Analyze Dark Web and Security Threats

Samar Ansh and Satwinder Singh

July 5, 2022

Analyze Dark Web and Security Threats

Samar Ansh[1][0000-1111-2222-3333] and Dr. Satwinder Singh[2][1111-2222-3333-4444]

[1] PG Student, Department of Computer Science and Technology, Central University of Punjab, Bathinda, India 151401
[2] Head of Department, Department of Computer Science and Technology, Central University of Punjab, Bathinda, India 151401
samar@cybercare.info

Abstract: Deepest area of data storage where data mining and data management is not possible without Tor (network) Policy is known as Dark web. Dark web is paradise for Government and Private sponsored cyber criminals. In another words Dark web is known as underworld of Internet used for sponsored and organized cyber crime. Tor network at the entry relay /guard user original source IP replaced with local IP (i.e. 10.0.2.15) by default by default and every user machine ID (IP) recognize as local IP (10.0.2.15). A single source IP allocated for each user without collision make user anomaly or invisible over Internet. Tor browser work similar VPN by default as function to hide original source IP, but advantage is Tor network's volunteer devices are use as tunnel to establish communication and offer freedom from surveillance of user activity. Tor browser offer circuit (IP Route) for user activity, where circuit allot available Tor IP at exit relay for user. Dark web use same IP at entry relay around the world, but at exit relay IP is different and available based on country. In Dark web network data transfer as encapsulation of packet / massage is placed after 3-layer of different encryption.

Keywords: Darkweb, Darknet, Deepweb, Tor, Project Onion, VPN, Cyber Threat

1. Introduction

Dark Web is a tool of the internet with the most challenging and unpredictable methods found and used by cybercriminal, terrorist, and sponsored spies to achieve their goal of illicit purposes. The cyber-crime that occurs within the Dark Web is similar to the crime of real world. Web services, on the other hand, are major challenges for identifying criminals because of their sheer scale, unpredictable environment, and anonymity. Dark Web is a tool of the internet with the most challenging and unpredictable methods found and used by cybercriminal, terrorist, and sponsored spies to achieve their goal of illicit purposes. The cyber-crime that occurs within the Dark Web is similar to the crime of real world. Web services, on the other hand, are major challenges for identifying criminals because of their sheer scale, unpredictable environment, and anonymity. The dark web's traffic classification is crucial to categorizing real-time application execution. Analyzing Dark-web traffic aids in the early identification and monitoring of malware assaults (detection) before to the onslaught and harmful actions following the breakout. But, now these days cyber-criminals are contemporary using VPNs to replace original information with forge for illicit purposes. Different applications (ie- audio-stream, video-stream, browsing, chat, email, VoIP-calls, etc) and modes are used to threaten and achieve illicit purposes. Dark-web (Tor application) traffic detects 13-15% (predict) in compared with chrome and Firefox application. In the name of privacy existing users are switching to Tor browser regularly, approximately (0.5%) every year from Firefox /Chrome /Opera and IE. This is very dangerous, switching users either threatening to others else threatened by someone over the dark web.

1.1. Internet

The term "Open Internet" or "Open Web" is frequently attributed to a collection of content housed on web servers and accessible via any internet browser (Edge, Chrome, Opera, and Firefox). Furthermore, because the site proprietors have not yet established robot.txt guidelines, this content may be indexed and appears on result by the Internet search engines i.e. "Google, Bing, Yahoo etc". So that search engines do not display any result on web pages.

1.2. Deep-web

The word "deep web" means "behind the open or surface web and a few dark webs," and it means "below the open or surface web and a few dark webs." Only certain websites and article links may be searched and indexed by search engines. Websites and article links are used to rank search results based on relevance, inbound links, and keywords. Because search engines do not cache URLs (search engines scan the internet by viewing one web page, then the links on that page, and then the links on subsequent sites), data is not returned to the user. Almost every time you search within a website, you're going to a deep internet page. The deep web performs to detect content by links, but help to protect user personal information and maintain privacy. The material is password locked behind a paywall and is proprietary, includes personally identifiable information (PII), or is controlled by law to restrict access (such as email accounts, tax records, payment systems, etc.)

1.3. Dark-web

A Large volume of data is not indexed by search engines (Google, Bing, Yahoo etc) and need special application or authorization to access these data. In another word- information which exists on Internet (World Wide Web), but need special application (Tor) and authorization to access is known as Dark web. Almost 96% un-indexed data stored with Dark web which need Tor application to access.

1.4. Tor

The Onion Router (Tor) is open source application used to access dark web networks, eventually become a U.S. option. Tor is a browser created by the Naval Research Laboratory (Onion Router). Tunnel based network (Tor volunteer computer) used to establish communication for "hides user's identity (IP) and to protect user's Internet privacy. Tor browser create tunnel by volunteer device (5000) to establish secure communication (circuit) for Dark web and user become anonymously on the Internet to monitor or censor.

1.5. I2P

A peer-to-peer (P2P) anonymous communication network for project of invisible Internet which protect user from any type of surveillance. I2P also process encrypted entrance as dark web, but network is decentralized in P2P and offer end-to-end encryption and routes using user machine network. This is similar to Tor project but difference is in I2P we can use this for instant messaging, file store and web hosting.

I2P use uni-directional tunnel to deliver packets, but separate inbound and outbound proxy routers that is called garlic routing. Packets divided into smaller which distributes balancing across different peers, so performance is more efficient.

2. Literature Review

Definition of the Dark web is criminal-threats, technical and legal challenges with unknown global network structures where techniques of detection methods, algorithms, and tools are used by criminals on the Internet. In the market of Dark web, transactions taking place must be followed to find out about criminals' strategic methods. Multilayer structure without indexed, split of the Dark Web makes it very difficult to detect criminals and crime. Dark Web Environment much unexpected as daily the old sites continue to disappear while new sites are emerging, strong digital evidence is needed to assemble forensic legal frameworks to ensure victory barriers to arrest and criminal prosecution. Content over the Internet can't be access with standard search engines. Dark Web information is usually not available on it most people and is deliberately hidden on the standard Internet browser, known as the "CLEARNET". Onion Router (abbreviated Tor) is one of the main ways to access the dark web "which connects your internet. Tracks by combining your online traffic with data from multiple servers around the world to make you invisible" [39].

There are lots of opportunities in Dark web (Tor) for malicious players to "exchange illegal assets anonymously". The Dark Web is becoming increasingly valuable, particularly in terms of unlawful activities and services. Safety measures should address these issues and take steps to eliminate them. Developing technologies that can encrypt and anonymity (such as Dark Web and its specialized application) have put lawmakers and policymakers on the challenge of successfully battling dangerous online players. [37]

Area of Dark Web is sharp and new subject in the domain of study and research, the research available on this topic is finger countable. In this section of the research paper, I review books about the Dark Web, black market user base online, online compliance, committed crime, and digital communities. Dark Web is a world wide network that users access pair to pair secure connection (Silk Road). Silk Road market place, accessible at "http://silkroad6ownowfk.onion", was only accessible via Tor application and always featured a sequence of root domains followed by "onions". Dark Web started with ARPANET; an internet founder founded by the Pentagon in 1969. As computer interactions begin. Mystery network of Dark web in increase the range of remote access to launch emerge ARPANET". [25]

These networks eventually become a U.S. option. "Naval Research Laboratory", which presents a software (web browser) called Tor ("Onion Router"). Tor application (network), "hides user IP and identity with help of tor node to tor node connection to protect American workers overseas and their opponents. [8], [12], [34]

However, in 2014, Onion Router (Tor) software was made available for public use and Tor sites become a haven of criminals and terrorists for drug trafficking, children pornography, and terror activity. “A. A. AlQahtani and E.-S.-M. El-Alfy”, (2015) In fact, there are concerns about using the dark web because most people have access dark web for illicit intention, some have legitimate goal to run anonymously their business online. So why use the dark web? Generally, Dark-web used for a variety of important reasons, such as (a) protection of privacy rights to the intended of citizenship and public scrutiny form. (b) Opponent protection from political retaliation. (c) Whistling and news leaks (private information). (d) Cybercrime (Fraud, cyber-attack, etc.) (f) Sales of illegal goods and contents (child pornography, private file, illegal or fraudulent software, etc.) over the Dark-web market. [9], [14], [30]

Tor's anonymity is based on simple and equitable mechanisms. The technique works by passing a request to the site through at least three relays, which are machines that are chosen at random to check people into the Tor network [21].

Tor is named after execution that an extra layer of encryption includes for each computer to the signal that only it can decipher (one-way hashing). The request then travels to a computer known as a 'exit relay' (exit to the Tor network), which is where the receiver believes it originated. [6]

This makes users anonymous because of “exit relay” is probably calls on behalf of masses of various machine (users) and randomizing algorithms verbdictate which exit relay is used to communicate with host. There is more than 7000 computer work as volunteers/nodes to relay the Tor communication. [3]

In effect, each request identity replaces with the originator and is hidden among the many layers of the Tor. One of the main ways to access the Deep Web is Tor browser “blocks your online tracks by combining your online traffic with data from multiple servers around the world to make you invisible”. [39]

As computer interactions begin to grow, “the number of private networks is beginning to emerge on the side of the ARPANET”. [25].

Anonymous access social networks - often referred to as Dark-web - are becoming popular with criminals smuggling drugs, illegal weapons trading, fake IDs, ID-theft, and child pornography. In another word, the Dark web is a completely illegal online marketplace that has been set up offering all kinds of illegal services driven by hidden and anonymous money laundering. To establish this anonymity, communication is done by using special anonymous networks such as I2p and Tor browsers. [15]

Anonymous access social networks - often referred to as Dark-web - are becoming popular with criminals smuggling drugs, illegal weapons trading, fake IDs, ID-theft, and child pornography. In another word, the Dark web is a completely illegal online marketplace that has been set up offering all kinds of illegal services driven by hidden and anonymous money laundering. To establish this anonymity, communication is done by using Onion anonymous networks such as I2p and Tor browsers [Invisible Internet Project].

In addition to this, legal and technological issues, as well as illicit usage, will be investigated further. The deep web is on the search for anonymity, which is an important aspect of the deep web's corrupt and uncontrolled character. Despite the literature review a variety of contexts, its elements are destructive, adverse, and highly susceptive to a variety of illegal activities, and they need to be policed or stopped, due to lack of control and weak regulations. [35]

An early challenge for the Dark-web internet was that it became difficult to discover hidden web portal. The Hidden Wiki delivered the first wave of customers in 2004 [“Dark net Markets Are Not beyond the Reach of Law 2016”].

This site offers a list of all presently operational Dark Web portal (website), as well as user comments and information on what can be accessed through each web-site. Another option is to use Tor protocol and specific search engines i.e. “Ahmia”, which crawls any hidden site, and Grams, which is particularly designed to locate hidden sites providing criminal goods and services such as counterfeit money, narcotics, and firearms. [6]

I In order to perform real-world transactions, dark web markets adopted Bitcoin, a pseudonymous money that is difficult to trace as Tor. Bitcoin has become the Darknet's first official and standard currency. Bitcoin and others are virtual currencies that are uninsured and variable, first introduced in 2009. They're kept in digital wallets that are encrypted. Designs of Bitcoins to be very tough to identify back to the individual that spent them. [7]

Log files maintain the record of each transaction, but only mention the wallet ID, not the identity of the buyer or seller (“Yellin, Aratari and Pagliery, n.d” that allows you to purchase bitcoins, a consumer ought to log into a bitcoin change, inclusive of the famous “Mt. Gox”, where buyers and sellers alternate exchange local currencies for bitcoins. [40]

Bitcoins may also be mined by volunteering a computer's CPU time to solve challenging arithmetic challenges. While bitcoin is the money of choice on the Dark Web today, “Zerocoin” is a cryptocurrency in the stage of development that will be even more anonymous than the transactions of Bitcoin. [3]

The Dark web changed into advanced in small steps, and it became now not designed to be what it's far these days. Tor’s developer on the NRL wanted an easy manner for army employees to communicate overseas. Hidden Wiki’s founders develop an index for common users to higher apprehend and browse the content of the Dark web. Bitcoin to was introduced to facilitate paying anonymously. Developers of this technology have been a vision of privacy, now not ill-intentioned, however, their intentions have not stopped unlawful interest from blossoming inside the shadows developed via the Tor network.

Most Tor users, after all, are simply looking for anonymity and may be utilizing Tor for genuine purposes. Only 1.5 percent of Tor users browse the Dark Web content, spite of the fact that it generates a lot of traffic. [38]

The problem is the inspection of the Dark Web and Tor are virtually only. It is not possible to make an application or tool that maintains users nameless while additionally tracking their interest to ensure that they may be now not getting access to unlawful websites. Tor's developer would love to think that the browser especially includes the site visitors of newshounds valiantly writing stories from countries where there is no laws protective loose speech, however that isn't always the case. For most visitors to hidden dark web websites, the usage of Tor application is for accessing and distributing photographs of infant abuse and place order for unlawful drugs. Toddlers abuse money owed for the most important set of dark web network traffic. "Dr. Gareth Owen and Nick Savage, researchers at the college of Portsmouth", carried out a 6 month examination that discover hidden service and facility of Tor. They came to the conclusion that over 80% of Tor traffic queries to hidden services over sites seen in the investigation were routed towards recognised child exploitation websites. [36]

They did renowned that these records may not be a wonderfully correct illustration, on account that authorities organizations regularly use computers so as to automatically get entry to web portal containing pix of toddler abuse as a part of their research. It's nearly hard to tell how much of the 80 percent is due to police action and how much is due to traffic generated by a human at a computer. Although half of the child abuse visitors detected have been police pastime, a lot of personal visitors remains on the dark web network targeting toddler abuse websites. The fact is images of child abuse are not isolated over dark web networks. n 2014, as part of an attempt to reduce the spread of child abuse content, the Internet Watch Foundation conducted a research. Only 51 (approximately 0.2 percent) of the 31,266 web links featuring content (photos, video) of child abuse were on the portal of the Dark web. (There is no method or technique available to search all Dark web sites, and As a result, the data may be skewed to appear like a decrease-than-accurate percent of web portals containing pictures of child exploit which might be at the dark web). [3]

This graph represents that, while the dark web facilitates cybercrime, it isn't the most convenient way to do it. The drug trade is the most generally connected topic with the Dark Web, and it is a key component of Dark Web markets. In reality, A huge percentage of dark web networks and hidden web portals is represented in "Dr. Gareth Owen's" examination of dark web surfing conduct. [36]

These websites are genuinely less complicated for enforcement officials to infiltrate because the officials are capable of higher hiding their identity once they cross undercover. Moreover, the medicine must be bodily brought, which leaves a window open for classic policing to understand the dealers. One of the most important and most notorious darkish internet marketplaces became Silk Avenue. It was developed in 2011 by "Ross William Ulbricht" who concealed beneath the alias of "Dread Pirate Roberts" (DPR) [6]

It's far expected that DPR obtained commissions over $13 million from permitting carriers to apply his Silk Road platform. In the October of 2013, the FBI close down Silk Road. In the investigation of FBI, they decided that more than $1.2 billion in income had come about concerning 150,000 clients and 4000 companies. [27]

Those astounding numbers display the size of illegal change at the dark web. Ulbricht turned into tried and sentenced to existence in jail in might also of 2015. [6]

Silk Street's closure in 2013 is no longer the end of dark web markets. In reality, several more sprang up to take Silk Avenue's place. These sites are used for more than just selling narcotics. They sell something that businesses need to put on the internet. Just after couples of weeks in the 2013 credit card credential breach at target, dark web markets have been selling stolen credit score cards at a charge of $20 to $one hundred in keeping with the card. [6]

There is a clear call for a black market online, so it is not a problem that it will use on its very own. Governments must work to approve policy so that they will deal with the difficulties placed forth with the aid of the Tor network in an attentive and willing reminder.

When FBI techniques were less than effective, the demise of Silk Road became a story. Although the operator, Ross Ulbricht, was apprehended, the darknet market for illicit items has flourished since shut down Silk Road in October 2013 by the FBI. The market was once located along Silk Road in the same region but is now more diverse. There is a Reddit Black Market, The list of reliable sites is significantly broader than the trusted list, and the market catalogue is being updated to help users know credibility status. [32]

After shut down of Silk Road, users flocked to an earlier unknown location called the "Sheep Marketplace". This web portal governed the market of the Dark Web until a trader exploited the risk and stole $ 6 million worth of bitcoins. [32]

Silk Road 2.0 was introduced by the previous management of the first Silk Road on 6 November 2013. This happened just one month after the closure of Silk's first road. Service of Silk Road 2.0 was a very short duration. It was hacked and stole $ 2.7

million from users' bitcoins by a merchant in February 2014. [32]

There was an internal route of the online marketplace in the Dark web that deals with illegal drugs and weapons etc. The Federal Bureau of Investigation (FBI), shut down the Silk Road web portal in 2013. But web portals like the "mythical Hydra", resurfaced like Silk Road 2.0 mid-month. FBI fails to detect new Silk Road immediately and it took a couple of years to track down its administrator and servers. [17]

Till the day Silk Road become the first choice for criminals and not going to end. Since May 2016 Dark net market was operational and is considered most powerful" Black Markets Are Not Over Legal Access 2016". [11]

Therefore, when the government demolished Silk Road, the operation was clearly not a complete success, as it did little to deter others from creating new Dark Web markets, nor did it prosecute retailers or customers for their transactions place. As "Eric Jardine" explore and direct, these crimes over the Dark Web can be downplayed, but other programs appear and simply replace their place. [33]

There are two distinct black online hobby circuits that can be labeled rational as totally useful but with commendable features: whistling and hacktivism. Whistleblowing is an important part of what keeps democracy under test, however, it can potentially disclose strategies of government and resources if they are not eliminated through formal channels. The black internet was used by whistleblowers including "Chelsea Manning, Julian Assange, and Edward Snowden" to reveal the secrets of the authorities. [27]

If the media had used the official, congressional approach to make their grievances, the issues would have been addressed without the distribution of publicly divided records. Hacktivism is another non-black and white matter while the wishes of other hacktivators may be controversial, their methods are always annoying and most impactable, but illegal. For example, in October 2011, an anonymous hacktivist organization crashed into a web hosting provider known as Freedom Web hosting, using a paid Denial of Service (DDoS) attack. They did this with the help of tracking signatures of web portals that abuse children. This Freedom website is also hosted on the server. Criminals also stole the details of 1,500 customers from the city of Lolita, a website that abuses children, and hacked them online. [6]

Removing of The intention to remove the child abuse content from web portal is honorable, However, vigilante justice has no place online since culprits cannot be held accountable. In 2001 remark that the cyber crime is not much different from real world crime and affection - it has just been killed in a new way with virtual weapon. [18]

If Internet Protocol (IP) addresses cannot be identified, anonymity on the internet is guaranteed. Client software of Tor delivers Internet traffic via a global hidden network of tunnel to hides user information and avoids any monitoring activities. This feature of Dark web offer more suitable for cyber criminals, who are always trying to hide their tracks. [2],[23]

A well-known preferred channel for governments to exchange documents privately is Dark Web, so that journalists can pass through several regional surveys and opposition to avoid controlling authoritarian regimes. [20].

The Onion Route is a way of communicating technique anonymously with a computer network. Messages are encrypted and routed via numerous network nodes known as onion routers. Each onion root, like an onion peel, peels the encryption layer to disclose the route instructions and transmits a message (signal) to pre-sequenced next route, where same steps repeated till deliver to destination. This process prevents central nodes from knowing the source, location, and content of a message [25].

It's reasonable to believe that specialised sites facilitate the exchange of both mode physical and private information i.e. login credential to access on paid web pornographic sites, as well as PayPal credentials. [26]

The Assassination Market's website offer betting that predicts when a group can bet on the death of a particular person, and collect payments if the date is "correctly" estimated.[1]

This encourages murder since the murderer may benefit by placing an exact wager at the time of the suspect's death because he or she knows when the event will take place. Because the charge is to know the date rather than to commit an act of murder, it is very difficult to give a criminal charge for the murder. White Wolves and C'thuthlu are two well known website to hire assassin. White Wolves and C'thuthlu are two well known website to hire assassin. [21], [24]

Various types of firearms are offered by 'Euroarms' web portal that mostly delivers to customer's doors around in the Europe. The characters of these weapons are sold separately and the website should be recognized separately on the dark web. [4], [5], [22]

On the dark web, paedophilia, often known as CP (child pornography), is freely available. A specific law permits pornography on the website. The dark web offers a variety of web portals and forums for clients who wish to engage in pedophilia. [21]

Secure anonymity is critical for these individuals and the organizations that support them. It can save a real life. Although Tor Policies should be tailored to specific circumstances. (Marx 1999).

Like any other invented technology, anonymity may be used for both good and harm. The vast majority of individuals do not want their online persona to be linked to their offline persona. They may fear political or economic retaliation, persecution, or even death. Instead of using their own identities to defend themselves, many individuals prefer to talk in fictitious or anonymous words. The following are a few examples of when people use Tor's online anonymity. [25], [29]

Monitoring the Dark Web will continue to be difficult due to its structure and nature. Efforts to address it should concentrate on the topics listed below. [13], [19]

Europol authorities have raised worry that bitcoins are increasingly being used in illicit operations. Since its beginning in 2014, DDoS "4" Bitcoin (DD4BC), a cybercrime squad named after the Distributed Denial of Service (DDoS) assault, has hit more than 140 firms. This inspires other gangs, which leads to cyber extortion. According to Europol, the DD4BC organisation began threatening victims through email with DDoS assaults until a bitcoin ransom could not be paid. The emergence of Bitcoin coincided with the growth of cyber-terrorists on the dark web. [28]

crawlers / spiders are designed to collect dark web content. Spiders get access to password- protected websites and download files at random. Spiders are trained to download all of the site's “pdf, Word, HTML, PHP, CGI, links, images, ASP files, videos, and audios”. The forum capture programme about 15 and their format are determined by the spidering forum utility. A comprehensive forum comprises topics, authors, threads, postings, and timeline tags that allow members' interactions to be rebuilt. Occasional spidering forum and growing updates are perfected on the basis of research requirements. Using computer-assisted language approaches, the forum's content is gathered and analysed in English, Arabic, French, Chinese, and Spanish. Specific multimedia collection techniques and files from web sites and web sites and spidering 8 Samar Ansh: Analyze Dark Web and Security Threats have been developed. [31]

2.1. VPN / Non-VPN

Research team of “Gerard Drapper Gil, Arash Habibi Lashkari, Mohammad Mamun and Ali A. Ghorbani”- analyzed and generated dataset for 8 different contents (Audio-File streaming, Web browsing, Internet chatting, Electronic mail, VoIP, Video-File streaming, File transfer, and peer-to-peer) used over Internet connection of VPN /non-VPN traffic in 2016 "Detect and Analyze VPN Traffic and Time-Features".

2.2. Tor/ Dark-web

“Arash Habibi Lashkari, Gurdip Kaur, and Abir Rahali”. Analyzed and generated dataset with different applications used illicit purpose ie- audio-stream, video-stream, web-chat, browser, email, voip calls, transfer, p2p communication etc. over Tor browser traffic in 2020 "Approach to analyze and characterize the Dark Web Traffic using technique Deep Image Learning".

3. Problem Statement

As name Dark web / Deep web has a mysterious secret over the internet. There is finger countable research paper available. The technique used for committing a crime by criminals due to the Tor browser hiding the location and identity of the user. This article will help to detect and characterize applications used over the Tor browser and VPN application. In 2016 research was conducted by CIC for detection and characterize of applications used over Tor browser and VPN. In the past 2 years, activity over Dark Web has increased approximately 300%, and recently in 2020 more than 22 billion new records were added to the dark network. In the Dark Web forum, an estimated 10% of users post resources as cybercrime sellers and an estimated 90% of posts from buyers looking to contract.

This research work proposes for novel techniques to detect, analyze and characterize VPN and Tor applications together as the real representative of dark-web and VPN traffic by amalgamating out three public datasets namely- ISCXTor2016, ISCXVPN2016, and CICDarknet2020 to generate a complete dark-web. Tor and VPN traffic are covered in this dataset, respectively with details of the user application.

Another side we don’t know active users over the Dark web and mode of operands (Application) for illicit activity. A major part of this paper is to collect information about the following: -

a. Traffic Load on Dark web (Tor browser),
b. Traffic redirect with VPN service and
c. Applications used with Non-Tor browser and Non-VPN.

Article covers wide area of Dark web activity and security threats for a security provider.

4. Objectives

Dark-web is designed and developed with the idea of invisibility over the internet to spy on the user of the Internet by Gov. of USA is still a mystery for other than USA Gov.'s law enforcement agencies where cyber criminals perform the illicit activity. The objectives of this dissertation are:

a. To collect the peer reviews of a research paper from various security platforms.
b. To extract the features from a peer-review dataset of scientific papers.
c. To analyze and signalize the Dark-web traffic and activity using Deep Image Learning.

5. Methodology

Dataset

Dataset name "CICDarknet2020" is combination of ISCXVPN2016 and ISCXTor2017 and available on public domain for Dark web traffic from "Canadian Institute for Cyber security". The dataset consists (sample) of 141530 instances where 46782 instances are marked as VPN, 1392 instances marked as TOR, and the rest of instances are labeled as Non-TOR / Non-VPN (Open Internet). The dataset offers 85 features including source and destination IP with port numbers that recognize the flow of packets with a volume of data. "CICDarknet2020" dataset has been processed to get mature data in the desired format, then it is divided into two parts testing (20%) and training (80%). The data was gathered and prepared for analysis and characterization to detect service (communication) based on source and destination IP. Those 86 attributes of the prepared dataset can be divided into five groups:

I. Attributes based on TOR.
II. Attributes based on VPN.
III. Attributes based on NON-VPN /NON-Tor.
IV. Attributes based on Source and destination IP.
V. Attributes based on the protocol.

The first group (Tor) is on non-routable anonymous IP-based communication with P2P communication for users. Tor IP can be detected as IP of source or destination. The fifth group performs to detect and analyze the type of application running on port no. with protocol over communication. The instances of Tor (dark web) are the lowest in the dataset. The group of NON-VPN /NON-Tor (Open Internet) has a larger no. of instances in the dataset. The experiments have been carried out using Jupyter Notebook installed on Microsoft Windows Operating System 11.

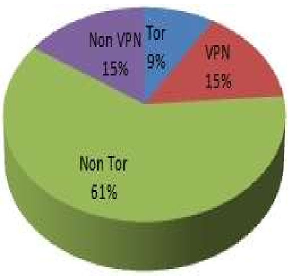
6. Machine Learning Models

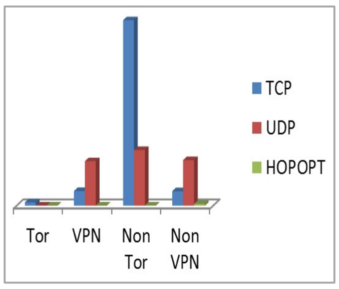
Machine learning models are classified as either supervised or unsupervised. The model is separated into two categories if it is supervised: regression and classification. We'll go with various models (Logistic Regression, Random Forest, Gradient Boosting, Ada Boosts, K-Nearest Neighbors, Decision Tree) of machines to signify and categorize in the sections. The detection accuracy of these models will be evaluated and the best among them will be considered.

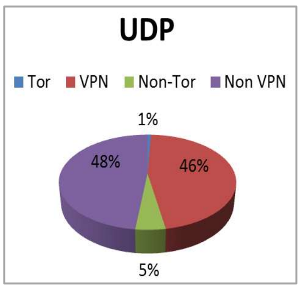
7. Results and Discussion

This portion outline the findings of the Decision Tree model creation utilising a variety of machine-learning methodologies to fulfill desired solution and found model of BAG-DT is best-optimized result.

Dataset "CICDarknet2020" processed and found actual user of Tor (Dark web) is 9% of sample 141530 users where Non-Tor user on top and runner with 61%.







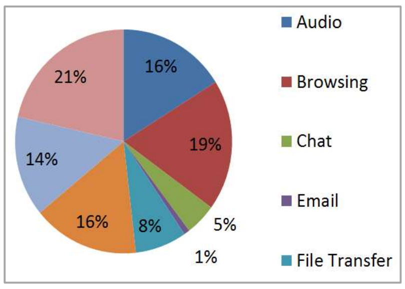


Figure 1. all Internet Traffic. Figure 2. Applied protocol. Figure 3. UDP protocol applied on Internet. Figure 4. Dark web users activity.

Each type of communication used TCP to establish communication and deliver data safely. The% of TCP protocol is higher compares to UDP, but in VPN and Non-Vpn UDP protocol is widely used to transfer (Download) data as shown in graph image (b) and (c).

Tor (Dark web) is known for illicit market (Illegal Software and Data Selling). Analyzing and compare uses of file transfer protocol (UDP) based of available Dataset "CICDarknet2020" and surprise Non-VPN user proceed with 48% and Dark web used UPD just 1% of total File Transfer activity as shown in Graph (c).

Analyze data display Tor browser user access TCP, UDP and HOPOPT protocol as shown in below Table 1.

Table 1. *protocol applied over the Dark web.*

TCP	UDP	HOPOPT
93%	5%	2%

Analyzed dataset of Dark web to understand and explain activity of over Tor network. There are 8th different type of activity found with multiple available application can run over Dark web. Dark web user access VOIP service (21%) mostly and that's why VOIP is highest traffic and Internet surfing in on 2nd position over Tor network. Even Onion (Tor) network surveillance free, but traffic of email is lowest (1%) within Tor. Lowest traffic of email and highest traffic of VOIP show that Tor network not trustable for their users and not going to share storable data (Email 1% and chat 5%) over Dark web network.

Based on Tor activity report, we pretend cyber criminal's are browsing illegal content and avoiding storage communication to avoid any trail and footprint.

8. Conclusions

Precisely, this paper includes a broad definition of Dark web criminal threats, unknown network architecture, detecting methods, algorithms, and tools provide technical and forensic hurdles and techniques used by finding crime and criminals on the Dark Web. Cyber criminals have instant intelligence mandatory ways to see your-self within Dark Web, but their no. of active user is 1.4% only. As a result, the stakes are tremendous. Lastly law and security an international boarder agency is one of the biggest obstacles function. Hidden web size requires more effective ways to reduce potential threats Dark Web. Black market and transactions taking place Tracks must be tracked in order to discover offenders using modern procedures. Unindexed, separated from multilayer the formation of the Dark Web makes it difficult to see charges. Dark Web ecosystem being much unexpected as daily the old sites keep disappearing as new sites emerge, strong digital evidence is needed gathering in law enforcement agencies to ensure victory obstacles to arresting and prosecuting criminals.

This study proposes, models, implements, evaluates, and reports on an efficient autonomous Dark web traffic detection system (DTDS). The proposed system characterizes the performance of supervised machine-learning techniques, including decision tree ensembles models were evaluated on a modern and inclusive dataset (CIC-Darknet-2020) involving a large number of captured cyber attacks and available services provided by Dark web organized into four classes (VPN, TOR, Non-VPN, Non-TOR).

References

1. Abbasi and H. Chen, ``Affect intensity analysis of dark Web forums,'' presented at the IEEE Intell. Secur. Informat., May 2007.
2. M. Akhoondi, C. Yu, and H. V. Madhyastha, ``LASTor: A low-latency AS-aware tor client,'' presented at the IEEE Symp. Secur. Privacy, May 2012.
3. Clemmitt, M. 2016. "The Dark Web." Accessed August 30, 2016. http://library.cqpress.com/ cqresearcher/document.php?id=cqresrre2016011500.
4. A. Alipoaie and P. Shortis, ``From dealer to doorstep-Howdrugs are sold on the dark net,'' GDPO Situation Anal., Swansea Univ., Global Drugs Policy Observatory, Swansea, U.K., Tech. Rep., 2015.
5. Darknet Markets Are Not beyond the Reach of Law. 2016. Accessed August 30, 2016. https:// darkwebnews.com/darknet-markets/darknet-not-beyond-law/.
6. Finklea, K. 2015. "Dark Web." Accessed August 30, 2016. https://www.fas.org/sgp/crs/misc/R44101. pdf.
7. Biddle P, England P, Peinado M., Willman B. (2003) The Darknet and the Future of content Protection. In: Feigenbaum J. (eds) Digital Rights Management. DRM 2002. Lecture Notes in Computer Science, vol 2696.Springer, Berlin, Heidelberg.
8. https://anshchoudhary.wordpress.com/2017/03/14/therise-and-challenge-of-dark-net-drug-markets/.
9. https://www.comparitech.com/blog/vpn-privacy/how to access the deep web-and darknet, Paul Bischoff, 2018.
10. https://www.technadu.com/dark-web- history/52017/.
11. K. Jaishankar. (2016). International Journal of Cyber Criminology (IJCC), ISSN: 0973-5089 January – June 2016. Vol. 10 (1): 40–61. DOI: 10.5281/zenodo.58521.
12. Rathod, "Darknet Forensics", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 6, Issue 4, July - August 2017, pp. 077079, ISSN 2278-6856.
13. Senker. (2016), Cybercrime & the Dark Net: Revealing the hidden underworld of the internet, London: Arcturus Publishing, ISBN 9781784285555.
14. A. A. AlQahtani and E.-S.-M. El-Alfy, "Anonymous connections based on onion routing: review and a visualization tool,'' ProcediaComput. Sci., vol. 52, pp. 121128, Jan 2015.
15. Arash, Habibi, Lashkari, Gurdip, Kaur, Abir, And Rahali https://ieeexplore.ieee.org/document/9251210
16. Darknet Traffic Big-Data Analysis and Network Management to Real-Time Automating the Malicious Intent Detection Process by a Weight Agnostic Neural Networks Framework.
17. Mac, Ryan. 2014. "Feds Shutter Illegal Drug Marketplace Silk Road 2.0, Arrest 26-Year-Old San Francisco Programmer." Forbes, November 6.
18. Grabosky, Peter. 2001. "Virtual Criminality: Old Wine in New Bottles?" Social & Legal Studies 10: 243–49.
19. Ciancaglini, Vincenzo, Marco Balduzzi, Max Goncharov and Robert McArdle. 2013. "Deepweb and Cybercrime: It's Not All About TOR." Trend Micro Research Paper. October.
20. Gehl, Robert W. 2014. "Power/Freedom on the Dark Web: A Digital Ethnography of the Dark Web Social Network." New Media & Society, October 15. http://nms.sagepub.com/content/early/2014/ 10/16/1461444814554900.full#ref-38.
21. Greenberg, Andy. 2013. "Meet the 'Assassination Market' Creator Who's Crowd funding Murder with Bitcoins." Forbes, November 18.
22. Love, Dylan. 2013. "There's a Secret Internet for Drug Dealers, Assassins, and Pedophiles." Business Insider, March 6.
23. Paganini, Pierluigi. 2012. "The Good and the Bad of the Deep Web." Security Affairs, September 17.
24. Pocock, Zane. 2014. "How to Navigate the Deep Web." Critic, Issue 03, March 19.
25. Tor Project. 2014a. "Tor: Overview." www.torproject.org/about/overview.html.en
26. Westin, Ken. 2014. "Stolen Credit Cards and the Black Market: How the Deep Web Underground Economy Works." LinkedIn, August 22.
27. Rudesill, D. S., Caverlee, J., & Sui, D. (2015). The deep web and the darknet.
28. "PGP Encryption." [Online]. Retrieved December 16, 2019, form https://www.techn adu.com/pgpencryption-dark-web/57005 /.
29. Journal of Computer and Communications, 2019, 7, 30-43, ISSN Online: 2327-5227.
30. Chen, H. and Yang, C. (2008). Intelligence and security informatics. Berlin: Springer.
31. Willard, P. and Bellamy, C. (2012). Principles of methodology. London: SAGE.
32. Swearingen, J. 2014. "A Year after Death of Silk Road, Darknet Markets are Booming." Accessed August 30, 2016. https://finance.yahoo.com/news/death-silk-road-darknet-markets-142500702.html.
33. Jardine, Eric. 2015. "The Dark Web Dilemma: Tor, Anonymity and Online Policing." Accessed December 14, 2016. https://www.cigionline.org/sites/default/files/no.21.pdf.
34. Godawatte, K., Raza, M., Murtaza, M., & Saeed, A. (2019, Dec 5-7). Dark Web along with The Dark Web marketing and surveillance [Paper presentation]. PDCAT 2019: Gold Coast, Australia.
35. González, P. (2013). Fingerprinting Tor. Information Management & Computer Security, 73-90. Guitton, C. (2013).
36. Owen, Gareth and Nick Savage. 2015. "The Tor Dark Net." Accessed December 13, 2016. https://www. ourinternet.org/sites/default/files/publications/no20_0.pdf.

37. Beshiri, A., & Susuri, A. (2019). Dark Web and its impact in online anonymity and privacy: A critical analysis and review. Journal of Computer and Communications. 7. 30-43. https://doi.org/10.4236/jcc.2019.73004.
38. Ward, M. 2014. "Tor's Most Visited Hidden Sites Host Child Abuse Images." Accessed August 30, 2016. http://www.bbc.com/news/technology-30637010.
39. Hodson, 2014. "Invisible Internet".
40. Yellin, Pagliery and Aratari. "Darknet Market