Unveiling the Hidden: Exploring Challenges in Dark Web Investigation Using Measurement Sensors

Vinod Babu Bollikonda[1,*], KVD Kiran[1]

[1]Department of Computer Science and Engineering, Koneru Laxmaiah Education Foundation, Vaddeswaram, AP, India

E-Mails: 2002031023@kluniversity.in; kirancse@kluniversity.in

Abstract

This study is centered on the possible methods to analyze and investigate dark web crimes by technical and non-technical users such as law enforcement agencies. Also, the study focuses on learning anonymity procedures used by malicious actors to hide their identity on the dark web and identify the challenges to making a network-level investigation. The other objective is to study the proven methods to determine the hidden services directory (HSDir), active marketplaces, crawling and indexing of the dark web pages. Methods: A Proof of Concept (PoC) experiment explores multi-level anonymity techniques used by malicious actors. Level one involves using a commercial VPN to hide system details, and level two employs a hypervisor, MAC changer, proxy server, and the Tor network. The results reveal the complexities of Tor anonymity and provide insights into the methods employed by malicious actors. The proposed methodology offers a comprehensive approach to understanding and investigating dark web crimes, combining website fingerprinting, open-source intelligence, and threat intelligence data. Findings: Investigation teams face challenges as the proven and tested methods of previous works in this study, such as network-level bulk datasets and webpages fingerprinting dataset analysis, are technology-intensive and non-technical users will face challenges. Usage of Anonymous tools and techniques used at the host level (VM), Mac change, VPN and Tor network complicates the investigation to track and trace the activities. Tor browser has hopped through random nodes to anonymize the connection before connecting to the marketplace. MAC Changer will change the Mac address flashed on the network card by the device manufacturer to anonymize the system-level details. Novelty: Identified the requirement of a comprehensive and novel methodology that is adaptable to investigate dark web crimes by the technical and non-technical teams of law enforcement an agency is proposed in this study. This methodology includes website fingerprinting, OSINT and threat intelligence data collected from various sources. This methodology shall evolve with phase-wise steps of proven techniques such as crawling, indexing, attribute-based analysis, and dataset creation to obtain actionable intelligence proposed in this paper to investigate and eradicate dark web crimes.

Keywords: Dark web; anonymity; hidden services; cybercrimes; tor

1. Introduction

The International Telecommunication Union (ITU) estimated in 2021 that approximately 4.9 billion people, around 63% of the world's population, are using the internet. On a negative note, this growth has opened new avenues in cybercrime [1] due to internet technologies distributed and anonymous nature. According to János Besenyő, Attila Gulyas, [2] the content over the internet is distributed and referred to as Surface, Deep and Darkweb. Any type of internet content crawlable and indexed by the search engines such as Google, Bing and Yahoo is known as surface web; Deep web holds classified and special-purpose data such as corporate workflow, emails, medical records,

DOI: https://doi.org/10.54216/JCIM.150113
Received: February 04, 2024 Revised: April 25, 2024 Accepted: July 27, 2024

banking transactions, and government data accessible only with specific login credentials. The dark web has often referred to as a subset of the deep web. According to Matthew Robert Shillito [3] dark web hosts illegal content and services that are most harmful, including classified information, malware, pornography, red rooms, narcotic drugs, arms and ammunition, scams, money laundering, hire to-kill and more. Law enforcement agencies and other enterprises are facing challenges in eradicating crimes. The primary objective of this paper is to study dark web crimes and their types and intricacies of anonymity methods used to access the dark web and to identify the technical and non-technical investigation approaches and analytical methods helpful to law enforcement agencies in the eradication of dark web crimes. For instance, one prominent case is the WannaCry ransomware attack in 2017, which affected hundreds of thousands of computers worldwide. This attack exploited a vulnerability in the Windows operating system, encrypting users' files and demanding ransom payments in Bitcoin. The widespread nature of this attack and the significant financial losses incurred by individuals and organizations underscore the serious threat posed by cybercrime. Furthermore, the proliferation of hacking groups and the increasing sophistication of their tactics highlight the evolving nature of cybercrime. Organizations are constantly targeted by these malicious actors, who exploit vulnerabilities in systems and networks to gain unauthorized access, steal sensitive data, or disrupt operations. The high-profile breaches of companies like Equifax, Yahoo, and Sony serve as reminders of the far-reaching consequences of such cyber-attacks. Further to this study, we aim to define the scope for a comprehensive methodology to obtain actionable intelligence for investigating and eradicating dark web crimes.

2. Analysis Method

The thematic analysis method was used to research the qualitative data collected from various research papers. The data were examined to understand the anonymity intricacies and its impact and identify the proven methods and approaches for analysis and investigation.

2.1 Dark network – popular technologies

The dark web is accessible via a specialized network such as Freenet, The Invisible Internet Project (i2p) shown in Figure 1 and The Onion Router (TOR) protocol shown in Figure 2. The dark network is decentralized and runs over the computers referred to as nodes, which are connected to share the computing, storage and network resources using different algorithms and protocols. ToR is one of such preferred mentioned below.

2.2 Freenet

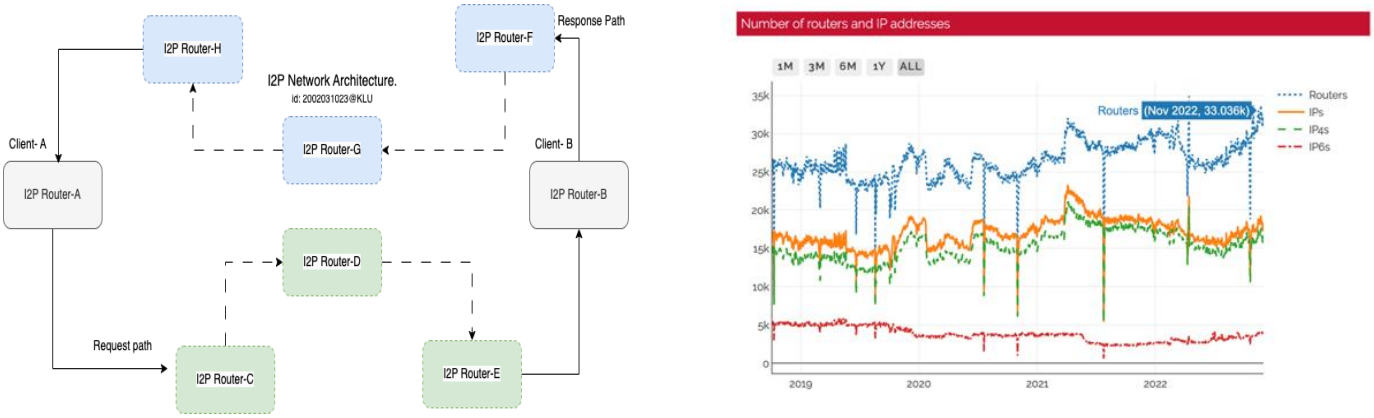
Freenet is a peer-to-peer network protocol designed and developed by Ian Clark [4] for discrete file sharing and anonymous and secure communication. Its primary objective is censorship-free file sharing and publishing primarily used. The freenet is an open-source software application (https://github.com/ freenet/fred). It works in two operational modes, and each node voluntarily contributes a certain amount of storage space to the freenet. The file is uploaded into storage space contributed by nodes. The 'manifest key' is a URI returned to the user to retrieve and reconstruct the encrypted and stored original document. It secures the file using an encryption method and divides each into 32KB blocks before inserting it into the freenet. A manifest key is also known as the manifest block that contains the decryption key and block hashes required to retrieve files.

2.3 The Invisible Internet Project (I2P)

Roberto Magán-Carrión et al. [5] The Invisible Internet Project (I2P) is an open-source technology developed to anonymously render and access hidden and discrete internet services. Though I2P is like ToR, I2P provides faster accessibility and better Anonymity than ToR for a few applications. I2P is a network designed on a group of virtual routers. The unique features of I2P have attracted a diverse user base, including both legitimate users and malicious actors seeking to exploit its advantages for illicit purposes. This necessitates a deeper understanding of I2P's inner workings and the development of effective methodologies for investigating crimes related to this network. For instance, consider the scenario where an I2P connection is established between two clients named 'Vin' and 'Bob'. Let's delve into the intricacies of their communication process by examining the following sequential steps.

DOI: https://doi.org/10.54216/JCIM.150113
Received: February 04, 2024 Revised: April 25, 2024 Accepted: July 27, 2024

1. Initialization: The process begins with the initialization of both Client Vin and Client Bob, ensuring that they are ready to establish a connection.
2. I2P Router 1: The flow chart shows the first I2P router that Client Vin's communication passes through. This router functions as a relay, ensuring anonymity, and secure transmission of data.
3. I2P Router 2: The flow chart illustrates another I2P router that the communication encounters. This router serves as a redundancy measure, providing an alternative path in case of failure or congestion in the primary route.
4. I2P Router 3: Like I2P Router 2, this router acts as a failover path, further enhancing the reliability and resilience of the communication channel.
5. Failover Decision: At each I2P router, there is a failover decision point. If the primary path fails or becomes congested, the communication can switch to the alternative path provided by the failover routers.
6. Client Bob: Finally, the communication reaches Client Bob, ensuring secure and private transmission.



(a) (b)

Figure 1. (a) I2P Network Architecture; (b) Usage statistics.

2.4 The Onion Router (TOR)

L. Basyoni et al. [6] The Onion Router (Tor) network takes first place in hosting and propagating dark web hidden services and illegal activities. The Onion Routing project in 2002 was developed by the United States Naval Research Laboratory ('NRL') to protect and anonymize US intelligence communication. ToR is a first-of-its-kind network with a complete decentralized hidden services deployed on distributed network nodes spread across the globe hosted by volunteers.

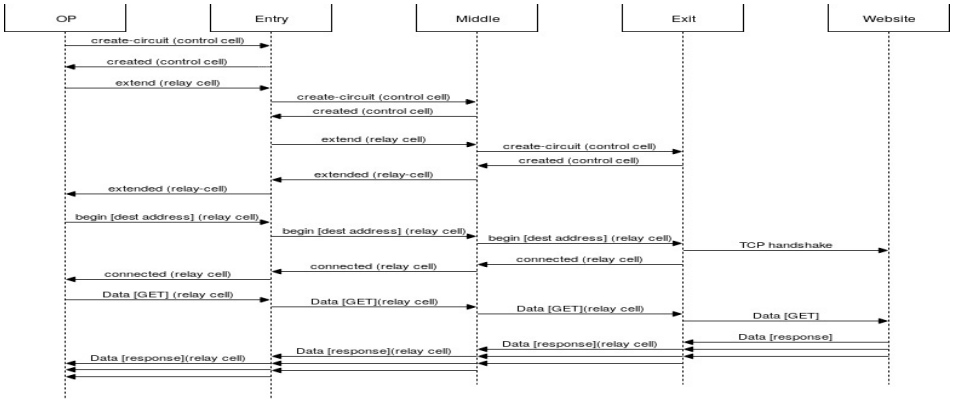


Figure 2. Distributed network of active tor nodes

Tor metrics indicate the increased usage of the user connected directly to the ToR network as relay nodes. The below

DOI: https://doi.org/10.54216/JCIM.150113
Received: February 04, 2024 Revised: April 25, 2024 Accepted: July 27, 2024

table provides the top 10 countries of 'mean daily users' connected from December 2022 to December 2023 mention in Table 1.

Table 1: Top-10 countries by relay users.

S.No	Country	Mean daily users	Usage %
1	United States	536548	21.35
2	Germany	300841	11.97
3	Finland	118191	4.70
4	Russia	107020	4.26
5	India	103050	4.10
6	France	88818	3.53
7	Indonesia	86136	3.43
8	Netherlands	82289	3.27
9	U K	62891	2.50
10	Brazil	44408	1.77

3. Darknet Markets Techniques and Crime

3.1 Darknet markets and crimes

Upulie Handalage et al. [7], the dark web is the hub for nefarious activities of criminals such as financial fraudsters, gamblers, paedophiles, terrorists, and state-sponsored threat actors. The dark web marketplaces are the hidden services deployed over the ToR network. Dark web marketplaces are broadly classified into two types (i.e.) Market Places and Vendor shops where Market places are the business-to-business (B2B) and business-to-customer (B2C) online e-commerce platforms on the dark network that hosts the sales of illegal products such as drugs, fraudulent data, stolen products, classified information from individuals and organized crime groups and vendor shops are the individuals or organized crime groups own online shops over the dark network to sell illegal products and services.

3.2 Dark Web Crimes

Martin J et al. [8], the dark web is the hub of illegal activity, including illicit drug trading. Persi Paoli, Giacomo et al. [9], illegal Arms and Ammunition. Bruggen, Madeleine et al. [10], Child Pornography (CP), Human Trafficking. Asalah Altwairqi [11] Cannabis and other narcotic drugs, Terrorism Frauds – Money Laundering. Crypto scams, illegal Organ Trading

3.3 Anonymous Techniques and Tools

Anonymization is a method of data traffic modification on the network to protect user identities from threat actors. The idea behind any anonymization method is to eradicate the possibility of identifying the two endpoints in communication without tampering with or disturbing the data integrity and usefulness. According to T. Farah and L. Trajković. [12] has referred few anonymization techniques such as Truncation, Reverse truncation, Enumeration, Permutation, Prefix-preserving, Pseudonymization, Binning, Hashing, Black marker, Precision degradation, Time unit destruction and others. Fields and processes for anonymization are suggested as follows in Table 2.

Table 2: Anonymization Methods

S.No	**Anonymization Parameter**	**Purpose**
1	IP Address	Truncation,Reverse truncation, Permutation, Prefix-preserving pseudonymization, Black marker.
2	MAC address	Truncation, Reverse truncation, Structured pseudonymization, Black marker.
3	Timestamps	Precise degradation, Enumeration, Random-time shift, Black marker.
4	Counter	Precision degradation, Binning, Random noise addition, Black marker
5	Port number	Binning, Premutation, Black marker.

DOI: https://doi.org/10.54216/JCIM.150113
Received: February 04, 2024 Revised: April 25, 2024 Accepted: July 27, 2024

Montieri, Antonio [13], and others have focused on the vital issue of maintaining privacy using Anonymity Tools (AT) and the possibility of Traffic Classification (TC) of anonymous services such as TOR, I2P and JonDonym. AT supports traffic classification and analysis, and it is possible to fingerprint traffic details to the granular level by a distinction based on the traffic types and hidden services inference. Comprehensive analysis classified it into three parts Anon network (L1), Traffic Type (L2) and Application/hidden service (L3) using machine learning (ML) classifiers. It is further stated that the results of Hierarchical Approach (HC) methods have yielded higher accuracy in determining whether traffic is L1/L2 or L3 traffic and which specific application type.

Kaur, Shubhdeep & Randhawa, Sukhchandan. (2020) [14], Proven anonymization tools and techniques used to access the dark web created more complexity in conducting analysis and moving forward on the investigation. The list of tools used for anonymization includes Anonymizers, Virtual Private Networks (VPN), Proxies (Inbound and outbound), Virtual Machines (with MAC changing ability), PGP Key (to maintain confidentiality and integrity of messages), XMPP / IRC chatroom tools. The anonymous access method differs based on the type of dark web visitor. Usually, malicious actors use commercial virtual private network services or other anonymizer tools over the internet to access Freenet, i2p or ToR-based dark web marketplaces or blogs to connect. It makes it more complicated for law enforcement agencies to investigate illegal activities.

3.4 Anonymity - Proof of Concept (POC)

To identify the real-time challenges in dark web crime investigation, a practical test as proof of concept was conducted using the below setup and steps mentioned in Table 3.

Table 3: Proof of Concept – Software List.

S.No	System Name	OS/Software - Type/Version
1	Host System	MacBook Pro – macOS 13.0.1
2	Windows Virtual Machine	Downloaded Windows virtual machine image provided by Microsoft from https://github.com/magnetikonline/linux-microsoft-ie-virtual-machines. Last seen on 10th October 2022.
3	VMware Fusion Work Station player	Type-2 Hypervisor. https://www.vmware.com/in/products/workstation-player/
4	Nord VPN	Nord VPN Version 7.16.1 (214) – Commercial Licensed VPN service.
5	Tor browser	Version 12.0 - https://www.torproject.org/download/
6	Wireshark	Packet Sniffer Software. https://www.wireshark.org/download.html

Anonymity is a double-edged sword used for positive and negative purposes. In this study, various challenges are observed in dark web crime investigation.

Step – 1: Host Level Anonymization

In this proof of concept, a hypervisor tool is employed to deploy a windows Virtual Machine (VM). To ensure host level anonymization, the TMAC software by Technetium [15], which is a Mac address changer, is utilized. The virtual machine (VM) utilized in this study was equipped with a default virtual Network Interface Card (NIC) named 'Ethernet0 2', which had a MAC address of '00-0C-29-AB-61-60'. To achieve the desired level of anonymity, we employed the 'Change MAC Address' feature in the Technetium software, which generated a random MAC address of '88-96-76-CB-33-A6'. This random MAC address was then applied to the VM, as illustrated in Figure 3, resulting in the desired level of anonymity.

DOI: https://doi.org/10.54216/JCIM.150113
Received: February 04, 2024 Revised: April 25, 2024 Accepted: July 27, 2024

Figure 3. Modified MAC Address

Step – 2: Network Level Anonymization.

To enhance anonymity, a commercial VPN service was utilized to select a desired geographic location and anonymize the host IP address. In addition, an open VPN client software with a freevpnbook proxy was implemented on a virtual machine, adding an extra layer of anonymity. Prior to VPN encryption, network connections were monitored using the 'netstat' command utility on the virtual machine. The netstat details below reveal a direct HTTP connection without any form of encryption or tunneling, as illustrated in Figure 4.

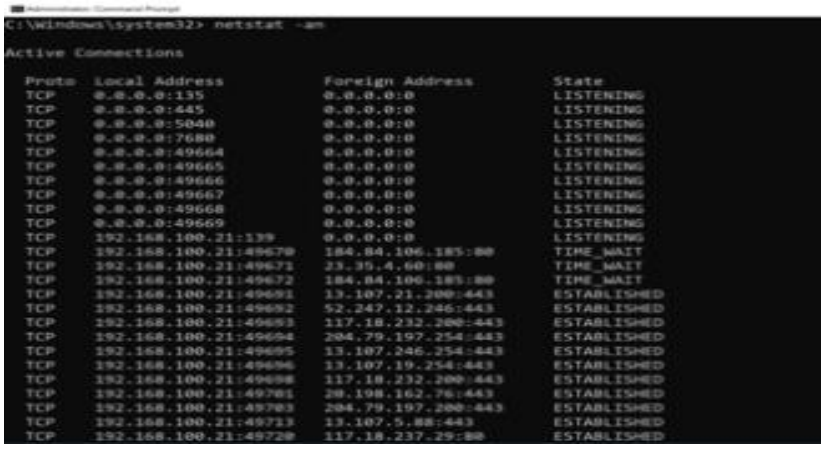


Figure 4. Network Anonymization proof with VPN IP

Step – 3: Route hopping to maintain communication anonymity.

The Onion Router (ToR) browser was utilized to access the alphabay market [17] on the Onion site. By leveraging the relay nodes within the Tor network, communication with the destination was established through an anonymous path, deviating from the default route provided by the Internet Service Provider (ISP). The utilization of encrypted network connections and communication packets, as observed through the analysis on Wireshark [18], highlights the increased complexity in eavesdropping Tor communication.

DOI: https://doi.org/10.54216/JCIM.150113
Received: February 04, 2024 Revised: April 25, 2024 Accepted: July 27, 2024

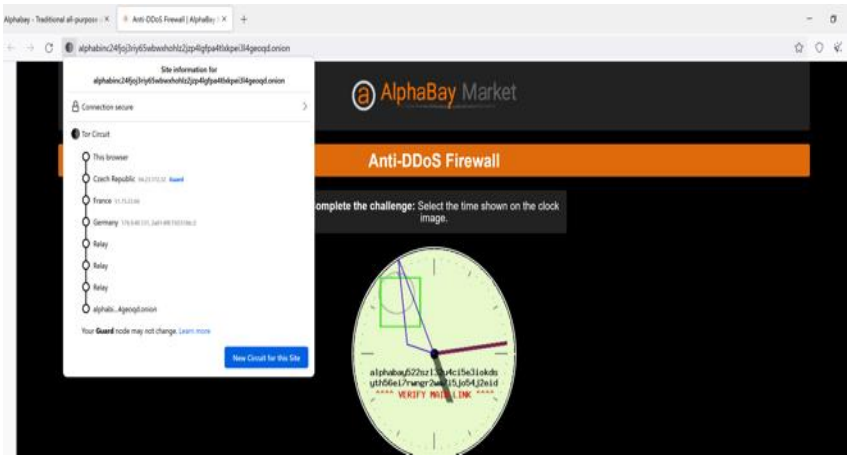


Figure 5. Route hopping and anonymous path navigation

3.5 Host Level-OS Anonymity

The utilization of anonymity techniques at the host system level, such as commercial VPN services and virtual machines, presents a challenge in identifying the true identities of threat actors and malicious users operating on the dark web. Consequently, traditional packet sniffers like Wireshark and Tcpdump are ineffective in capturing the original details of the system.

3.6 Guest Level -OS Anonymity

Virtual machine (VM) is used to access the dark web, and these VMs are usually destroyed after the activity on the dark web to wipe off the tracks. As an alternative to VM, TAILS live operating system (Live OS) with built-in VPN service can be connected to the host system without installation using a removable USB storage drive to make tracing more difficult. Hence, dark web access through the guest OS system applies another layer of anonymity, and it is challenging to identify the exact source.

3.7 Network Level Anonymity

Tor browser communication over entry, relay and exit nodes establish the encrypted virtual private network (VPN) connection. It will pass through the last best-known route to the destination updated on the Distributed Hash Table (DHT) of the ToR network. Hence, eavesdropping on the Tor network to trace malicious activity is challenging.

3.8 Application-Level Anonymity

Dark web applications typically host their services on Tor nodes and establish connections through Tor relay nodes called Rendezvous Points (RPs). The onion addresses and relevant information of these applications are meticulously maintained in the Hidden Services Directory (HSDir), which also acts as a gateway to access the Hidden Services (HS). The communication path to reach the hidden services directory is designed to ensure anonymity by employing route hopping. This intentional complexity poses challenges in tracing intercepted data at the network level, especially when attempting to locate the rendezvous point. In Figure 6, the Hidden Services Directory is represented by the 'green color' node, intricately interconnected with various other nodes responsible for hosting dark web services on the Tor network.

DOI: https://doi.org/10.54216/JCIM.150113
Received: February 04, 2024 Revised: April 25, 2024 Accepted: July 27, 2024

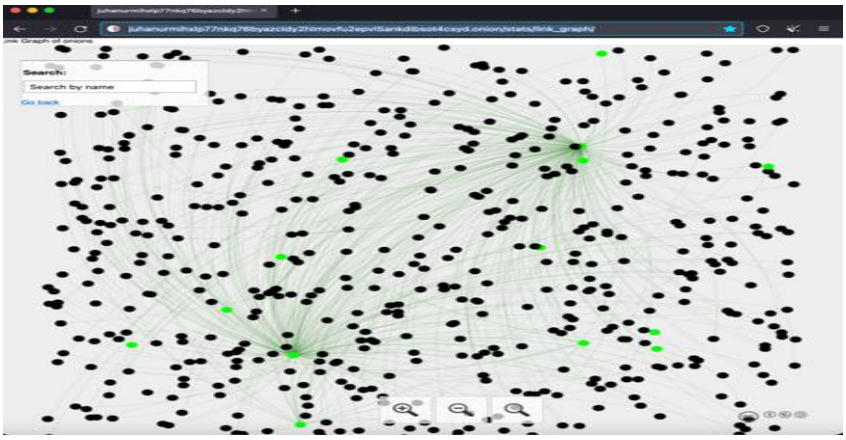


Figure 6. Hidden Services Directory (HSDir).

3.9 First Level Anonymity

In the context of this research, the first level of anonymity was established through the utilization of the commercial VPN 'NORD.' This VPN effectively concealed system details such as the IP address assigned by the Internet Service Provider (ISP) and the MAC address at the host machine.

3.10 Second Level Anonymity

Moving to the second level of anonymity, the research employed the type-2 hypervisor VMware Fusion software on the host machine. A Windows virtual machine was deployed, and the MAC changer software was utilized to alter the MAC address. Additionally, to further obfuscate the IP address assigned by NORD VPN, the research employed the combination of OpenVPN and a freevpnbook proxy server. The Tor browser was utilized to establish connections to Tor nodes, enabling access to dark web sites.

3. 11 Third Level Anonymity

The third level of anonymity in this research involves the utilization of the TOR network, which anonymizes the routing path to the destination application by randomly hopping through TOR nodes. This level of anonymity applies to both surface websites and onion websites on the dark web. By leveraging the TOR network's random node hopping capability, the research ensures that the origin of the connection remains obscured, enhancing the overall anonymity of the user's activities. Though the 'exit' node to surface website communication is unencrypted, using this to obtain the actionable intelligence might be challenging.

4. Measuring the Effectiveness of Anonymization

Evaluating the effectiveness of anonymization methods used to connect to the dark web via Tor presents inherent challenges. To assess the level of anonymity provided by these methods, various approaches can be employed. These include network traffic analysis, statistical analysis, end-to-end timing attacks, fingerprinting resistance, and usage pattern studies, such as analyzing market place postings and other blog communications. Each approach offers valuable insights into the effectiveness of anonymization techniques while also uncovering potential vulnerabilities. In this case, measurement sensors that can provide actionable intelligence include network traffic analysis, statistical analysis, end-to-end timing attacks, fingerprinting resistance, and usage pattern studies.

DOI: https://doi.org/10.54216/JCIM.150113
Received: February 04, 2024 Revised: April 25, 2024 Accepted: July 27, 2024

4.1 First Level Anonymity Measurement Sensor

To assess the effectiveness of first level anonymity, researchers can employ IP address tracking techniques. By monitoring the outgoing IP address from the host machine, researchers can determine whether the assigned IP address matches the IP address of the VPN server. This measurement sensor provides insights into whether the VPN successfully hides the original IP address, validating the first level of anonymity. Challenges: One challenge in measuring first level anonymity lies in differentiating between the actual IP address and the IP address assigned by the VPN. Additionally, VPNs may vary in their capabilities to effectively hide system details, potentially impacting the level of anonymity achieved.

4.2 Second Level Anonymity Measurement Sensor

Measuring the effectiveness of second level anonymity requires a multi-faceted approach. Researchers can employ MAC address tracking techniques to verify if the MAC address of the host machine matches the MAC address presented to the network. This sensor validates the successful alteration of the MAC address. Additionally, researchers can analyze the outgoing IP address from the virtual machine to determine if it matches the IP address provided by the VPN and proxy server. Lastly, monitoring Tor network connections and analyzing the routing path can confirm if the Tor browser effectively connects to random nodes, strengthening second level anonymity. Challenges: The challenges in measuring second level anonymity stem from accurately tracking MAC address changes and verifying the integrity of the IP address assigned to the virtual machine. Additionally, assessing the randomness and effectiveness of Tor node hopping can be complex due to the decentralized nature of the Tor network.

4.3 Third Level Anonymity Measurement Sensor

Tracking the effectiveness of third level anonymity involves monitoring the routing path and connections made through the Tor network. Researchers can employ network traffic analysis techniques, such as packet sniffing, to examine the sequence of nodes and the randomness of the routing path. Additionally, statistical analysis can be applied to measure the distribution of connections to surface websites and onion websites, ensuring a balanced and randomized selection. Challenges: Measuring third level anonymity poses challenges due to the decentralized nature of the Tor network and the dynamic selection of nodes. The ability to accurately track and analyze the routing path while accounting for potential network disruptions or malicious nodes can be complex. Overall, measuring the different levels of anonymity presents challenges related to accurately tracking system details, validating the effectiveness of anonymization techniques, and accounting for the dynamic nature of network connections. Overcoming these challenges requires a combination of technical expertise, sophisticated measurement tools, and an understanding of the underlying technologies involved.

5. Result and Discussion

This study revealed various researchers' work on Freenet, I2P and ToR based dark web crime analysis and investigation procedures, and most of these studies are technically intensive, and non-technical analysts and law enforcement investigators might need support in adapting the approaches. It is possible to circumvent the network level anonymity by application layer interception such as dark website/blog/ marketplace messages in plain text. It is observed during this study that DNMs such as vice city, alphabay, Andromeda, agora, and other marketplaces advise using the PGP key for safe communication. Andrew C Dwyer et al.[19] Darknet Marketplaces (DNM) have adapted to using PGP encryption to overcome this. Figure 7 shows how the generation of encrypted PGP key pair (public & private key), malicious actors use similar PGP keys in their postings on the dark web to showcase the integrity and authenticity of the transaction to their customers.

DOI: https://doi.org/10.54216/JCIM.150113
Received: February 04, 2024 Revised: April 25, 2024 Accepted: July 27, 2024

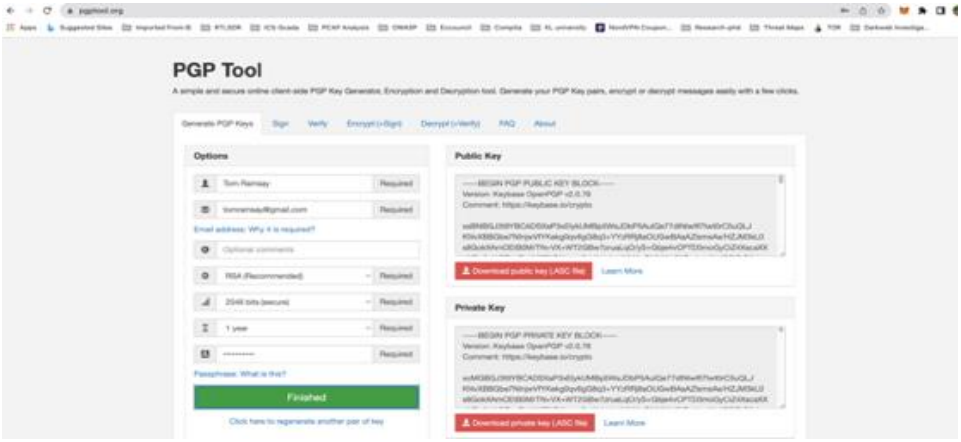


Figure 7. PGP key generated on PGP tool [20]

C. Cilleruelo, L. de-Marcos, J. Junquera-Sánchez et al. [21] has analyzed to identify the interconnection between I2P and TOR darknet content and service providers. This research was conducted on a dataset created by combing the I2P and TOR network-based websites and services. The dataset used for this research contains 49000 dark net services. This study proved the interconnection between dark networks. The dataset used in this study is helpful to law enforcement agencies (LEA) in investigating the existing domains and their active connections based on the attributes. LEAs can use methods used in this study to define the starting point of crawling and monitoring. Bergman, Jesper & Popov, Oliver. et al. [22] referred to 58 peer-reviewed articles through the systematic literature review (SLR) method and identified the prevalence and characteristics of dark web crawlers over anonymous communication networks (ACN) such as I2P and IPFS. From SLR knowledge, it is understood that on the surface and dark web, 'python' programming language with web scraping libraries such as selenium or scrapy are used widely. For this paper, a set of experiments were carried out and evolved a crawling and scraping model useful in digital investigations to obtain clues and evidence. Further, the advantages of the proposed method and the future scope of work were explained. Dark web crawling is challenging as it involves accessing content not indexed by traditional search engines and is often hidden behind layers of encryption and anonymity. From the above studies, it is understood that there is no one-size-fits-all algorithm for crawling the dark web, as it largely depends on the specific objectives and resources available. The sequential steps mentioned below would be appropriate to acquire the marketplace data from crawling and indexing.

- Active URLs list: Create a list of URLs that you want to crawl. URLs can be created manually or through automated tools.
- Target URL Identification: Application prompt for specific keywords to matching with active urls content.
- Crawler: There are various tools available for setting up a crawler on the dark web, including custom scripts and open-source software like Scrapy and Beautiful Soup. It is essential to configure the crawler to navigate the Tor network and follow any specific rules or protocols set by the target websites.
- Data storage: As the crawler navigates through the dark web, it will collect data from the target websites. Storing this data securely and appropriately is essential, as it may contain sensitive or confidential info.
- Analyze and filter data: Once the data has been collected, it can be analyzed and filtered to extract useful information. This can involve using natural language processing techniques, machine learning algorithms, and other tools to identify patterns and trends.
- Indexing Data: By using "AND" conditions, you can create filters that require all specified attributes to match. For instance, you could define a filter that searches for records where the "username," "PGP key," and "wallet address" all match specific values. This would narrow down the search results to entries that satisfy all the defined conditions. On the other hand, "OR" conditions allow you to create filters that require at least one of the specified attributes to match. This means that if any of the attributes, such as "username," "PGP key," or "wallet address,"

DOI: https://doi.org/10.54216/JCIM.150113
Received: February 04, 2024 Revised: April 25, 2024 Accepted: July 27, 2024

matches the given value, the record will be included in the search results. By combining "username" with "or" conditions, you can search for records where either the "user name," "PGP key," or "wallet address" matches the defined value. This provides flexibility in the search criteria, allowing you to retrieve results that meet any of the specified conditions.

- Interpret and visualize results: Finally, the results of the crawling and analysis can be interpreted and acted upon. This can involve making decisions based on the insights gained from the data or using it to inform further research or investigations.

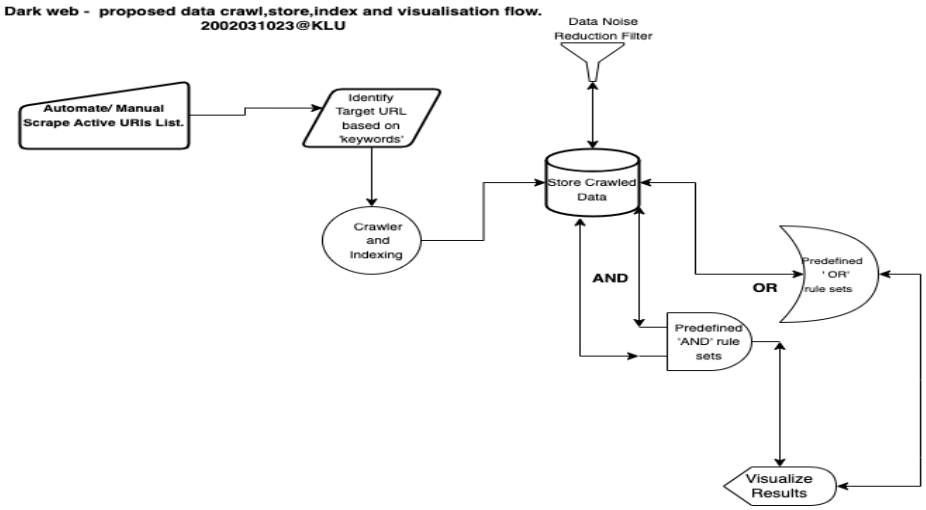


Figure 8. Proposed process of dark web crawling, storing, indexing and visualization

Meng, Yitong. [23] proposed a novel website response fingerprinting (WRFP) depended on the response time feature. This approach simultaneously monitors the hidden services webpages and their types and the hosted web servers. Experimental analysis with test results showed that the WRFP classifier effectively classified various web pages, subpages, and their types. Website fingerprinting is a valuable component that enhances the effectiveness of dark web investigation techniques. It involves analyzing the unique patterns and characteristics of websites to identify them, even when encryption is in use. This technique proves particularly useful for law enforcement agencies seeking to identify and track illegal activities on the dark web. By fingerprinting websites, agencies can create a database of known illegal marketplaces and monitor their activities for potential investigations. OSINT: Open Source Intelligence refers to gathering information from publicly available sources. OSINT on the dark web can monitor and collect information on dark web marketplaces, forums, and other platforms. Researchers and law enforcement agencies can analyze discussions, user profiles, and interactions to gain insights into illicit activities, such as drug trafficking or cybercrime, and identify potential threats or individuals involved. Threat Intelligence: Dark web marketplaces are known for facilitating the sale of illegal goods and services. Threat intelligence is vital in proactive investigation by identifying emerging threats, understanding criminal networks, and predicting potential cyberattacks. By analyzing data from dark web sources, investigation agencies can proactively identify vulnerabilities, track cybercriminals, and develop countermeasures.

- Operation Silk Road: The most infamous cases involving the dark web was the takedown of Silk Road, an online marketplace known for facilitating the sale of illegal drugs and other illicit goods. Law enforcement agencies utilized various techniques, including website fingerprinting, to identify and track the activities of Silk Road.
- Operation Bayonet: Alphabay was another prominent dark web marketplace that was shut down by authorities. It facilitated the buying and selling of drugs, stolen data, and other illegal items. In investigations like these, techniques such as website response fingerprinting could have played a role in monitoring and classifying the webpages and subpages within the marketplace.

DOI: https://doi.org/10.54216/JCIM.150113
Received: February 04, 2024 Revised: April 25, 2024 Accepted: July 27, 2024

- Darkode: Darkode was a notorious online forum where cybercriminals traded hacking tools, stolen data, and engaged in other illegal activities. Law enforcement agencies infiltrated and dismantled Darkode, relying on various techniques, including website fingerprinting, to gather intelligence on the forum's activities and identify its members.
- Operation Onymous: Europol played a key role in shutting down several major dark web marketplaces, including Hydra and Agora. Web fingerprinting was likely used to identify and track these platforms.

This study brought an understanding that investigation at the application level would be a feasible approach for investigation. We further focus on designing and developing a comprehensive methodology that can provide valuable intelligence to non-technical users and other teams working on eradicating dark web crimes. The proposed method should cover the below mentioned.

1. Integration of intelligence data collected from analysis on the marketplaces fingerprinting.
2. To design, develop and integrate the OSINT methodology for data collection, such as Information Extraction (IE), Attribute Extraction (AE) and correlation on various social media.
3. Integration network-level threat intelligence data acquired using proven methods or new approaches.
4. The proposed method should integrate the data collected from the hidden services directory (HSDir) to obtain actionable intelligence.

6. Conclusion and future work

To enhance the applicability of this research in real-life scenarios, it is important to provide specific instances or situations where various techniques, such as the examination of ToR nodes and website fingerprinting, have been effectively utilized in dark web investigations. This will highlight the practicality and effectiveness of these methods in addressing the challenges posed by the dark web. Having studied various research papers on the subject, it is evident that procedures like the collection and analysis of ToR nodes network traffic, website fingerprinting, compromised tor nodes for website traffic analysis, utilization of Hidden Service Directory (HSDir) nodes, PGP keys, and the use of honey traps, such as fake marketplaces, persona profiling, pattern analysis, and stylometry on actual marketplaces, have been employed in dark web investigations. However, to ensure a comprehensive analysis and investigation, future work should focus on developing a methodology that encompasses all the necessary steps. This includes gathering information about active targets (illegal traders) on marketplaces, identifying the attributes of traders and their postings on popular marketplaces. Additionally, the research should address challenges such as marketplace identification, acquisition and analysis of Tor datasets, crawling and indexing of dark web pages, deanonymization of dark web traffic, and gathering threat intelligence.

References

[1] Nukusheva, Aigul, et al. "Formation of a legislative framework in the field of combating cybercrime and strategic directions of its development." Security Journal 35.3 (2022): 893-912. https://doi.org/10.1057/s41284-021-00304-3

[2] János Besenyő, Attila Gulyas, The Effect of the Dark Web on the Security, Journal of Security and Sustainability Issues 11(2021), no. 1, 103-121, DOI 10.47459/jssi.2021.11.7, https://journals.lka.lt/journal/jssi/article/1510/info.

[3] Matthew Robert Shillito. (2019). Untangling the 'Dark Web': an emerging technological challenge for the criminal law, Information & Communications Technology Law. 28(2): 186-207. https://doi.org/10.1080/13600834.2019.1623449.

[4] Clarke, Ian & Sandberg, Oskar & Wiley, Brandon & Hong, Theodore. (2001). Freenet: A Distributed Anonymous Information Storage and Retrieval System. Lecture Notes in Computer Science. 2009. DOI 10.1007/3-540-44702-4_4.

[5] Roberto Magán-Carrión, Alberto Abellán-Galera, Gabriel Maciá-Fernández, Pedro García-Teodoro, Unveiling the I2P web structure: A connectivity analysis, Computer Networks, Volume 194, 2021, 108158, ISSN 1389-1286, https://doi.org/ 10.1016/j.comnet.2021.108158.

DOI: https://doi.org/10.54216/JCIM.150113
Received: February 04, 2024 Revised: April 25, 2024 Accepted: July 27, 2024

[6] L. Basyoni, N. Fetais, A. Erbad, A. Mohamed and M. Guizani, "Traffic Analysis Attacks on Tor: A Survey," *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, Doha, Qatar, 2020, pp. 183-188, doi: 10.1109/ICIoT48696.2020.9089497..

[7] Handalage, Upulie & Prasanga, Tereen. (2021). Dark Web, Its Impact on the Internet and the Society: A Review. 10.13140/RG.2.2.11964.36484.

[8] Martin, J., Munksgaard, R., Coomber, R., Demant, J. and Barratt, M. (2019) 'Selling drugs on dark web crypto markets: differentiated pathways, risks and rewards, British Journal of Criminology.

[9] Malathi S, Arockia Raj Y, Abhishek Kumar, V D Ashok Kumar, Ankit Kumar, Elangovan D, V D Ambeth Kumar*, Chitra B & a Abirami (2021) Prediction of cardiovascular disease using deep learning algorithms to prevent COVID 19, Journal of Experimental & Theoretical Artificial Intelligence, DOI: 10.1080/0952813X.2021.1966842.

[10] van der Bruggen, M., Blokland, A. (2021). Child Sexual Exploitation Communities on the Darkweb: How Organized Are They? In: Weulen Kranenbarg, M., Leukfeldt, R. (eds) Cybercrime in Context. Crime and Justice in Digital Society, vol I. Springer, Cham. https://doi.org/10.1007/978-3-030-60527-8_15.

[11] Kumar, V.D.A., Sharmila, S., Kumar, A. et al. (2023). A novel solution for finding postpartum haemorrhage using fuzzy neural techniques. Neural Comput & Applic. 35(33), 23683–23696

[12] T. Farah and L. Trajković, "Anonym: A tool for anonymization of the Internet traffic," 2013 IEEE International Conference on Cybernetics (CYBCO), Lausanne, Switzerland, 2013, pp. 261-266, doi: 10.1109/CYBConf.2013.6617434. https://ieeexplore.ieee.org/document/6617434

[13] Montieri, Antonio & Ciuonzo, Domenico & Bovenzi, Giampaolo & Persico, Valerio & Pescapè, Antonio. (2019). A Dive into the Dark Web: Hierarchical Traffic Classification of Anonymity Tools. PP. 10.1109/TNSE.2019.2901994.

[14] Kaur, Shubhdeep & Randhawa, Sukhchandan. (2020). Dark Web: A Web of Crimes. Wireless Personal Communications. Wireless Personal Communications: An International JournalVolume 112Issue 4Jun 2020 pp 2131–2158 https://doi.org/10.1007/s11277-020-07143-2.

[15] Ambeth Kumar, V.D. (2016). Human Life Protection In Trenches Using Gas Detection System. Journal of Biomedical Research. .27 (2), 475-484

[16] Openvpn is a VPN client solution, https://openvpn.net/, to connect with any VPN server. Free proxy/VPN server configuration downloaded from https//www.vpnbook.com/, last visited on 10th February 2023.

[17] Kumar, I., Kumar, A., Kumar, V.D.A. et al. (2022) Dense Tissue Pattern Characterization Using Deep Neural Network. Cogn Comput 14, 1728–1751.

[18] Wireshark is network packet analyzer software for network communication analysis, https://www.wireshark.org/download.html. They were last visited on 10th February 2023.

[19] Dwyer, Andrew & Hallett, Joseph & Peersman, Claudia & Edwards, Matthew & Davidson, Brittany & Rashid, Awais. (2022). How darknet market users learned to worry more and love PGP: Analysis of security advice on darknet marketplaces. https://doi.org/10.48550/arXiv.2203.08557

[20] Ambeth Kumar, V.D. Ramakrishnan,M. (2013). Temple and Maternity Ward Security using FPRS. Journal of Electrical Engineering & Technology, 8(3), 633-637.

[21] C. Cilleruelo, L. de-Marcos, J. Junquera-Sánchez and J. -J. Martínez-Herráiz, "Interconnection Between Darknets," in IEEE Internet Computing, vol. 25, no. 3, pp. 61-70, 1 May-June 2021, doi: 10.1109/MIC.2020.3037723. https://ieeexplore.ieee.org/document/9291465.

[22] S. Hemamalini ,V. D. Ambeth Kumar ,R. Venkatesan,S. Malathi. (2023). Relevance Mapping based CNN model with OSR-FCA Technique for Multi-label DR Classification. Journal of Fusion: Practice and Applications, 11 (2), 90-110.

[23] C. S. Manigandaa,V. D. Ambeth Kumar,G. Ragunath,R. Venkatesan,N. Senthil Kumar. (2023). De-Noising and Segmentation of Medical Images using Neutrophilic Sets. Journal of Fusion: Practice and Applications, 11 (2), 111-123.

[24] Ambeth Kumar, V.D. (2017). Automation of Image Categorization with Most Relevant Negatives. Pattern Recognition and Image Analysis, 27(3), 371–379.

DOI: https://doi.org/10.54216/JCIM.150113
Received: February 04, 2024 Revised: April 25, 2024 Accepted: July 27, 2024