays used for illicit communication.
- To observe and analyze Tor network in order to identify illicit traffic.

3 LITERATURE REVIEW

In the past decades there is a lot of work done against on traffic analysis of Tor network, previous researchers had done traffic analysis on either entry node or on exit node and in some case both S. Chakravarty, M. V. Barbera et al. [5] worked on effectiveness of active traffic analysis attack against Tor network using a statistical correlation method and Cisco NetFlow data to reveal a source of anonymous traffic which done in two phases and they monitor both enter and exit node relay data. [9] S. Chakravarty, G. Portokalidis et al. shows using two decoy servers they inject traffic pattern that exposes bait credentials for decoy services and deployed prototype implantation into the Tor network. Much research on traffic analysis happens on entry or exit points but this [6] R. Jansen, M. Juarez, et al. research conducted solely with middle relays and also worked on website fingerprinting to detect onion service usage. [8] Y. Gilad and A. Herzberg give methods to identify clients without eavesdrop on the communication to the server and also without relying on the traffic pattern using different network attacks and side channels attack based on two scenarios. P. Mittal et al. [7] showed that Tor (anonymity system) provide efficient service to its users by using full use of forwarding capacity and also this facility sometime leaks information about Tor relays in the circuit so, they present stealthy attacks based on throughput information can reduce uncertainty about bottleneck relay of any circuit whose throughput is observed to identify guard relays and whether 2 concurrent TCP connection belong to the same user. Tor is always vulnerable against traffic analysis attack S. J. Murdoch and G. Danezis [10] present new traffic analysis technique shows which nodes are being used by Tor having a partial view of the network, this research gives a very good and brief idea about how to reduce the anonymity provided by Tor.

The actual creator of Tor P. Syverson et al. [1] talks about second-generation onion router (Tor) and gives a brief idea about how Tor network work and motto behind creating this extraordinary low latency, popularly used anonymous network and also talk about limitation in original design with improvements. Hidden server now a days known as onion servers are very crucial part of the Tor network because it allows clients(users) to interact with onion services L. Øverlier and P. Syverson [2] shows attacks on these hidden servers which reveals the location, there are the first actual intersection attacks on any anonymous deployed network. [4] P. Winter, A. Edmundson et al. studied and conduct an online survey of 517 users and 17 semi-structured interviews of Tor users on how they use onion services, network communication of Tor, problems regarding onion addresses and improvements needed in Tor and onion service. [3] Remembering onion service address is difficult so, J. Victors et al. introduce Onion Name System (OnionNS) which allows Tor users to reference any onion service by a meaningful globally unique verifiable name by the administrator. The researcher also get idea about ad-hoc network so [13] papers talk about cluster routing in traditional ad-hoc network.

4 RESEARCH METHODOLOGY

4.1 Architectural Diagram

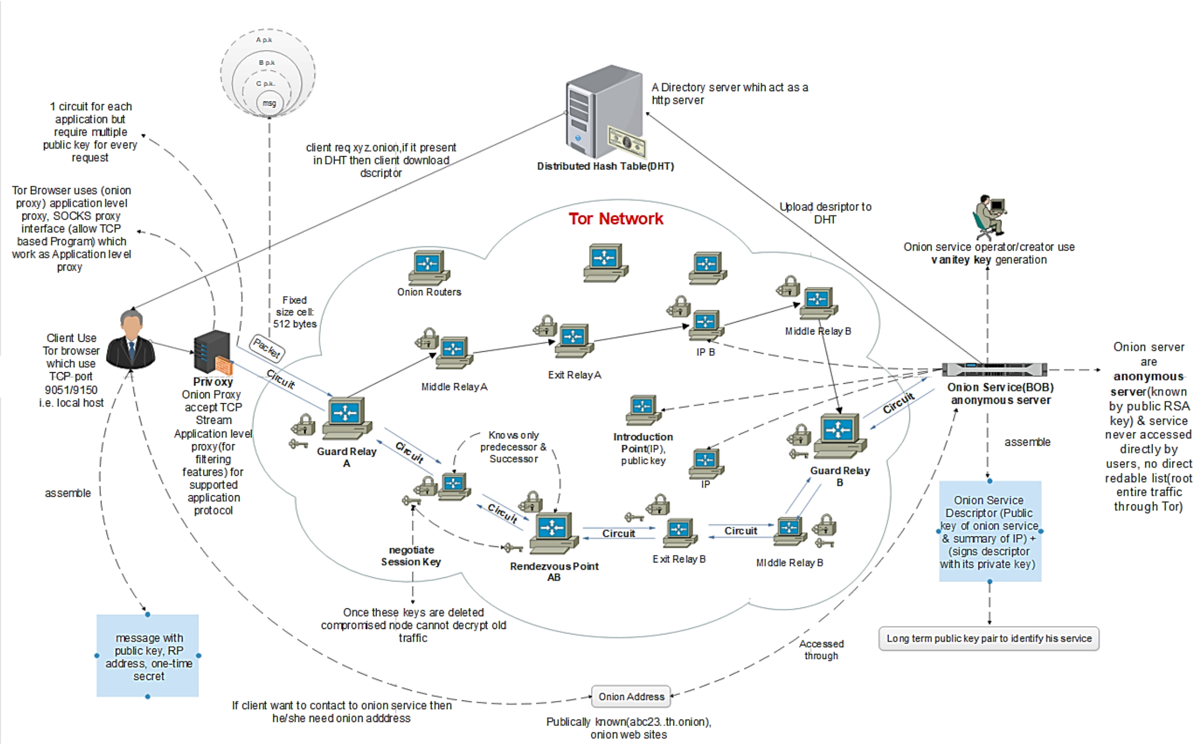

A p.k
B p.k
C p.k.
msg
1 circuit for each application but require multiple public key for every request
Tor Browser uses (onion proxy) application level proxy, SOCKS proxy interface (allow TCP based Program) which work as Application level proxy
client req xyz.onion,if it present in DHT then client download dscriptor
Distributed Hash Table(DHT)
A Directory server whih act as a http server
Upload desriptor to DHT
Onion service operator/creator use vanitey key generation
Tor Network
Onion Routers
Fixed size cell: 512 bytes
Packet
Circuit
Client Use Tor browser which use TCP port 9051/9150 i.e. local host
Privoxy
Onion Proxy accept TCP Stream Application level proxy(for filtering features) for supported application protocol
Guard Relay A
Middle Relay A
Exit Relay A
IP B
Middle Relay B
Onion Service(BOB) anonymous server
Onion server are anonymous server(known by public RSA key) & service never accessed directly by users, no direct redable list(root entire traffic through Tor)
Introduction Point(IP), public key
IP
Guard Relay B
assemble
Knows only predecessor & Successor
Circuit
Circuit
Circuit
Circuit
negotiate Session Key
Rendezvous Point AB
Exit Relay B
MIddle Relay B
Onion Service Descriptor (Public key of onion service & summary of IP) + (signs descriptor with its private key)
assemble
message with public key, RP address, one-time secret
Once these keys are deleted compromised node cannot decrypt old traffic
Long term public key pair to identify his service
Accessed through
Onion Address
If client want to contact to onion service then he/she need onion adddress
Publically known(abc23..th.onion), onion web sites

Fig. 4.1. Tor Architectural diagram

Onion routing protocol (Tor) is out-turn of a P. Syverson, M. G. Reed and D. Goldschlag which is created for protecting the user's anonymity while using the internet. Tor network is different from the traditional network, the backbone of Tor is onion servers and volunteer relays. Working of Tor is very stiff because first it needs onion servers which provide different onion services (OS) to Tor users; onion service creates a public key to advertise its existence inside Tor because a public key worked as an onion address through which users interact with different onion sites. OS creates an OS descriptor and uploads it to the Distributed Hash Table (DHT), now client/user download the DHT and knows about 16 characters long onion service address which is derived from services public key, after knowing onion address now client request to the DHT and ask services for connection. If the onion service exist and free to receive connection then client learns about onion service public key (onion service address) and IP-address, meanwhile client picks a random relay to build a circuit and assemble an introduction message (which contain one-time secret, address of Rendezvous Point) encrypt it with public key of OS and send it back to an OS till now 1st half is completed. 2nd half involves actual communication, OS receives the message and decrypts it with its own private key and learns about RP, OS creates a circuit through RP and further communication done through that circuit, here RP tells the client that connection established. Here the important part is RP doesn't know the OS and also the client, it only worked as a tunnel between client and OS, and 6 hopes are used in entire communication.

This research is about finding relays location using traffic analysis of Tor, the main agenda for this research is criminal uses dark web and deep web for accessing govt. confidential data, gain access to censored information, and other illegal online activities in which the black market utilizes the Tor infrastructure. The expected outcome of this research is to find relay location inside the Tor network with the help of traffic analysis to identify cyber-criminal illicit activities and malicious payload which gives how many numbers of relays used for illicit communication, and their location. As discussed Tor is very popular among those cyber buddies who hire hackers or criminals to do illicit activities which scathe govt. or other legitimate organizations, it directly affects the black market because it revels the relay location inside the Tor network so, for govt. defense department it becomes easy to trace them. To achieve the goal of the research, the researcher divided the process flow into two different phases, phase-1 and phase-2. Both uses in the analysis of tor network phase-1 are about simulation setup of tor middle relay and gathering logs, analyze it and payload injection in normal network second phase directly deals with actual tor network and traffic analysis of Tor, network attacks, payload injection and result analysis.

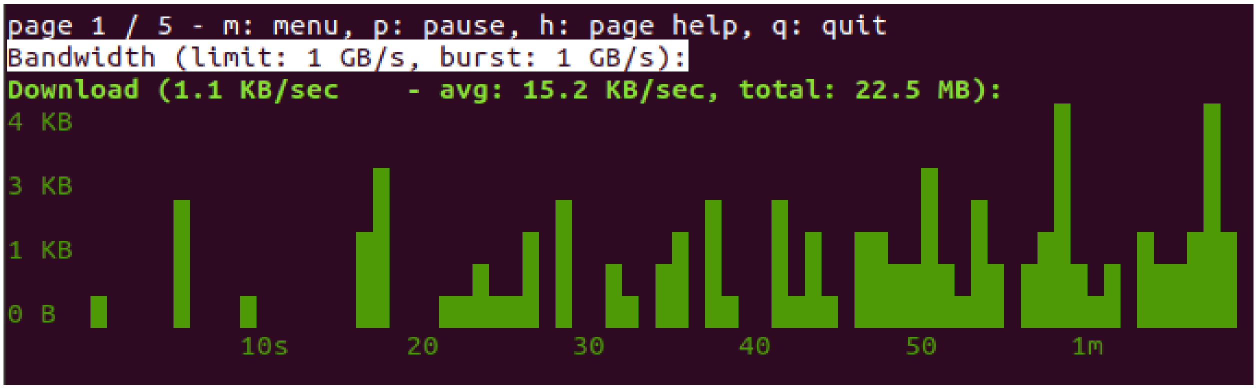
4.2 Detailed Operational Plan

Whole research is work in two phases both phases covers the analysis of Tor network, the first phase is a simulated test bed environment in which tor middle relay is set up in two different Operating System (OS), Kali GNU/Linux kali-rolling version 2019.4, Ubuntu Bionic-Beaver version 18.04.3 LTS and to monitor relay utilization in Tor NYX version 2.1.0 is installed which gives very good idea about relay working in a graph format it also gives inbound and outbound connections details because NYX is a very useful tool to check relay utilization and if user want to modified then they can, by this setup researcher get good idea about behaviour of middle relay and analyse logs which gather data regarding the inbound and outbound connection. This setup is to run day and night for gathering good and truthful information which generate results and saves and maintain logs, this results use in analyzing tor network for finding crucial information, here payload is vital part of the research and test-bed setup because payload helpful in locate the relay location, its written in python language, first researcher test the payload in normal network to gather router location (ip-address) to check whether it successfully penetrate the normal network or not, according to the researcher speculation if the payload failed to penetrate the normal network then it won’t be able to penetrate complex and strong Tor network. To get the ip-address of router which are connected in peer-to-peer network researcher perform ethical attack here researcher make an assumption to check whether payload is able to bypass the firewall of network and router without revealing itself, if successfully worked and give list of routers ip-address in the whole network then this payload is mounted in Tor network, here payload is injected with https/http request to perform ethical attack, using those statistic researchers generate result which is helpful in phase two.

Phase two is an actual emulated bed setup in which researcher performs the real task on live Tor network, here phase-1 statistic and results are guiding researcher in performing the attacks on Tor network, phase two describes the live Tor network in which middle relay is also needed to observe and capture traffic for analysis purpose. According to the statistic of each relay, the researcher generates the results and using it, graph is generated to describe the process. Here payload plays the vital role in process because payload is used for generating an attack on Tor network and furthermore it’s also used in monitoring the behavior of itself, here researcher assumes that payload is strong enough to penetrate the Tor network and give at least 1st relay location i.e. first middle relay location place after guard relay. Here main agenda of this is researcher first check the possibility of the payload whether it’s powerful enough to give location of relay then researcher attaches the payload with http/https request and send it to the live Tor network. Here payload is programmed in such a way that it revert back ip-address of the particular relay and spreading automatically inside the Tor network. The payload design in such a way that it creates a persistence connection between two relay and create a circuit and each relay in a circuit ping back its own ip address to the adversary. If tor mechanism won’t allow the payload to enter in network then researcher also have a solution i.e. it ping back its last location from where it is discarded. This gives idea where payload needs to be upgraded in order to enter the network.

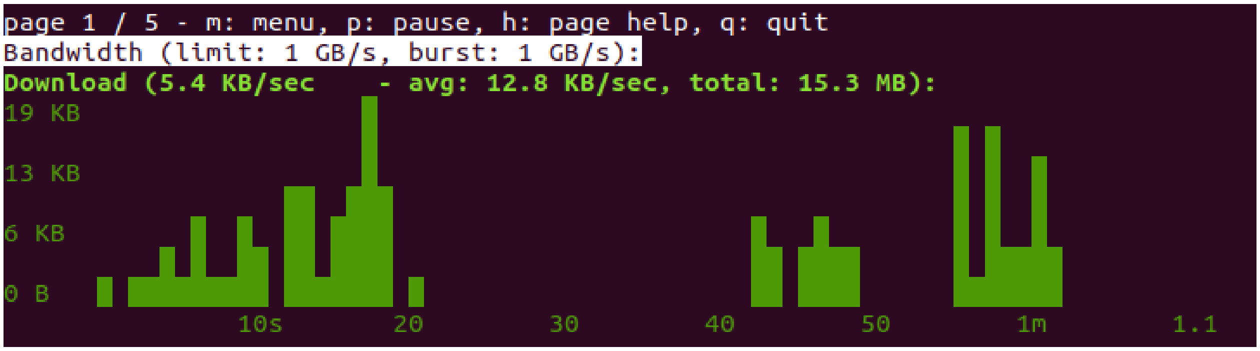
5. ANALYSIS AND FINDINGS

Graph 1 Download speed of In-bound connection



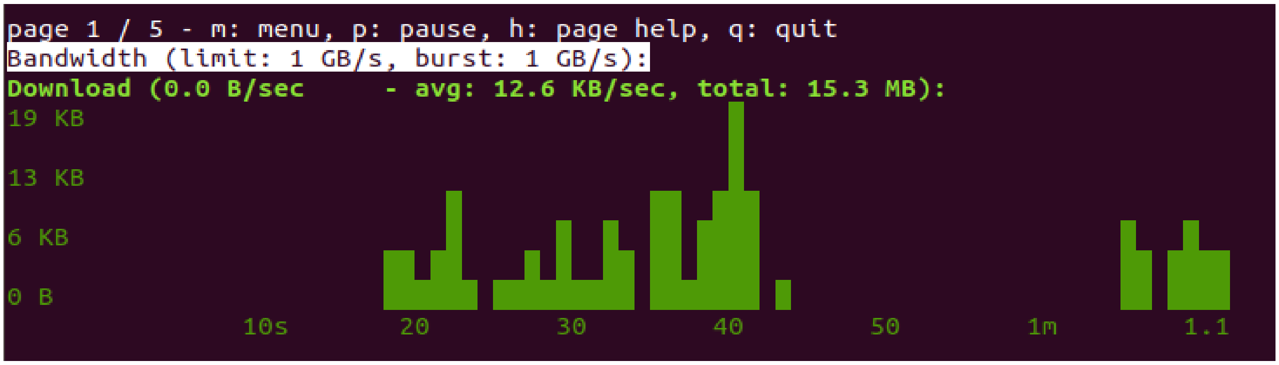
```
page 1 / 5 - m: menu, p: pause, h: page help, q: quit
Bandwidth (limit: 1 GB/s, burst: 1 GB/s):
Download (1.1 KB/sec    - avg: 15.2 KB/sec, total: 22.5 MB):
4 KB

3 KB

1 KB

0 B
              10s       20        30        40        50        1m
```

Graph 2 Download speed of In-bound connection



```
page 1 / 5 - m: menu, p: pause, h: page help, q: quit
Bandwidth (limit: 1 GB/s, burst: 1 GB/s):
Download (5.4 KB/sec     - avg: 12.8 KB/sec, total: 15.3 MB):
```

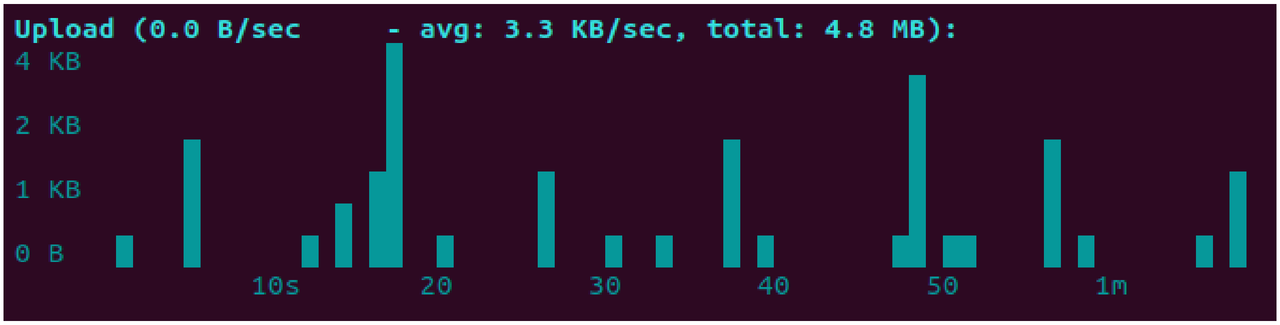
Graph 3 Download speed of In-bound connection



```
page 1 / 5 - m: menu, p: pause, h: page help, q: quit
Bandwidth (limit: 1 GB/s, burst: 1 GB/s):
Download (0.0 B/sec      - avg: 12.6 KB/sec, total: 15.3 MB):
```

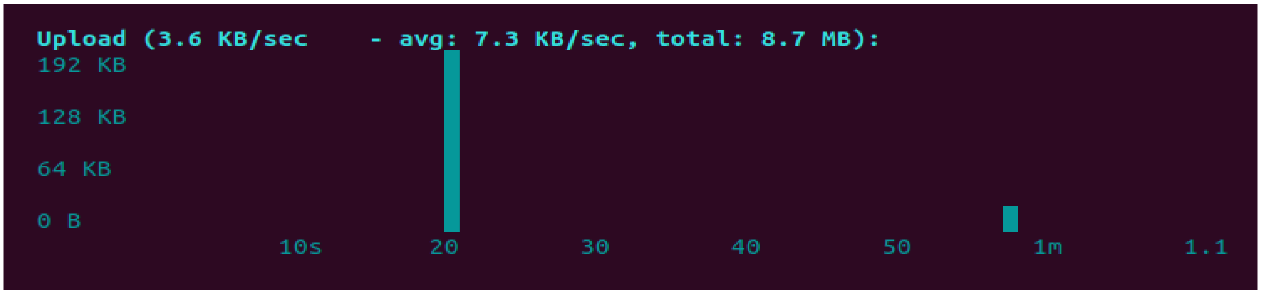
This above graph 1, 2 and 3 represents the in-bound connections download speed. Left side of the graph represents the speed in kilo bytes (kb) and below horizontal line represents time in seconds. This graph describes how many kbps of data is connected to tor middle relay. This is the graphical representation of the connection is made to tor middle relay (non-exit relay). This graph is continuously fluctuating with internet speed.

Graph 4 Upload speed of Out-bound connection

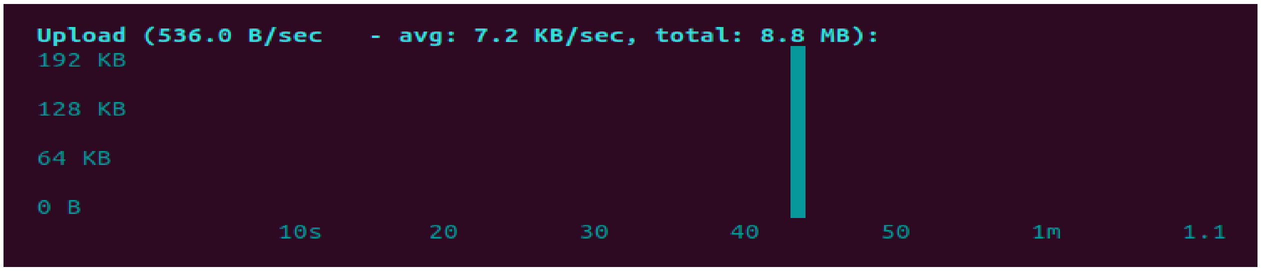


```
Upload (0.0 B/sec      - avg: 3.3 KB/sec, total: 4.8 MB):
```

Graph 5 Upload speed of Out-bound connection



Graph 6 Upload speed of Out-bound connection

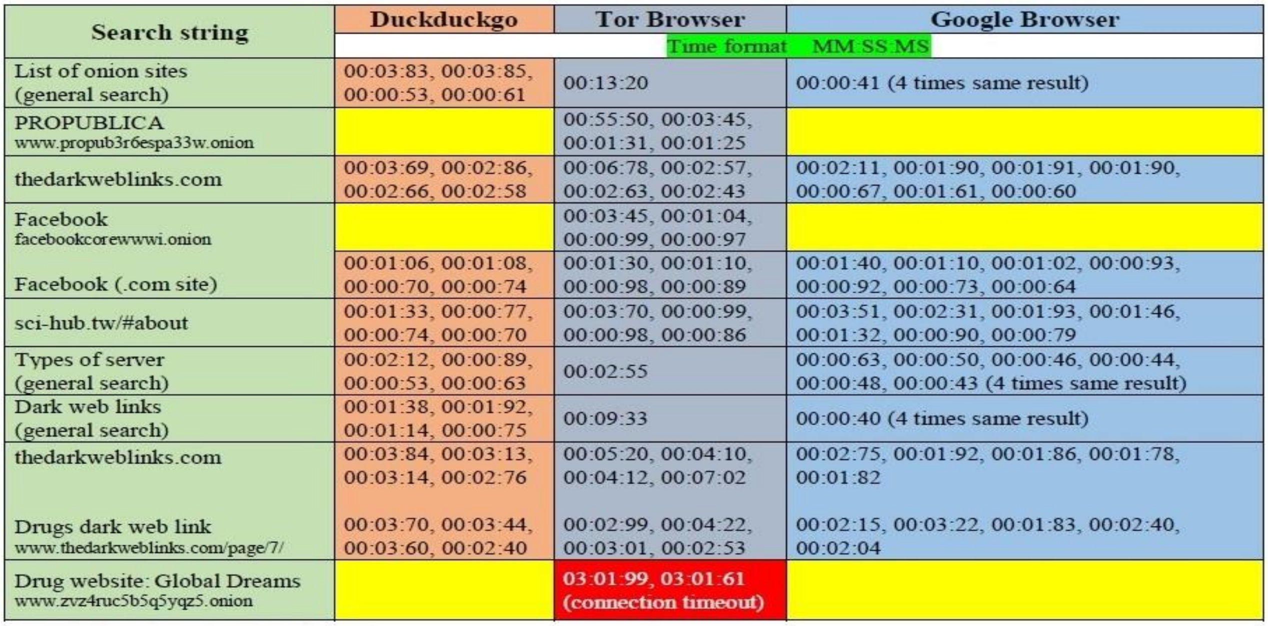


Graph 4, 5 and 6 represents the outbound connection of Tor middle relay same as in in-bound, vertical line represents the speed in kb and horizontal line represents the speed in seconds. This graph represents the rate of out-bound connection w.r.t connection established and leaving the tor non-exit relay.

Graph 1 and graph 2 is a graphical representation of the connection movements in tor network. To represents this graphical form of connection researcher use nyx tool.

NOTE: The graph 1, 2, 3, 4, 5 and 6 all are tested in Ubuntu OS.

Table: 1 Comparison between Browsers



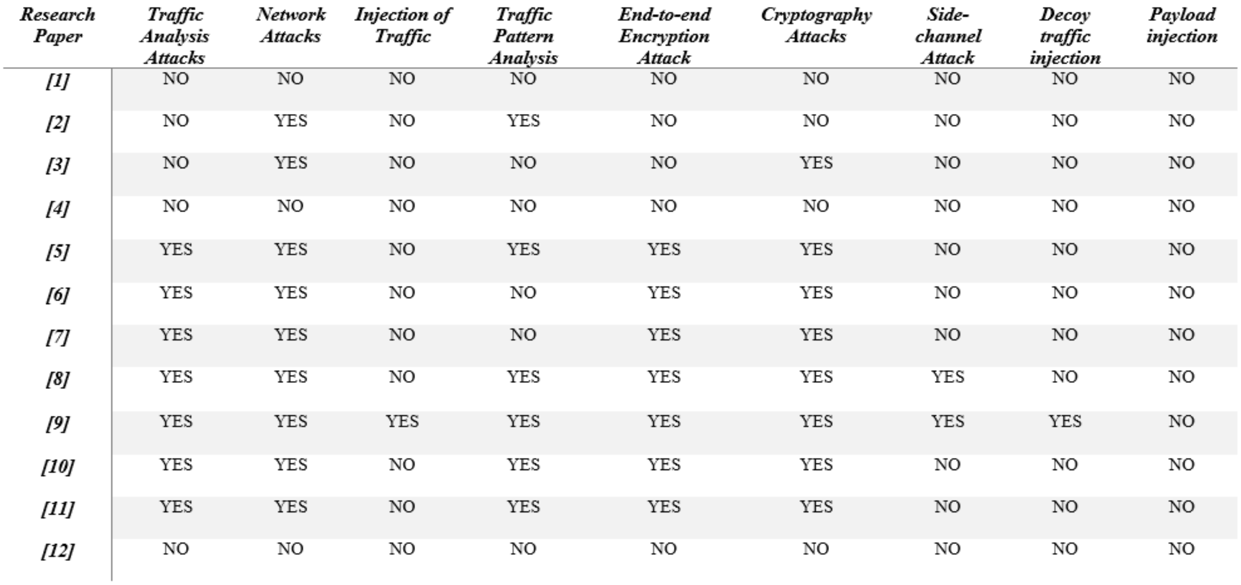
Search string	Duckduckgo	Tor Browser	Google Browser
		Time format MM:SS:MS	
List of onion sites (general search)	00:03:83, 00:03:85, 00:00:53, 00:00:61	00:13:20	00:00:41 (4 times same result)
PROPUBLICA www.propub3r6espa33w.onion		00:55:50, 00:03:45, 00:01:31, 00:01:25	
thedarkweblinks.com	00:03:69, 00:02:86, 00:02:66, 00:02:58	00:06:78, 00:02:57, 00:02:63, 00:02:43	00:02:11, 00:01:90, 00:01:91, 00:01:90, 00:00:67, 00:01:61, 00:00:60
Facebook facebookcorewwwi.onion		00:03:45, 00:01:04, 00:00:99, 00:00:97	
Facebook (.com site)	00:01:06, 00:01:08, 00:00:70, 00:00:74	00:01:30, 00:01:10, 00:00:98, 00:00:89	00:01:40, 00:01:10, 00:01:02, 00:00:93, 00:00:92, 00:00:73, 00:00:64
sci-hub.tw/#about	00:01:33, 00:00:77, 00:00:74, 00:00:70	00:03:70, 00:00:99, 00:00:98, 00:00:86	00:03:51, 00:02:31, 00:01:93, 00:01:46, 00:01:32, 00:00:90, 00:00:79
Types of server (general search)	00:02:12, 00:00:89, 00:00:53, 00:00:63	00:02:55	00:00:63, 00:00:50, 00:00:46, 00:00:44, 00:00:48, 00:00:43 (4 times same result)
Dark web links (general search)	00:01:38, 00:01:92, 00:01:14, 00:00:75	00:09:33	00:00:40 (4 times same result)
thedarkweblinks.com	00:03:84, 00:03:13, 00:03:14, 00:02:76	00:05:20, 00:04:10, 00:04:12, 00:07:02	00:02:75, 00:01:92, 00:01:86, 00:01:78, 00:01:82
Drugs dark web link www.thedarkweblinks.com/page/7/	00:03:70, 00:03:44, 00:03:60, 00:02:40	00:02:99, 00:04:22, 00:03:01, 00:02:53	00:02:15, 00:03:22, 00:01:83, 00:02:40, 00:02:04
Drug website: Global Dreams www.zvz4ruc5b5q5yqz5.onion		03:01:99, 03:01:61 (connection timeout)	

Analysis of Tor browser w.r.t other browser is showing up here. Each column represents different information with different values and also each column colour represents different information. Here Yellow box represents that .onion site is not able to open in normal browsers, and Red box represents connection time out. Green colour represents search string and Orange, Grey; Blue colour represents a different search engine. Here some test uses college internet and some test uses home internet (GTPL network).

NOTE: There are some conditions which were considered while performing the practical

The researcher did not use any VPN (Virtual Private Network) while performing the task, there are other factors which also have to be considered in this practice like human error, Internet speed, Website responding time, well-known sites take less time than less known sites.

Table: 2 Attack Observation



Research Paper	*Traffic Analysis Attacks*	*Network Attacks*	*Injection of Traffic*	*Traffic Pattern Analysis*	*End-to-end Encryption Attack*	*Cryptography Attacks*	*Side-channel Attack*	*Decoy traffic injection*	*Payload injection*
[1]	NO	NO	NO	NO	NO	NO	NO	NO	NO
[2]	NO	YES	NO	YES	NO	NO	NO	NO	NO
[3]	NO	YES	NO	NO	NO	YES	NO	NO	NO
[4]	NO	NO	NO	NO	NO	NO	NO	NO	NO
[5]	YES	YES	NO	YES	YES	YES	NO	NO	NO
[6]	YES	YES	NO	NO	YES	YES	NO	NO	NO
[7]	YES	YES	NO	NO	YES	YES	NO	NO	NO
[8]	YES	YES	NO	YES	YES	YES	YES	NO	NO
[9]	YES	YES	YES	YES	YES	YES	YES	YES	NO
[10]	YES	YES	NO	YES	YES	YES	NO	NO	NO
[11]	YES	YES	NO	YES	YES	YES	NO	NO	NO
[12]	NO	NO	NO	NO	NO	NO	NO	NO	NO

The researcher study Tor network and on the basis of some research paper researcher make a list of common attacks which can be used to analyze or penetrate the Tor network. The observation table is the analysis of the researcher's work.

6. CONCLUSION

This research paper gives a good idea about working of the Tor network, how the client/user connects to the Tor network and actual communication happen inside Tor which helps and guides readers to further analysis of Tor and future work. The researcher talks about how payload helpful in the entire research. This research is based on identifying relay location with the help of payload by binding it with the http/https request and sends it to the Tor network and analyzes the behaviour payload and also getting ip-address of a relay in a circuit. Here traffic analysis of a Tor plays a vital part to understand the Tor network in order to perform network attacks.

7. REFERENCES

1. F. Rochet and O. Pereira “Dropping on the Edge: Flexibility and Traffic Confirmation in Onion Routing Protocols*” Proceedings on 18th Privacy Enhancing Technologies Symposium (PETS2018),* Barcelona, Spain, July 24–27, 2018.

2. J. Victors, M. Li, and X. Fu “The Onion Name System: Tor-powered Decentralized DNS for Tor Onion Services”. *Proceeding on Privacy Enhancing Technologies symposium 2017(1),* January 2017.

3. L. Øverlier and P. Syverson. “Locating Hidden Servers”. *IEEE Symposium on security and Privacy,* claremontresort-Oakland, California, USA, May 21-24, 2006.

4. N. Dutta and HKD Sarma, “A probability based stable routing for cognitive radio Adhoc networks”, *Wire.Net, (Springer),* vol. 23(1), pp. 65-78, 2017.

5. N. Dutta and IS Misra, “Mathematical modelling of HMIPv6 based network architecture in search of an optimal Performance”, *IEEE 15th ADCOM,* Guwahati, India, pp. 599-605, 2007.

6. N. Dutta and IS Misra,”Multilayer hierarchical model for mobility management in IPv6: a mathematical exploration”, *Wire. Pers. Comm (Springer),* vol.78 (2), pp.1413-1439, 2014.

7. N. Dutta, HKD Sarma and Z. Polkowski, “Cluster based routing in cognitive radio Adhoc networks: reconnoitering SINR and ETT impact on clustering”, *Com. Com., (Elsevier),* pp. 10-20, vol. 115, 2018.

8. P. Mittal, A. Khurshid, J. Juen, M. Caesar, and N. Borisov “OSINT Analysis of the TOR Foundation”.

9. P. Mittal, A. Khurshid, J. Juen, M. Caesar, and N. Borisov “Stealthy Traffic Analysis of Low-Latency Anonymous Communication Using Throughput Fingerprinting” *Proceedings of the 18th ACM conference on Computer and Communications Security,* Chicago, Illinois, USA, October 17 - 21, 2011.

10. P. Winter, A. Edmundson, L. M. Roberts, A. Dutkowska-Zuk, M. Chetty, and N. Feamster “How Do Tor Users Interact With Onion Services?” *Proceedings of the 27th Usenix Security Symposium,* Baltimore, MD, USA, August 15–17, 2018.

11. R. Dingledine, N. Mathewson, and P. Syverson “Tor: The Second-Generation Onion Router”, *13th USENIX Security Symposium,* San diego, CA, USA, August 9-13, 2004.

12. R. Jansen, M. Juarez, R. Gálvez, T. Elahi, and C. Diaz "Inside Job: Applying Traffic Analysis to Measure Tor from Within" *Proceedings of the 25th Symposium on Network and Distributed System Security (NDSS '18)*, San Diego, CA, USA, February 18-21, 2018.

13. S. Chakravarty, G. Portokalidis, M. Polychronakis, and A. D. Keromytis "Detecting Traffic Snooping in Tor Using Decoys" *Proceedings of the 14th International Conference on Recent Advances in Intrusion Detection*, Menlo Park, CA, USA, September 20-21, 2011.

14. S. Chakravarty, M. V. Barbera, G. Portokalidis, M. Polychronakis, and A. D. Keromytis "On the Effectiveness of Traffic Analysis Against Anonymity Networks Using Flow Records" *Proceedings of the 15th Passive and Active Measurements Conference (PAM 2014),* Los Angeles, CA, USA, March 10-11, 2014.

15. S. J. Murdoch and G. Danezis "Low-Cost Traffic Analysis of Tor" *Proceedings of the 2005 IEEE Symposium on Security and Privacy,* The Claremont Resort, Oakland, California, USA , May 8-11, 2005.

16. S. Sathwara and C. Parekh "Distributed Denial of Service (DDoS) Attacks Comparative Impact Analysis and Mitigation Techniques: A Survey" IJARIIE-ISSN (O)-2395-4396, Vol-3 Issue-2 2017.

17. Y. Gilad and A. Herzberg "Spying in the Dark: TCP and Tor Traffic Analysis" *Proceedings of the 12th Privacy Enhancing Technologies Symposium (PETS 2012),* Vigo, Spain, July 11–13, 2012.