Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

A Review on Classification of Tor-Nontor Traffic and Forensic Analysis of Tor Browser

Srusti D. Mehta
M.E in Computer Engineering (Cyber Security)
Gujarat Technological University
Ahmedabad, Gujarat

Deepak Upadhyay
Assistant Professor
Gujarat Technological University
Ahmedabad, Gujarat

***Abstract*— Tor browser is for normal use as well as for the malign activities. The Tor is also called as Onion Routing. The Onion Routing system is seen as best anonymity tool out there and is used by nearly 2.5 million daily. Law Enforcement Agencies need to monitor and to investigate crimes hidden behind the anonymity provided by the Tor browser and there are also many other professions in which it would be helpful or necessary to keep an anonymous online profile. It includes Law enforcement officers, Journalists, Militaries, IT professionals, Business executives. In this paper a review on Classification of Tor-Nontor traffic and the Forensic Analysis of Tor Browser are there. The results can be helpful for Law Enforcement Agencies in cases where a Tor browser user is under an investigation.**

Keywords*— *Tor; Onion Routing; User anonymity; Private browsing; Tor traffic; Nontor traffic; Forensic Analysis; Windows OS.

I. INTRODUCTION

The deep web is a subset of internet that is not indexed by the major search engines it means that you have to visit those places directly instead of being able to search for them. So, there are not directions to get there, but they are waiting if you have an address. The deep web is largely simply because the internet is too large for search engines to cover completely. So, the Deep Web is the long tail of what is left out. The dark web is the World Wide Web content that exists on darknets, overlay networks which use public internet but require specific software, authorization or configurations to access. The dark web is a small part of the deep web, the part of the web not indexed by search engines, although sometimes the term “deep web” is mistakenly used to refer specifically to the dark web.

Anonymity and privacy are two main elements to protect freedom of speech. Goal of anonymity is to protect all the information which can reveal real identity of user information like real name, location, IP address etc. The goal of privacy is to make sure that any organization or entity does not collect or store any personal or private information like user browser history, location information, account details etc. without user’s knowledge. Tor project was initiated in 1995 by US Naval Research Laboratories and the main goal of their project was to separate identification information from routing and to design an anonymous communication network for military communication. Tor browser is a modified version of Mozilla Firefox with some extra features for anonymity and privacy. Some of these features are the Tor launcher, Tor button, no script and HTTPS-Everywhere. By default, browsing is configured for private mode with the option to clear browsing activity and its related artifacts such as cookies and other browsing related data after closing of the browser [1].

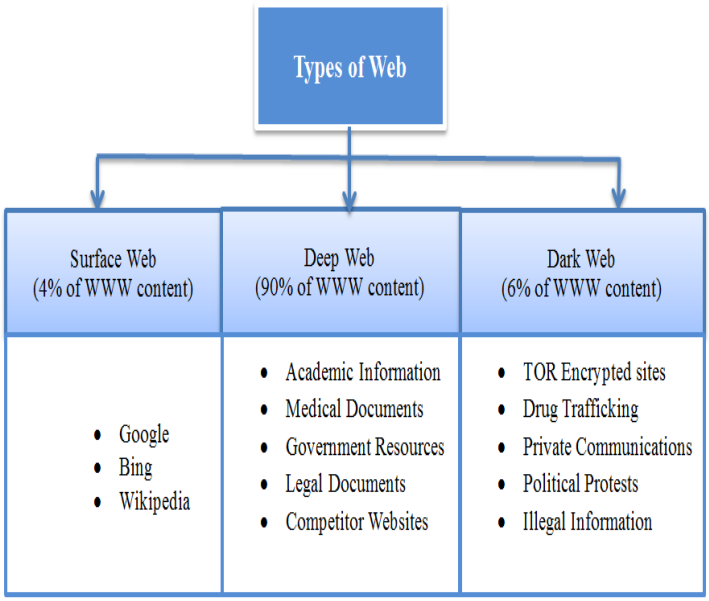

Figure 1: Types of Web

II. TOR WORKING METHODOLOGY

Tor is free and open source software for enabling anonymous communication. Tor name is derived from an acronym for the original software project name “The Onion Router”. So, Tor is also known as Onion Routing. Web analytics have increased over the last decade, so has the number of user trying to maintain their online anonymity. Due to this, majority of the users use Tor browser for normal use as well as malign activities. The Onion Routing system is seen as best anonymity tool out there and is used by nearly 2.5 million daily. Law Enforcement Agencies need to monitor and to investigate crimes hidden behind the anonymity provided by Tor. In Tor browser, data encryption is done in a layered (i.e., ‘onion-like’) manner so that only the last/exit proxy is able to decrypt the original client data and forward it to the final destination, while the intermediate TOR proxies remain unaware of the data content as well as the identity of the final destination. While connected to the Tor network, activity will never be traceable back to your IP address. Similarly, your Internet Service Provider (ISP) won’t be able to view information about the contents of your traffic, including which

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

website you're visiting. It's very difficult, if not impossible, to become truly anonymous online, but Tor can certainly help you get there. All of your traffic arriving at its destination will appear to come from a Tor exit node, so will have the IP address of that node assigned to it. Because the traffic has passed through several additional nodes while encrypted, it can't be traced back to you.

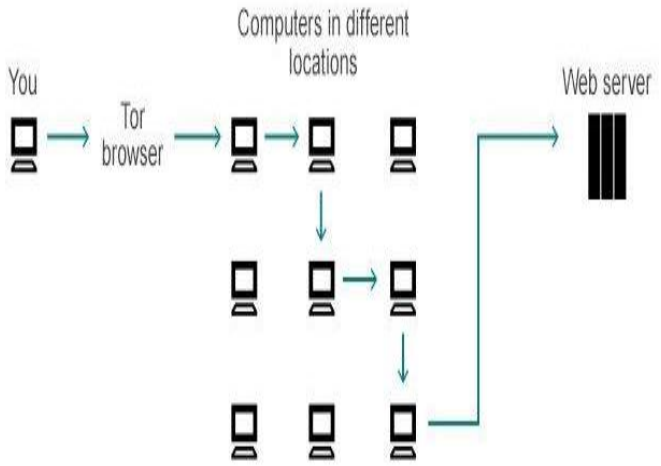

Figure 2: Working of Tor [6]

However, one of the issues lies in trusting the operator of the exit node. If you're visiting an unencrypted (non-HTTPS) website, it's possible the node operator can track your activity and view your information. They could collect data such as which webpages you're viewing, your login information, the content of your messages or posts, and the searches you perform. Although, there's no way to trace that information back to you or even back to the entry node.

III. LITERATURE REVIEW

Jadoon, Iqbal [1] objective is Forensic analysis of Tor Browser on Windows 8.1. This browser is not perfect but still with all these weaknesses, it is good enough because it provides both privacy and anonymity at the same time. It offers features like tor button, no script and HTTPS-Everywhere which further improve its anonymity and privacy. Jadoon, Iqbal analyzed system registry, memory and hard disk for all the artifacts that Tor browser leaves on user system when browser is open and after it is closed. Jadoon looked for the artifacts about Tor installation, usage and browsing activities. Tor browser leaves many artifacts on user system especially in system memory. In this research, all possible artifacts are recovered from host system. Jadoon also considers different scenarios which a forensic investigator can face during investigation. Recovering relays information from memory and hard disk is very important. This information will be very helpful in backtracking Tor user. This research will also be very help for researches and investigators with limited budget. In future research network forensics of the Tor browser and in forensics analysis of orfox which is android version of this browser. Orbot is another android app which works as Tor proxy. Forensics analysis of this application is also including in future research goals [1].

Muir M, Leimich P [10] the Tor browser, though, can leave behind digital artifacts in Windows 10 which can be used by an investigator. The limited research in the area of Tor Forensics discovered during the Literature Review suggested that a need for live forensics had become increasingly important. A new methodology was proposed, which favoured a live forensic analysis followed by a static analysis of the virtualized test machine. This resulted in a more through forensic process, allowing the maximum level of knowledge about the browser to be gained in the limited timeframe available. This paper outlines an experimental methodology and provides results for evidence trails which can be used within real-life investigations [10].

Rebecca Nelson, Atul Shukla [12] identifies the digital artifacts and their locations that could be recovered from Google Chrome, Mozilla Firefox, their respective private modes, and TOR. This research not only extends the current field of digital forensics for which artifacts can be found in which locations, but also confirms various claims in regards to the privacy of private browsing modes. Rebecca Nelson, Atul Shukla successfully recovers several artifacts of interest, e.g. browser history, cookies, and form autofill information in all public browsing sessions. Analysis will be conducted using tools commonly used by law enforcement agencies around the nation instead of proprietary or application-specific file carvers. Research project will focus on Windows 7 and the latest version of the TBB (v7.0.5). Future research in the realm of web browser forensics should continue to analyze which artifacts can be recovered from commonly used web browsers to aid in digital forensic investigations [12].

Alfredo Cuzzocrea, Fabio Martinelli [2] proposes a machine learning technique able to identify whether a user is using the Tor network. As secondary result, Alfredo Cuzzocrea, Fabio Martinelli evaluates the effectiveness of the proposed technique in the discrimination of the kind of service used in the Tor network. Apply state-of-art machine learning algorithms in order to evaluated method on real-world data.
As future work Alfredo Cuzzocrea, Fabio Martinelli plan to follow three distinct directions:

(i) investigating whether formal methods techniques can be useful in order to achieve better performances;
(ii) extending the target application environment to Clouds
(iii) studying how synopsis-oriented paradigms, like those proposed in different but- related contexts, may contribute to make traffic analysis framework more efficient [2].

Natalija Vlajic, Pooria Madani [11] look at the performance of the TOR browser. In its default settings, the TOR browser provides little if any protection against four most common forms of user tracking.

1] Fingerprinting (IP address)-based tracking
2] Storage (cookie)-based tracking
3] Session (URL rewriting)-based tracking
4] Cache (ETags)-based tracking

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

Four different techniques of user tracking and then looked at the possibility of using each of these techniques to track anonymized TOR visitors to a public website. Subsequently, point to specific additional steps and measures that should be taken by a TOR user in order to minimize or fully eliminate the possibility of being tracked online [11].

Eduardo Fidalgo, Enrique Alegre objective [4] is explore the automatic classification of images uploaded to Tor darknet. Methodology is Semantic Attention Keypoint Filtering (SAKF). A filtering strategy that removes non-significant features at a pixel level that mainly do not belong to the object of interest or foreground. Improvements over the baseline results. Further investigations will be focused on the feasibility of combining SAKF method with deep features, together with the automatic selection of the blurring factor proposed in [4].

Mhd Wesam Al-Nabki, Eduardo Fidalgo [9] objective is identifying the Most Suspicious Domains in the Tor Network. Methodology is a new algorithm, named ToRank. ToRank, that ranks hidden services in Tor better than the known algorithms used for the Surface Web. Mhd Wesam Al-Nabki, Eduardo Fidalgo thoroughly analyze the content present in Tor, creating a dataset, DUTA-10K, that extends the previous Darknet Usage Text Address (DUTA) dataset. Proposal obtains a higher harm to the Tor network robustness than all of them, what indicates its superiority for this problem [9].

Traffic classification has been the topic of many research efforts, but the quick evolution of Internet services and the pervasive use of encryption makes it an open challenge. In this paper, Arash Habibi Lashkari, Gerard Draper Gil [3] present a time analysis on Tor traffic flows, captured between the client and the entry node. They define two scenarios, one to detect Tor traffic flows and the other to detect the application type: Browsing, Chat, Streaming, Mail, Voip, P2P or File Transfer. As future work extend dataset and further study the application of time-based features to characterize encrypted traffic. Also, to extend ISCXFlowMeter application to extract the other features such as Flow-based and Packet-based to experiment the combination of these features sets [3].

IV. CONCLUSION

In this review, Classification of Tor-Nontor traffic and the Forensic Analysis of Tor Browser are there and these results can be helpful for Law Enforcement Agencies in cases where a Tor browser user is under an investigation. Forensics tools are available for all major browsers but there are no specific tools for this browser. There are also many other professions in which it would be necessary or helpful to keep an anonymous online profile. It includes Journalists, Law enforcement officers, Business executives, Militaries, IT professionals. In upcoming versions of browser it will also be helpful for Tor browser developer to improve security and privacy of their browser.

REFERENCES

[1] Abid Khan Jadoon, Waseem Iqbal, Muhammad Faisal Amjad, Hammad Afzal, Yawar Abbas Bangash, “Forensic Analysis of Tor Browser: A Case Study for Privacy and Anonymity on the Web”, Forensic Science International, Elsevier, March 2019.

[2] Alfredo Cuzzocrea, Fabio Martinelli, Francesco Mercaldo, Gianni Vercelli, “Tor Traffic Analysis and Detection via Machine Learning Techniques”, International Conference on Big Data (BIGDATA), IEEE, 2017.

[3] Arash Habibi Lashkari, Gerard Draper Gil, Mohammad Saiful Islam Mamun and Ali A. Ghorbani, **“**Characterization of Tor Traffic using Time based Features”, 3rd International Conference on Information Systems Security and Privacy, ICISSP 2017.

[4] Eduardo Fidalgo, Enrique Alegre, Laura Fernández-Robles, Víctor González-Castro, **“**Classifying suspicious content in tor darknet through Semantic Attention Keypoint Filtering**”**, Digital Investigation, Elsevier, May 2019.

[5] https://privacy.net/what-is-tor/

[6] https://www.google.com/search?q=tor&sxsrf=ACYBGNQLOE1LFLakrWQR3-dlabeH_vV4g:1570238076861&source=lnms&tbm=isch&sa=X&ved=0ahUKEwipyqLW-IPlAhUJiXAKHesYBeoQ_AUIEygC&biw=1366&bih=657#imgrc=uJ8dQW549fE_VM:

[7] https://www.sans.org/reading-room/whitepapers/forensics/tor-browser-artifacts-windows-10-37642

[8] Kota Abe and Shigeki Goto, **“**Fingerprinting Attack on Tor Anonymityusing Deep Learning**”,** Proceedings of the APAN – Research Workshop, 2016.

[9] Mhd Wesam Al-Nabki, Eduardo Fidalgo, Enrique Alegre, Laura Fernandez-Robles , **“**ToRank: Identifying the Most Influential Suspicious Domains in the Tor Network**”**, Expert Systems With Applications, Elsevier, January 2019.

[10] Muir M, Leimich P, Buchanan WJ, “A forensic audit of the Tor Browser Bundle”, Digital Investigation, Elsevier, March 2019.

[11] Natalija Vlajic, Pooria Madani & Ethan Nguyen, **“**Clickstream tracking of TOR users: may be easier than you think**”**, Journal of Cyber Security Technology, Taylor & Francis, August 2018.

[12] Rebecca Nelson, Atul Shukla and Cory Smith, “Web Browser Forensics in Google Chrome, Mozilla Firefox, and the Tor Browser Bundle”, Digital Forensic Education, Springer, 2019.