iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: III Month of publication: March 2023

DOI: https://doi.org/10.22214/ijraset.2023.49599

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue III Mar 2023- Available at www.ijraset.com

The Dark Web: Privacy and Anonymity

R. Kavitha[1], R. Kapilsurya[2], V. Shanmugam[3], R. Saran Kumar[4]
[1]Department Of ECE, University College of Engineering, Kanchipuram.
[2, 3, 4]Department of Cybersecurity, Paavai Engineering College, Namakkal

Abstract: The Internet as a whole is a network of multiple computer networks and their vast infrastructure. The network consists of websites accessible through search engines such as Google, Firefox, etc. It's called the Surface Web. The Internet is further subdivided into the Deep Web – content that is unindexed and inaccessible to traditional search engines. The Dark Web is considered part of the Deep Web. It is accessible via TOR. Participants on the dark web are anonymous and hidden. Anonymity, confidentiality and the possibility of not being detected are three factors offered by special browsers like TOR and I2P. In this article, we discuss the impact of the Darknet in different social areas and provide results. It gives the number of daily anonymous darknet users (with TOR) in Kosovo and worldwide at the time showed the impact of hidden service sites and got search engine results from Ahimia and Onion City Dark Web. Anonymity on the dark web is not sufficiently verified. TOR strives to provide anonymous activity and aims to provide anonymous activity. Here are the results for showing the number of users and their location. calculations are based on IP addresses by country code, and accesses the sources of those addresses and reports the numbers in aggregate form. In this way,dark web users are indirectly represented. Another key factor is the number of anonymous network users on the dark web. In such a network, users are counted based on requests from directory clients (via TOR metrics) and the relay list is updated. Indirectly, the number of users calculated for the anonymous networks.

I. INTRODUCTION

Many people think that the Internet and the web are synonymous. In fact, they are two different terms with common elements. The internet consists of various networks and its vast infrastructure. It achieves the connection of 1 million computers by creating a network where any computer can communicate with other computers as long as they are connected to the Internet [1]. The Web, a medium, provides access to information. Conceptually, the web is content consisting of websites accessed through search engines such as Google, Firefox, etc. This content is known as the "Surface Web" (Figure 1) [2] [3] [4] [5].

The other part of the Internet is the Deep Web (Figure 1), which refers to the category of its content, which for various technical reasons is not indexed by search engines and cannot not be searched by us via traditional search engines It includes information about private networks and intranets (institutions, universities, companies, commercial databases, etc.), sites with query contents or search forms. The Deep Web is subdivided into the Dark Web (Figure 1). content is intentionally hidden and inaccessible by standard web browsers [2] [4].

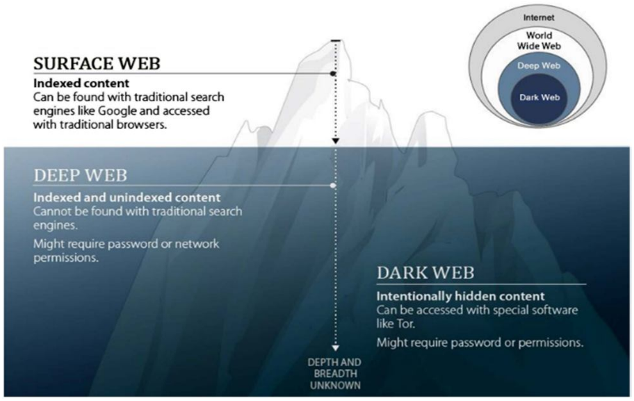
SURFACE WEB
Indexed content
Can be found with traditional search engines like Google and accessed with traditional browsers.
Internet
World Wide Web
Deep Web
Dark Web
DEEP WEB
Indexed and unindexed content
Cannot be found with traditional search engines.
Might require password or network permissions.
DARK WEB
Intentionally hidden content
Can be accessed with special software like Tor.
Might require password or permissions.
DEPTH AND BREADTH UNKNOWN

Figure 1. The Internet layers

The publishers of these sites on the dark web are anonymous and hidden users access the dark web to share data with little risk and go unnoticed (anonymously). Anonymous user access is essential for the Dark Web, which has recently been supported by encrypted tunnels for surveillance protection. Dark web content is powered by The Onion Routing (TOR). This is the anonymous network and accessible through the TOR browser.

The TOR project was started by the US Naval Research Laboratory in 2002 to enable anonymous communication online. Invisible Internet Project (I2P) is another network on the Internet that has traffic on its borders, used for anonymous communication, encryption of user traffic, etc. It offers greater robustness and reliability than the network [1] [5]. In the dark web, users are less likely to be tracked, which facilitates both legal and illegal activities in the traffic of this part of the Internet [2]. Darknet can be accessed through anonymous and decentralized nodes of certain network groups (TOR or I2P). TOR is the name of the software that we install on our computer and which takes care and manages the computer network to which it is connected. It allows users to access websites through a virtual tunnel, and people and organizations can distribute data over public networks without compromising their privacy. TOR allows users to route their traffic through the "user's computer" so that the traffic cannot be traced back to the original user and hides their identity. To transfer data from one layer to another, TOR creates "repeaters" on computers that transmit information through its tunnels across the world of Encrypted information is placed between relays. TOR traffic passes through three relays in its entirety before being routed to its final destination [2]. The "output relay" is called the final relay. The relay IP address corresponds to the TOR traffic source. By using TOR, the software was able to hide dresses from the user. Browsing a website through TOR can reveal links to a given web page (with TOR's output relay IP address in the background). For added security, anonymity, and privacy during communications, individuals should use email, web chat, or similar communication platforms hosted in TOR [2].

II. RELATED WORK

There is a growing number of research papers and projects related to the Darknet. In terms of the work, importance and substance related to , this project is at the center of the improvement of the national supervision of [6]. The arms swap and the emergence of child pornography can easily be done with the help of the dark web. The network analysis distribution uses the TOR network, allowing users to easily penetrate the process of anonymity. Therefore, in order to provide an in-depth analysis of , various literature works have studied the improvement of , thus providing other TOR and routing principles of various intelligence systems in the United States. It can not only activate Web Dark process for legitimate purpose but also activate Web Dark process for illegal purpose. The description of system privacy and the appropriate analysis of trackers- users of the network are easily described for the analysis of facts and the pursuit of research using the research framework ISI- works [7] . The literature review was carried out based on the A detailed study of the different parts of the dark web which are explained appropriately. The study also helps in describing relevant facts about the research done by the researcher. In another work by Barnett et al., [8] The role of spiders (defined as software programs used to transmit information over the World Wide Web) and by manipulating the registration process can be easily accessed in order to obtain accurate and desired information about kinds different can be easily collected, Search. 4044 Social network analysis (SNA) is a topic of interest in [9], and 4044 is performed with the aim of obtaining a graph-based approach, so the analysis of network 4044 groups becomes easier with the description of groups or population strength. The impact of social interactions is extensively described through the use of social networks, allowing easy identification of real-world networks. 4044 Various SNA techniques have been developed to examine forum postings and web site 4044 relationships. The focus is on understanding the "Dark Web" and its unique properties [10]. A detailed coding scheme has been developed to assess extremist websites and terrorist content. Sentiment and impact analysis can identify violent and radicalized sites that pose a significant threat [11]. Terrorism informatics is the application of advanced information fusion, analysis techniques and methodologies to process, integrate, manage and analyze the diversity of terrorism-related information for international/national security applications. Technology comes from disciplines such as computer science, math, science, statistics, social science, public policy and linguistics. Studies have shown that terrorism involves a large amount of information from different sources, languages, data types, information fusion and analysis, such as text mining, data mining, language translation, data integration, video and image processing can help detection and prevent terrorism- [12]. Identifying fraud and theft is important nationally and internationally, as criminals can escape using fake identities, and smugglers can also enter the country with fake visas or passports [13] . In Internet Fraud, Cyber Hacking, Hacking, Illegal Trade, Hate Crime, Virus Spread, Internet Pornography, Internet Privacy, Theft of Confidential Information and Cyber Terrorism, Drug Trafficking and Terrorism Without Borders and is a worldwide security problem.

III. TECHNIQUES, ATTRIBUTES, ACCESSING AND COMMUNICATION IN THE DARK WEB:

Anonymity on the darknet [14] is derived from the Greek word “anonymia”, meaning hiding one's identity from others. When we perform an action on the network, our fingerprints are stored as data on the Internet. If Internet Protocol addresses cannot be traced, anonymity can be said to be guaranteed. TOR clients route global internet traffic through a network of volunteer servers. This hides user information and prevents any possibility of monitoring activity.

The dark web also has a negative impact, allowing criminals to commit cybercrimes and cover their tracks [15]. It is considered an appropriate channel for governments to exchange secret documents, journalists to circumvent censorship, and dissidents to escape the possibility of authoritarian regimes. Onion 1 technology enables anonymous communication over a computer network. Messages are sent encrypted (using asymmetric encryption) and then they are sent through certain network nodes called onion routers. When a message is sent to an onion router, each onion router removes the encryption layer in the same way the onion skin is removed, so as not to discover the routing instructions, so the message is sent to another router, and the process is repeated until it is sent to a specific destination (Figure 2). This technique secures intermediate nodes to form "notifications" of message source, destination, and content [1] [14] [16].

IV. ONLINE PRIVACY IN THE DARK WEB

TOR is used to enable private, anonymous and secure communications and activities for specific purposes [2] [14]. In the following are given some examplesthat they are related to above mentioned elements:

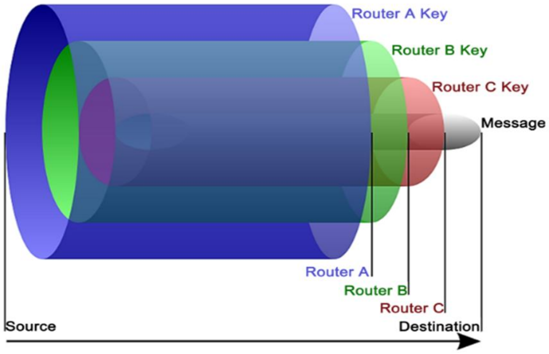

Figure 2. The message routing by using the onion technique

1) Anti-censorship and political activity. To avoid censorship and reach certain destinations or content blocked in one way or another, TOR considers it an adequate tool. This allows individuals to access content that may be blocked in certain parts of the world. To prevent this, some governments create rules for using TOR or block access to TOR for a specific period of time. TOR is also used by dissidents to secure and keep their communications and locations anonymous. One such case is the dissident movement in Iran and Egypt[2].
2) Sensitive communication. If individuals wish to access chat rooms or forums, and prefer sensitive communications for personal or business purposes, this is enabled by TOR. It is used to protect children on the line (internet browsing) from abuse (by hiding their device's IP address). Companies can use this tool to secure their projects and protect competitors' spies [2] [17] [18].
3) Disclosure of information. Journalists can use TOR to communicate with informants and dissidents [2] [18]. Personally, I think it is possible to communicate and share documents anonymously with TOR editors (eg New Yorker's Safe). Edward Snowden used Tail (a operating system for anonymity) running in TOR. He reported to reporters and communicated , requesting the release of classified documents regarding the US defense program. Snowden released top secret document which describes how the National Security Agency (NSA) attempted to de-anonymize users to using the TOR browser [2].

V. DARK WEB IN THE GOVERNMENT, MILITARY AND INTELLIGENCE

The dark web can be an online playground for criminals due to the anonymity provided by other software like Tor and I2P. However, as previously stated, there are many areas where researching and using the dark web can be beneficial. This applies not only to citizens and businesses seeking privacy online, but also to certain branches of government, namely law enforcement, military, and intelligence. Dark web anonymity can be used to protect military command and control systems in the field from adversary identification and hacking. The military can use the dark web to research its operating environment and identify activities that pose operational risk to troops. For example, there is evidence that the Islamic State (IS) and the support group seek to exploit the anonymity of the dark web for activities other than information sharing, recruiting and spreading propaganda., using Bitcoin to fund's operations. In the fight against ISIS, the Department of Defense (DOD) can monitor these activities and use various tactics to foil terrorist plots [19].

The military can use TOR software for covert or covert computer network operations, such as shutting down websites or denial of service attacks, or intercepting and suppressing enemy communications. Another use could be military deception or psychological operations, with the military using the Dark network to sow disinformation about troop movements and targets for counterintelligence purposes, or to disseminate information to discredit claims. rebels. These activities may support ongoing military 4044 operations or be conducted independently [20]. The Defense Advanced Research Projects Agency (DARPA) of the Department of Defense is conducting a research project called Memex to develop a new search engine capable of capturing patterns and relationships in online data to aid law enforcement and other stakeholders to detect illegal activities. Commercial search engines such as Google and Bing use algorithms to present search results by popularity and ranking, and capture only about 5% of the Internet [20]. By scanning 4,444 websites often overlooked by commercial search engines and capturing 4.444 million sites hidden on the dark web, the Memex project ultimately aims to build a more comprehensive map of Internet content. Specifically, the project is currently developing technology to “find related signals related to prostitution advertisements on popular websites” [21]. It is intended to help 4,444 law enforcement agencies in their investigations into human trafficking [21]. Similar to the military's use of the Dark Web, it is no secret that the Intelligence Community (IC) uses it as a source of public intelligence, although many relevant details are classified. According to Admiral Mike Rogers, Director of the National Security Agency (NSA) and Commander of U.S. Cyber Command, they "spend a lot of time looking for people who don't want to be found"[22] . Investigations into the NSA's XKeyscore program - one of programs revealed by Edward Snowden's classified information leaks - reportedly indicated that any user attempting to download TOR was automatically given an electronic fingerprint on , potentially allowing agencies to identify us considered untraceable [23]. While specific IC activities related to the Deep Web and Dark Web can be classified as , a least one Intelligence Advanced Re-Search Projects Activity (IARPA) related program can be related to searching for data located on the Deep Web Store[24]. Traditional tools such as detection based on the signature would not allow researchers to predict cyber threats; thus, managers are rethinking rather than anticipating and mitigating these attacks [25]. CYBER Attacks Automated Unconventional Sensor Environments (CAUSE) Program aims to "develop and test new automated methods to predict and detect CYBER attacks earlier than existing methods." [26]. It can use patterns of actor behavior and factors such as black market sales to help predict and detect cyber incidents [26].

VI. PAYMENT ON THE DARK WEB

Bitcoin is a currency often used in transactions on the dark web [27]. It is a decentralized digital currency that uses anonymous, peer-to-peer transactions [28]. Individuals typically obtain bitcoins by accepting them as payment, exchanging them for traditional currencies, or “mining” them [29]. When bitcoins are used for financial transactions, the transactions are recorded in a public ledger known as the blockchain. The information recorded in the blockchain is the bitcoin address of the sender and the recipient. An address does not uniquely identify any particular bitcoin; rather, the address simply identifies a particular transaction [30]. User addresses are associated and stored in wallets [31]. A wallet contains an individual's private key [32], which is a secret number that allows that individual to spend bitcoins from the corresponding wallet [33], similar to a password. and the cryptographic signature of the transaction are used to verify the transaction [32]. Wallets and private keys are not registered in public ledger; this is where using Bitcoin adds privacy. The wallets can be hosted on the network by software or hardware devices [34] of computers or mobile devices.

VII. RESULTS AND DISCUSSION

The results are based on research questions (RQs) and focus on TOR statistics and various reliable (anonymous) dark web privacy reports and information. Through them, arguments about anonymity and privacy are given for different situations. We give eight QRs as follows:

A. (RQ1) How many users use TOR software in Kosovo?

Based on the data we generated from TOR measurements, we found that Kosovo TOR software had nearly daily users between January and December 2018. This number increased and decreased by over the course of this period (Fig. 3).

B. (RQ2) (How many anonymous internet users are there in the world?

According to the TOR indicator, more than 4 million users worldwide anonymously used the Internet daily between January and December 2018, and this number decreased after the first two months of 2018 and the second month of the same year (Figure 4).

C. (RQ3) Impact of activity (%)(Hidden services sites - sites) In the darknet?

Based on findings from the University of Portsmouth,researchers worked with 40 relays (computers) running on the TOR network, and they collected more than 45,000 TOR hidden service websites. The researchers concluded that 2% of them targeted child abuse and 83% of visitors visited these websites. Another study of TOR hidden service sites was implemented by King's College London through search engines such as Ahmia and Onion City (for dark web). The researchers directly identified 5,205 sites. 1. Of these, 557 were identified as by illegal content. According to the TOR project, traffic to hidden services is estimated at 3.4%. Between March 2016 and March 2017, 50,000 - 60 existed.

D. (RQ4) Can anonymity be verified in the Dark Web and can we say that it is the anonymous content?

It cannot be said that anonymity is fully verified on the dark web. TOR was designed to allow anonymous activity, but researchers and security experts have worked hard to develop tools through which they can identify and anonymize hidden individuals or services. There are many cases of Anonymous (examples), but to broaden this research question, consider two of them:

1) The FBI took control of Freedom Hosting 5 in 2013, even though it was infected by Malware designed to identify visitors. Since 2002, the FBI has used the "Computer and Internet Protocol Address Verifier" [2], which is malware on the Freedom Hosting Network's hosted service, although it uses proxies or anonymous services such as TOR to identify and authenticate suspects and their locations.

2) In 2017, hackers belonging to Anonymous reactivated and took control of Freedom Hosting II, a dark web hosting service and predecessor of Freedom Hosting. They claim that more than 50% of content on Freedom Hosting is related to sensitive content. The user who posted this data on Freedom Hosting can easily be identified. security experts have concluded that Freedom Hosting II hosts 1,500-2,000 hidden services (of which nearly 15%-20% are considered active sites) [2].

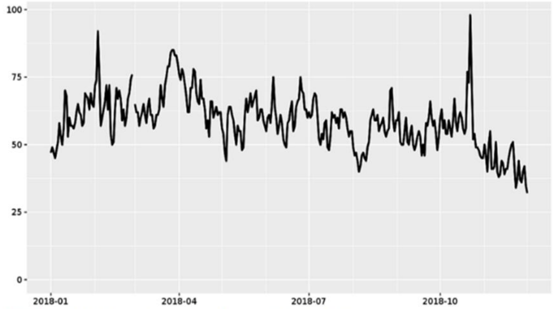

Figure 3. The TOR daily users in Kosovo during January to December 2018[3].

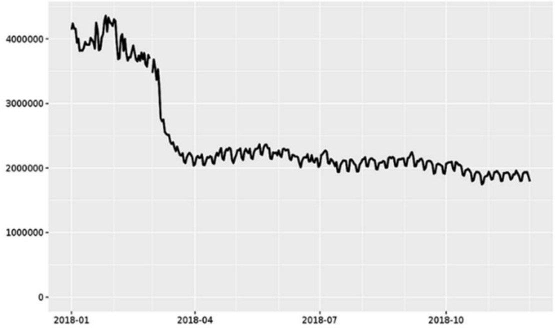

Figure 4. The TOR daily users in the world that they have used the Internet anonymously during January to December 2018[4].

E. (RQ5) How the number of users is retrieved from the directory requests through TOR and in what way does their calculation become?

There are mechanisms in TOR that assume that client makes an average of 10 requests per day. A TOR client, if connected to the Internet 24/7, can make about 15 requests per day, but not all clients stay connected to the Internet 24/7, so each client needs an average of 10 re- . The total number of user directory requests is divided by 10 and this is the number of users found. Another way to calculate the number of users, , is to assume that each query represents a customer browsing the web 1/10 (or 2 hours and 24 minutes) per day.

F. (RQ6) How do we know from which countries are the Dark Web users and in what way does their reporting become?

These directories break down IP addresses starting from by country code, where they are looked up and report the numbers in aggregate form. The digits indirectly represent dark web users. Since reports are made in forms like , it is believed that this is why TOR ships are associated with GeoIP databases.

G. (RQ7) How can censorship events be identified/calculated through TOR?

There is an anomaly-based censorship detection system6 that counts the number of users for a range of days and predicts the number of users for the next few days. If the current number of users estimated by the above system is high, it can be concluded that there may be a censorship event, otherwise not. For more details on this issue, please see the related report.

H. (RQ8) How can users be numbered in an anonymous network in the Dark Web (according to the TOR Metrics)?

According to TOR metrics the number of users is not counted directly, but directory requests are numbered periodically for clients in which case the relay list is updated. Therefore, according to the above, indirectly counts the number of users of an anonymous network.

VIII. CONCLUSIONS

Darknet networks like TOR provide many opportunities for malicious actors to trade legal and illicit "goods" anonymously. The dark web is a growing asset for, especially when it comes to illegal services and activities. Security mechanisms need to be aware of these issues 4044 and take steps to eliminate them. The evolution of technologies with encryption (security) and anonymity, such as the Darknet and its special software, challenges law enforcement and policymakers to effectively combat malicious actors operating in cyberspace. In this article separately discusses the impact of the dark web, its privacy and anonymity, through the results, it becomes anonymous users of this Internet segment in the Kosovo region the daily number of users and in the world and how many hidden services on the Dark Impact on the website. The results in this section were collected from the Ahimia and Onion City search engines (for the dark web). We concluded that anonymity cannot be fully verified on the dark web, even though TOR is dedicated to the net- work segment designed to provide anonymous activity. This is the reporting aspect of users in the country where are also being tracked. In this case, the directory breaks down the IP address based on the country code in the where the visit originated and reports the aggregate count. The digits indirectly represent dark web users. The number of users of dark web anonymous networks is not directly counted. This calculation is performed by the TOR metric, which counts client requests for the directory and, if so, updates the list of relays. Indirectly, the number of anonymous network users is calculated as a given case in the results of this article.

REFERENCES

[1] Chertoff, M. and Simon, T. (2015) The Impact of the Dark Web on Internet Gover-nance and Cyber Security.Centre for International Governance Innovation and Chatham House, 6, 1-18. https://www.cigionline.org/sites/default/files/gcig_paper_no6.pdf

[2] Finklea, K. (2017) Dark Web. Congressional Research Service, Washington DC, 10 March 2017, 1-19. https://fas.org/sgp/crs/misc/R44101.pdf

[3] Ilou, C., Kalpakis, G., Tsikrika, T., Vrochidis, S. and Kompatsiaris, I. (2016) Hybrid Focused Crawling for Homemade Explosives Discovery on Surface and Dark Web. Proceedings of the 11th IEEE International Conference on Availability, Reliability and Security, Salzburg, Austria, 15 December 2016, 1-6. https://doi.org/10.1109/ARES.2016.66

[4] Park, A., Beck, B., Fletche, D., Lam, P. and Tsang, H. (2016) Temporal Analysis of Radical Dark Web Forum Users. Proceedings of the IEEE ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), SanFrancisco, CA, USA, 880-883. https://doi.org/10.1109/ASONAM.2016.7752341

[5] Berghel, H. (2017) Which Is More Dangerous—The Dark Web or the Deep State? Computer,IEEE Computer Society, 50, 86-91. https://doi.org/10.1109/MC.2017.215

[6] Arora, M., Kanjilal, U. and Varshney, D. (2012) An Intelligent Information Retrieval: A Social Network Analysis. International Journal of Web Based Communities, 8, 213-222. https://doi.org/10.1504/IJWBC.2012.046263

[7] Wu, P. and Li, S. (2011) Layout Algorithm Suitable for Structural Analysis and Visualization of Social Network.Journal of Software, 22, 2467-2475. http://pub.chinasciencejournal.com/JournalofSoftware/18611.jhtml https://doi.org/10.3724/SP.J.1001.2011.03896

[8] Barnett, G. and Jiang, K. (2016) Resilience of the World Wide Web: A Longitudinal Two-Mode Network Analysis. Social Network Analysis and Mining, 6, 1-105.

[9] Egoryan, L. (2015) New Fields for Web-Analysis: Social Network and Blogs.Auditor, 1, 43-48.

[10] Davies, P. (2008) Information Technology. Oxford University Press, Oxford. https://www.worldcat.org/title/information-technology/oclc/374881000

[11] Clifton, B. (2012) Advanced Web Metrics with Google Analytics. Wiley, Hoboken,NJ. https://www.wiley.com/en-us/Advanced+Web+Metrics+with+Google+Analytics%2C+3rd+Edition-p-9781118168448

[12] Chen, H. and Yang, C. (2008) Intelligence and Security Informatics. Vol. 135, Springer, Berlin and Heidelberg, 1-460. https://www.springer.com/gp/book/9783540692072 https://doi.org/10.1007/978-3-540-69209-6

[13] Yang, C., Mao, W., Zheng, X. and Wang, H. (2013) Intelligent Systems for Security Informatics. Academic Press, Cambridge, Massachusetts, United States, 1-250. https://www.elsevier.com/books/intelligent-systems-for-security-informatics/yang/978-0-12-404702-0

[14] Jardine, E. (2015) The Dark Web Dilemma: Tor, Anonymity and Online Policing. Centre for International Governance Innovation and Chatham House, 20, 1-24. https://www.cigionline.org/sites/default/files/no.21.pdf

[15] Baravalle, A., Lopez, M.S. and Lee, S.W. (2016) Mining the Dark Web—Drugs and Fake IDs. Proceedings of the 16 th IEEE International Conference on Data Mining Workshops, IEEE, Barcelona, 12-15 December 2016, 350-356. https://doi.org/10.1109/ICDMW.2016.0056

[16] Chen, H. (2012) Dark Web—Exploring and Data Mining the Dark Side of the Web.