CYBERCRIMES IN THE DARKNET AND THEIR DETECTIONS: A COMPREHENSIVE ANALYSIS AND FUTURE DIRECTIONS

HumayraBinte Ali[1], Mohamed Abdelrazek[2], Shamsul Huda[3], MdMamunur Rashid[4], MdToufiqur Rahman[5], Amani Ibrahim[6]

[1]Australian Technical and Management College (ATMC), Melbourne
[2]Institute for AAAI, Deakin University, Burwood VIC 3125, Australia
[3]School of Information Technology, Australia
[4]School of Engineering and Technology CQUniversity, Australia
[5]Khulna university of Engineering and Technology (KUET), Bangladesh
[6]WIPRO, Australia

ABSTRACT

Although the Dark web was originally used for maintaining privacy-sensitive communication for business or intelligence services for defence, government and business organizations, fighting against censorship and blocked content, later, the advantage of technologies behind the Dark web were abused by criminals to conduct crimes which involve drug dealing to the contract of assassinations in a widespread manner. Since the communication remains secure and untraceable, criminals can easily use dark web service via The Onion Router (TOR), can hide their illegal motives and can conceal their criminal activities. This makes it very difficult to monitor and detect cybercrimes over the dark web. With the evolution of machine learning, natural language processing techniques, computational big data applications and hardware, there is a growing interest in exploiting dark web data to monitor and detect criminal activities. Due to the anonymity provided by the Dark Web, the rapid disappearance and the change of the uniform resource locators (URLs) of the resources, it is not as easy to crawl the Drak web and get the data as the usual surface web which limits the researchers and law enforcement agencies to analyse the data. Therefore, there is an urgent need to study the technology behind the Dark web, its widespread abuse, its impact on society and the existing systems, to identify the sources of drug deal or terrorism activities. In this research, we analysed the predominant darker sides of the world wide web (WWW), their volumes, their contents and their ratios. We have performed the analysis of the larger malicious or hidden activities that occupy the major portions of the Dark net; tools and techniques used to identify cybercrimes which happen inside the dark web. We applied a systematic literature review (SLR) approach on the resources where the actual dark net data have been used for research purposes in several areas. From this SLR, we identified the approaches (tools and algorithms) which have been applied to analyse the Dark net data, the key gaps as well as the key contributions of the existing works in the literature. In our study, we find the main challenges to crawl the dark web and collect forum data are: scalability of crawler, content selection trade off, and social obligation for TOR crawler and the limitations of techniques used in automatic sentiment analysis to understand criminals' forums and thereby monitor the forums. From the comprehensive analysis of existing tools, our study summarizes the most tools. However the forum topics rapidly change as their sources changes; criminals inject noises to obfuscate the forum's main topic and thus remain undetectable. Therefore supervised techniques fail to address the above challenges. Semi-supervised techniques would be an interesting research direction.

DOI: 10.5121/ijwest.2024.15202

KEYWORDS

Darknet, Cybercrimes, Data analysis, Deep Web, Surface web, Darknet Data Analytic, The Onion Router, Machine Learning

1. INTRODUCTION

The Internet which we surf every day, also known as the “Surface Web”, is actually a fraction of the total interconnected online world. On the other hand, the online data/sites which are un-indexed and unreachable to end users using normal widely-used web browsers or using ordinary search engines (Google, Yahoo, Bing) make up the “Deep Web” (1),(2) as mentioned in Figure 1. Much of the deep web content are used with malicious intent which is the hidden part of the Deep Web and is known as a Dark Web Figure 1. The bottom and deepest layer is only accessible by using specialized browsers (like the Tor browser), which is known as the Dark Net. The dark net is used by the criminals for most criminal activities including drug dealing, financial data breaches, assassination contracts, pornography business, dealing with human body parts, human trafficking, illegal arms transactions, hacking identity or credential information and conducting terrorism activities.

One of the major cybercrimes in the Dark net is selling confidential credential information. Cyber criminals steal sensitive and confidential information about people or business organizations using the different techniques and tools such as footprinting, scanning, enumeration and cyberattack by using different vulnerabilities of information systems or computer network that business organizations use. Then, attackers demand payment from the business organizations threatening the organizations for unauthorised releases of stolen confidential information or they sell that information to other people, which is also known as a data breach. A recent research conducted by Sam Smith from Juniper research shows that global cybercrime cost amounts to $2 trillion in 2019 which is a 300% increase from $500 billion in 2015 (3). This research indicated that 40% of the above cost is due to criminal activities of the Dark net which were classified as ‘Data breaches’.

Researchers and individuals in law enforcement are taking deeper investigation to properly understand the structure, content, malicious trades happening on the dark web, etc. Different techniques and tools are being used by researchers and law enforcement agencies. Mapping hidden services directory, monitoring customer data and social media sites, graph based Social Network Analysis (SNA) tool (4), (5), semantic analysis (6), (7), (8), keyword based Web graph analytic tools (9), Geo-Location Analayis of the Dark net products (10), Dark Net Volume Measurement and Content Analysis (11),(12), (13), (14), (15), (16) are some popular existing techniques.

Although there are a significant amount of existing works in the literature, the Dark net is the one of the biggest challenges that our digital society is facing. The study and analysis of the dark web is becoming a prominent research area. Due to the TOR technology and its anonymity, rapid disappearance and the change of the uniform resource locators (URLs) of the resources and links, monitoring the dark web is becoming more challenging for law enforcement agencies day by day. This study investigates existing techniques, identifies their limitations and offers a new research direction to address the aforementioned challenges in order to reduce the cybercrime and the illegal activities in the dark net.

2. PAPER ORGANIZATION

This paper is organized as explained below. Section 2 presents the literature review process. Section 3 reports the predominant illicit activities which are assumed to be the top controlling dark net market places. Section 4 presents the review of the Dark Net data analysis; here, the review process has been done based on the resources where the actual Dark Net data has been used, Section 5 presents the dark net challenges followed by the research direction and the conclusions of this study.

3. THE LITERATURE REVIEW PROCESS

For conducting our review process on Dark Net data analysis, we have used a combination of both the Systematic Literature Review (SLR) (17) and the Hermeneutic Literature Review (HLR) Approach (18). The main steps of our approach are: (i) identify research questions; (ii) select an appropriate search procedures; (iii) select inclusion and exclusion criteria for selecting articles; (iv) final articles selection; and (v) extraction of data and synthesis of data. Our own analysis to solve the open ended process that matches with the research work (18). The solution of the open-ended process can be described by applying the Hermeneutic Literature Review Approach (19) (20) (21). This approach for studying literature reviews is an inherently interpretive processes in which the reviewer engages in an ever deepening and through understanding of the relevant literature (22). Rather than relying on flawless understanding, the hermeneutic process keeps an emphasis on persistent process of developing apprehension and replacing of earlier theories by better theories and advanced paradigm [(23). Essentially, we have started the search procedure with the SLR approach and to find our focused target, we tuned up the keywords (papers) applying the circles of HLR.

3.1. Research Questions

The main focus of this study is to point out the most controlling Dark Net market places (RQ1) and the process of analysing those dark net data (RQ2) for achieving several targets. For example, to help law enforcement agencies in terms of finding the Dark Net criminals. The research questions for this SLR are shown in Table 1, along with the corresponding motivations.

3.2. Search Procedure

To get the most pertinent peer-reviewed papers, we devised a search method. The following is a description of our search process.

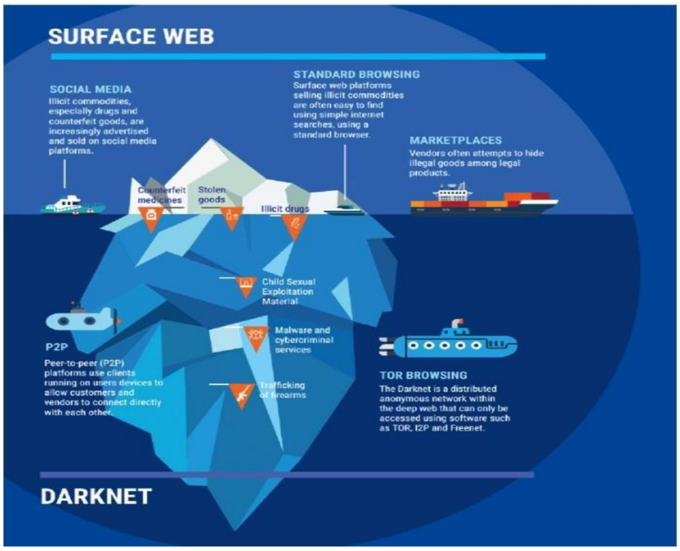

SURFACE WEB
SOCIAL MEDIA
Illicit commodities, especially drugs and counterfeit goods, are increasingly advertised and sold on social media platforms.
STANDARD BROWSING
Surface web platforms selling illicit commodities are often easy to find using simple internet searches, using a standard browser.
MARKETPLACES
Vendors often attempts to hide illegal goods among legal products.
Counterfeit medicines
Stolen goods
Illicit drugs
Child Sexual Exploitation Material
Malware and cybercriminal services
Trafficking of firearms
P2P
Peer-to-peer (P2P) platforms use clients running on users devices to allow customers and vendors to connect directly with each other.
TOR BROWSING
The Darknet is a distributed anonymous network within the deep web that can only be accessed using software such as TOR, I2P and Freenet.
DARKNET

Figure 1: The Layers of the Internet((2))

3.2.1. Search method

We searched six digital libraries—ACM Digital Library, IEEE Xplore, Scopus, ScienceDirect, SpringerLink, and Wiley Online Library—automatically to find possibly pertinent papers. With the search keywords presented in the following subSection, these libraries were looked up. We employed snowballing (24) as an addition to the automatic search to ensure the identification and selection of as many pertinent documents as possible.

3.2.2. Search terms

We used these keywords: dark net data analysis, dark web, illicit activities on the dark net, dark net market places, dark net challenges, thematic analysis of dark web data, deep web data. We excluded these keywords: TOR hidden services, Deanonymizing TOR, Dark Web crawling. After the search, we only selected those papers where the actual dark net data has been used.

3.2.3. Data sources

The digital libraries that were searched are displayed in Table 2.

TABLE 1: Research Quecusfostions

	Motivation
RQ1: What are the most controlling Dark net market places.	To find the most dominant market places on dark web.
RQ2: Data analysis Process.	To analyse the research works to find the data analysis process, their main contribution and the research gaps where the actual dark net data has been downloaded and used

Research Question

TABLE 2: Database Sources

IEEE Xplore	http://ieeexplore.ieee.org
Scopus	https://www.scopus.com/
ScienceDirect	http://www.sciencedirect.com
ACM	http://portal.acm.org
SpringerLink	https://link.springer.com
Wiley	http://onlinelibrary.wiley.com

Source URL

4. PAPERS SELECTION

For this section, we used the Hermeneutic approach. We are describing the main two circles of this search approach. After applying the keywords on the digital libraries, we found a good amount of papers. Initially, we discarded the papers where the main contributions are the technical analysis of dark web crawling and the acquisition approaches. Then, we came across a small number of papers. Among those, we figured out the top-ranked malicious activities on dark web sites. We found that “Human Trafficking”, “Terrorist Activities” and “Drug and Silk Road” are the most common crimes that people are dealing with by using the anonymity of the dark web sites. Then, our objective was to find out the works where actual dark net data has been used. So the number of papers reduced significantly. Then, we chose a small group of papers where the actual dark net data had been downloaded and analysed to gain some directions on the highest level of dark net crimes. The hermeneutic approach helped us to summarise the number of these papers on which we did our reviews to find the data analysis approach, their main contributions and the key research gaps on their works, etc. The following sections encompass the overview and the circles of HLR approach.

4.1. Hermeneutic Approach

When applying the SLR approach to the step of ”search procedure”, we faced a problem of reviewing literature of relevance to a specific problem and it seems to be a continuing open-ended process. As described before, we solved this problem by applying the Hermeneutic Literature Review Approach as presented in Figure 2. The basic hermeneutic process encompasses two major hermeneutic circles, which are: (i) the search and acquisition circle and (ii) the wider analysis and interpretation circle that are collaboratively convoluted. We found a huge number of paper on Dark net. It was not easy to separate the papers which are of our interest. Our interest is not the technical side of the dark net sites rather our interest is to find the works where the actual data has been used in the major dark net market places and to analyse their contribution areas. The hermeneutic process helps us to fine-tune our resources.

4.2. The Inner Circle

The circle in Figure 2 is a process of properly searching the database for the literature which is also hermeneutic circle (22). To produce high quality reviews, understanding searches and the identification of relevant literature is a prerequisite. According to the hermeneutic approach, Figure 2 shows the process of search and acquisition. The usual ways to go beyond the database are Snowballing and Citation. As seen in Figure 3, ”snowballing” is the practise of finding more publications by leveraging a paper’s citations or reference list.

5. MALICIOUS ACTIVITIES ON DARK WEB

We applied the hermeneutic approach on our objectives, which are: (i) To find the major profitable illicit activities on dark net and (ii) the research work on those selected activities where the actual dark net data has been investigated from 2004 to 2019.

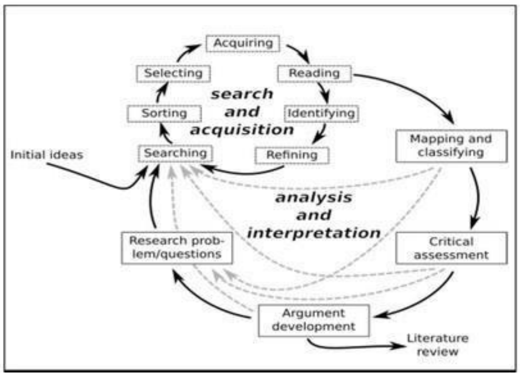


Figure 2: A hermeneutic framework with two main hermeneutic circles for the literature review procedure (22).

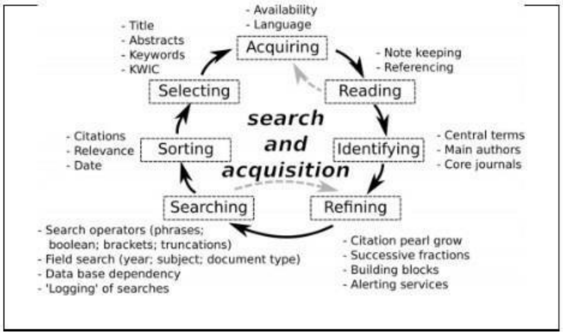


Figure 3: An overview of the many instruments and procedures connected to each stage of the hermeneutic circle of literature acquisition (22).

When we were doing the grouping of the resources using the hermeneutic approach, we extended the observed years and made it from 2004 to 2019. From the resources that we found from our SLR and HLR approaches, we found that the major illicit dark net market places are terrorist and extremist activities, drugs and silk road market places and human trafficking in the papers which are listed here: (25) (26) (27) (28) (29) (30) (31) (32) (33) (34) (35) (36) (37) (38) (39) (40) (41) (42) (43) (44) (45) (46) (47) (48) (49) (50) (51) Other illicit activities on dark web are: fake passports, stolen credit cards, exotic animals, fake identities, and prostitution (52).

5.1. Human Trafficking and Technology

Human trafficking has been shown to be a very profitable dark web market content (53),(47) (48), (49), (50), (51). There are almost 2.5 million victims of human trafficking globally, according to the United Nations Office on Drugs and Crime, and bringing the perpetrators to justice is the hardest part (54). According to estimates from the International Labor Organization (ILO), sexual exploitation alone generates an estimated 100 billion of the estimated 150 billion that human trafficking and other forms of exploitation generate annually (53).

Human trafficking is a global issue. Nowadays, young people have a very easy access to the various social media and messaging sites at anytime in any location. Generally, young people are more talented regarding the usage of internet and technological issues compared to their parents. It has gotten simpler for youngsters to be groomed into victims of human trafficking and sexual exploitation because the majority of parents do not thought to pay appropriate attention to what their kids are doing online. Educating ourselves and our children on the risks of the Internet, social media, and what we publish online can go a long way towards keeping our children safe (55).

Human Trafficking and Law Enforcement: It is challenging for the law enforcement to locate the criminals of the Dark web as TOR hides the identity of the dark web user. By implementing routeing, encrypting incoming data, and transmitting data irregularly over the network to volunteers around the world, the TOR browser protects both users and websites (56).the globe (56).

Criminals are using technology to commit crimes in an increasing number of ways. In the recent years, law enforcement agencies worldwide have become concerned about the use of technology against online profitable malicious activities. As an illustration, the International Association of Police Chiefs (IACP) in the USA had to seek assistance from the FBI’s cyber section due to the dearth of accessible cyber training for law enforcement personnel (LEOs).

With the advancement of human traffickers using technology to persuade victims and capture predators, law enforcement has begun to deploy the technology. These systems look for predators on both the ”surface web” and the ”dark web,” where they might remain anonymous and undetected. Memex is a technology that has been developed by the Defense Advanced Research Projects Agency (DARPA) to fight against human trafficking.

5.2. Drugs and Dark Web

We stated earlier that following our review approach, we found that Drugs are one of the most prominent dark net market places on Dark web (35) (36) (37) (38) (39) (40) (41) (42) (43) (44) (45) (46).

According to a Grand Forks report, the Dark Web was directly implicated in a worldwide drug trafficking network that originated in Grand Forks, North Dakota, when an 18-year-old who overdosed fatally on the dangerous synthetic opioid substance fentanyl citrate was discovered in 2015 (57). Several of the 10 people connected to the fatal overdose case were accused of selling illegal substances on the Dark Web in the early months of 2015.

After an investigation, it was found that the drug dealers were running their business via the Dark Web using the TOR browser to keep their anonymity. The job of law enforcement authorities is made harder by the illegal drug dealers’ use of modern technology to conceal their identities and their use of the Dark Web to sell their goods from a place where they can remain anonymous (57). The types of medications that can be obtained on the internet are depicted in Figure 4.

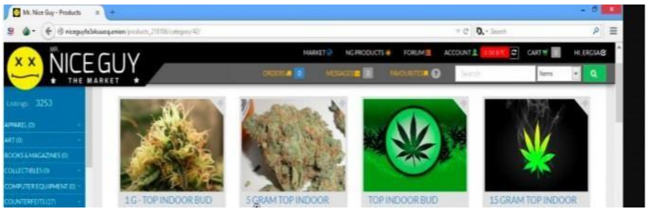


Figure 4: The drug products from Mr. Nice Guy URL:(Django, 2015)

5.3. Silk Road and Dark Web

A secret internet bazaar known as The Silk Road operated on the Dark Web. The site began operating in February 2011 and carried on until it was captured by the FBI in October 2013. The Silk Road website makes use of the TOR hidden service, making it inaccessible through regular web browsers and accessible only through the platform of the anonymous network. Further obscuring the identity of persons involved in transactions was the use of the peer-to-peer digital currency Bitcoin (58). This anonymous underground drug trafficking was conducted by Ulbricht (59) who escaped law enforcement by using the Dark Web to keep himself unknown. Figure 14 depicts photos from the Silk Road website, which indicate the various illegal goods that were offered for sale there.

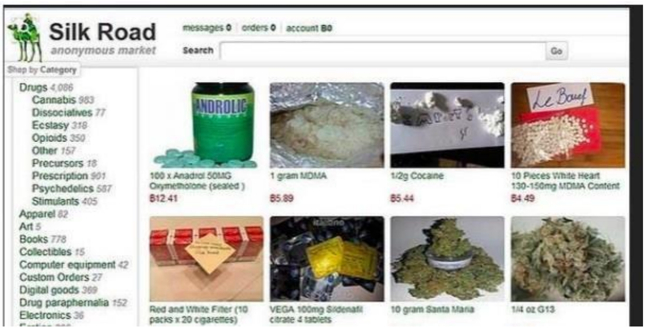


Figure 5: The Silk Road Products (58)

TABLE 3: Prominent Tor Websites for Keyword Searches

Keyword	**Extremist and Terrorism Promoting Websites**		
	Website	**URL**	**Category**
Terrorist	TorLinks Freenet FuckOff-And-Die.Com's Onion Portal "Name Unavailable"	http://torlinkbgs6aabns.onion/#po- litical http://freenet7cul5qsz6.onionhttp://3il6wiev2pnk7dat.onion http://uudllt7casd3cykd.onion	Discussion Board Discussion Board Discussion Board Discussion Board
Extremism	Contranumenism Manifestation Hack Canada	http://contra6am7tdml6h.onion http://havkcan12o41vmnv.onion	Organizer's Web- site Organizer's

			Web- site
Normal Security	Freenet	http://freenet7cul5qsz6.onion	Discussion Board

5.4. Dark Net Data Analysis on Terrorist Activities

The authors of (9) stated that manual investigations, which are time-consuming and ineffective, are used by both researchers and law enforcement. This problem inspired them to create a web crawler that automatically crawls webpages on the surface web (the internet) based on predetermined keywords, as shown in Table 3. Additionally, it followed the linkages to build the network map shown in Figure 6 by mapping the network. Then, they modified the same web crawler system to work on TOR websites. By considering these issues, they successfully developed ’Dark Crawler’ which can simultaneously access both the TOR and the normal surface web internet. The main contribution of this “Dark Crawler” is to successfully investigate the presence of extremist and terrorist contents on TOR and their mapped connections with other dark web crimes based on the keyword provided. Web graph, Degree Distribution and centrality are the algorithms that have been used to develop a web data analytic tool. The most popular website has been considered by calculating its centrality and the centrality is interpreted from in-degree and out-degree score comparison normalization. However, the major drawback of the system is that it lacks the automatic sentiment analysis,cannot classify between the extremist, non-extremist and neutral groups and does not support the multilingual option to analyze dark web crimes written in other than English.

According to (4), social network analysis (SNA) is a graph-based technique for examining a group’s or population’s network structure and how it affects social interactions. ”Dark Networks” (60) are social media platforms developed by criminal organ- isations. However, It is very challenging to track the presence of the websites and forums of the extremist and terrorist groups because they use the ephemeral natural of the dark net layer. The researchers of (26) outlined this problem of rapid disappearance and the change of the uniform resource locators (URLs) of the resource of extremist and terrorist groups online. The studies of (5) claimed that it has become obvious that there needs to be supplied systematic and automated techniques for recognising, gathering, and searching these sites due to the vast volume of sites, their dynamic and fugitive natures, their diverse languages, and noise.

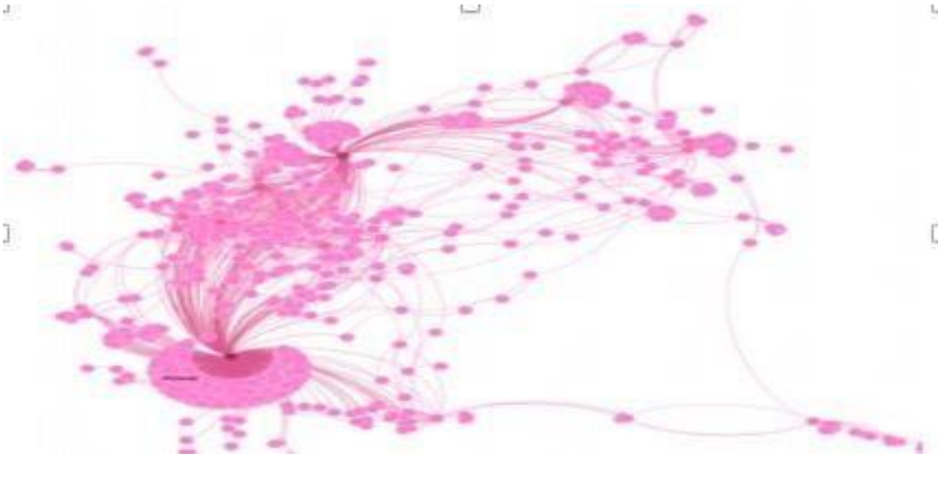


Figure 6: Collaborated Network with extracted TOR websites (9)

Following their objective, (5) performed a significant extension of their previous research. This includes broadening the area of data gathering, introducing a component for incremental spidering for routine data updates, improving the searching and browsing features, improving multilingual machine translation for Arabic, French, German, and Russian, and sophisticated social network analysis. They worked on 29 Jihadist forums, among, which, 17 are Arabic forums, 7 are English forums, 3 are French forums and the other 2 are in German and Russian, respectively. The total number of messages on the forums is about 13 million; approximately 3 million postings are added annually through incremental spidering. Their efforts have resulted in the creation of a Dark Web Forums portal, which is a framework for the data integration of web forums for searching and examining global Jihadist forums inside of a web-based information portal (5). The major drawback of their system is that it does not include a sentiment analysis engine, so that their developed system is unable to allow for deeper text mining and analysis.

For security agencies, the website's pages on improvised explosive devices serve as a valuable source of information. These websites offer various sorts and levels of information for the intelligence community using various communication genres (61). Although terrorist training and attack materials have a significant online presence, the author of (61) argues that it has been difficult to discover and analyse particularly violent content related to IED (improvised explosive device) webpages. While detecting the suspect websites, the law enforcement organisation should focus on IED Websites with a low false positive rate.

The genre classifier for categorising IED websites and Forums was created by the authors from (61) using a Complex Feature Extractor, Extended Feature Representation, and Support Vector Machine (SVM) learning algorithm. However, their major drawback is that this system cannot support multilingual feature analysis. In (62) reported the affect-analysis of Middle Eastern extremist organisation forums in comparison to US white supremacist group forums in another study in the field of measuring the presence of violence and hate speech by extremist groups. It is possible to further analyse extremist groups on the deep web using the methodologies they suggest as shown in Figure7. A wide range of content and sentiment analysis has been performed on white supremacist and extremist forums (6), (7), (8). A specific study by (63) examined the websites of American hate groups in-depth and discovered a lot of content relevant to recruiting, propaganda, and fund raising. (64) conducted a content study of 157 websites run by hate groups in the United States and discovered strong connections between some of the groups. (6) speculated that online resource sharing among white nationalist groups was common.

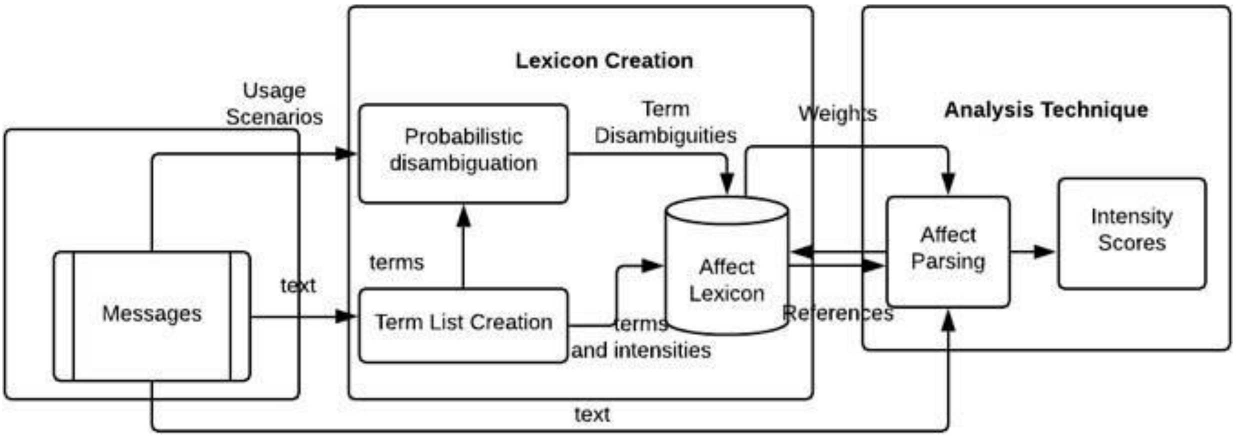


Figure 7: Affect Analysis System Design (62)

5.5. Dark Net Data Analysis on Drug Trafficking

In addition to the papers that we analysed in-depth here because of their high accuracy and application-centric character, there are a few additional studies that offer a country-specific viewpoint and are concentrated on drug trafficking and the dark net market known as Agora [8,9]. Because of the increased motivation made possible by the better anonymity the TOR Network provides, recent research (65) shows that the drug market represents the largest offer share in every TOR marketplace. The research of online drug markets in particular has gained a lot of attention in recent publications on the study of Internet organised crime, with an emphasis on the analysis of sellers' behaviour or the identification of fresh difficulties for LEAs (66; 67; 68; 69; 70).

By analysing the above research works, the researchers of (71) found the issue of dark net data privacy and safety in dealing with drugs. In (71) the AlphaBay, Nucleus and the East India Company marketplaces data have been explored. They provided a way for automatically browsing and gathering data from medicine offers on TOR marketplaces in order to focus an exploratory data analysis (EDA) on the data and, as a result, to generate a hypothesis to guide more in-depth research into the issue. Their interesting outcomes can be found from Figure 8, Figure 9 and Figure 10 which are depicted below. Their main drawback is that their developed setting is a short-term analysis on a reduced set of marketplaces.

As seen in Figure 8, the cannabis (natural, not synthetic) market is represented by Nucleus and Alphabay, with the ecstasy and stimulants sectors coming in second and third, respectively. The main European nations designated as shipping origins for pharmaceuticals bound for the East India market are displayed in Figure 9. Little bubbles in Figure 10 depict the sparse distribution of nations selling the relevant products, demonstrating the global availability of opioids, cannabis, stimulants, and benzodiazepines.

1)Geo-Location Analayis of the Dark net products: It is known that the Dark web forums related to any cyber criminal are only accessible through the TOR web browser. A huge range of cyber crimes are happening in the dark web which are drug trafficking, human trafficking, identity theft, money laundering, computer hacking, botnets, credit card frauds, gun sales, child pornography, and other related cyber-crimes. Some of this business make huge profits by destroying people's lives. For example, in 2013, the Silk Road business had been appraised around $1.2 billion and the founder has been arrested.

As these crimes are happening online and do not have any geographical territory and because there is no international law for cyber crime issues on the dark web, it is extremely important to investigate the geographical location in order to help the law

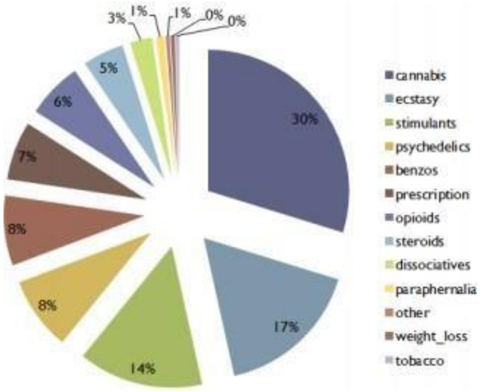


Figure 8: Overall Shares of the Drugs market (71)

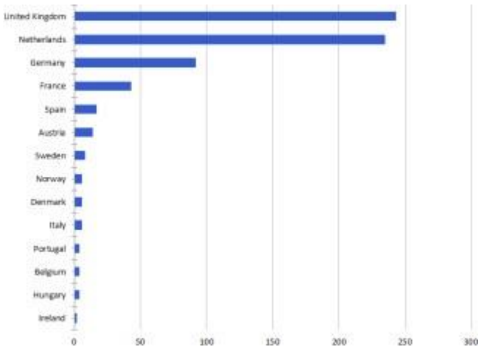


Figure 9: Eastindia offers the country of origin (71).

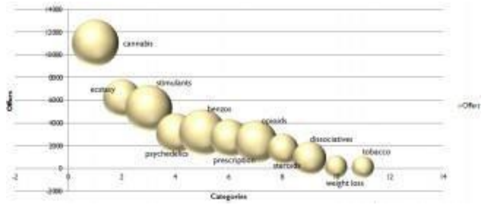


Figure 10: Categories size vs cross market vendors (71).

enforcement authorities. Unveiling the geographical distribution should be a big concern from the viewpoint of social anthropology to support and scrutinize the citizens when they are at risk. The researchers of (10), claimed they are the first to analyse the geo location of the Dark web forums. (10) asserts that the TOR and the Dark Web are crucial for defending the freedom of information and speech online, particularly in nations where the government or other strong forces work to stifle it. Strongbox or GlobaLeaks are two examples of political websites found on the Dark Web. From the known dark web forums, the researchers of (10), as presented in Figure. 11 successfully found the geolocation of some prominent dark web forums and sites by using their methods. They proposed to find the geo location by projecting the baselines on the forum crowd with the time zone (UTC). Their results from Figure. 11, Figure. 12 and Figure. 13 show that The Majestic Garden forum http://bm26rwk32m7u7rec.onionis a mostly American forum. The largest peak from the GMM is in the UTC+1 time zone, according to the Dream Market forum (http://tmskhzavkycdupbr) onion forum (Berlin, Paris, Rome). It was determined that this forum is largely European, with a few other exceptions.

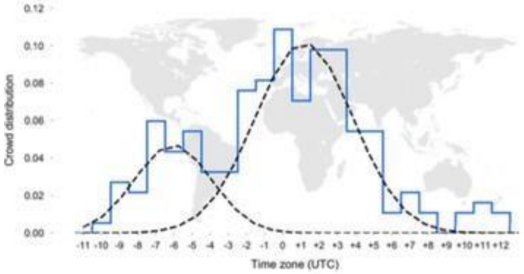


Figure 11: Dream Market forum,http://tmskhzavkycdupbr.onionOfficial forum of Dream Market Marketplace (10).

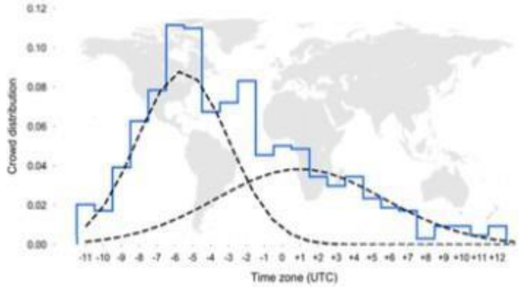


Figure 12: Pedo support community,http://support26v5pvkg6.onion(10).

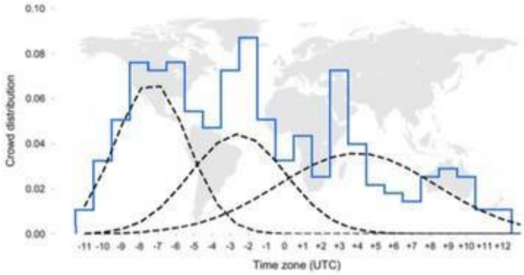


Figure 13: The Majestic Garden,http://bm26rwk32m7u7rec.onionPsychedelic forum (10).

Cryptomarkets are located on the dark net forum and operate as the online trading place of a variety of illegal goods. Agora, Evolution, SIlk road2 are top-ranked Cryptomarkets. These markets mostly sell narcotics, but they also sell other illegal items like doping products, counterfeit goods, guns, and financial or identity fraud. This research by (72) was conducted using the Evolution Cryptomarket to create a study of the trafficking network. They also sought to draw attention to the geographic trends in the trafficking of these goods. Thus, by applying the data analysis algorithms, they successfully evaluated the country specificity of the dark web drug marketing. Their developed system revealed that the English-speaking nations—such as the United States, the United Kingdom, and Australia—and Western European nations—such as the Netherlands—dominate the trafficking of illicit drugs on cryptomarkets. Cannabis, stimulants—such as cocaine and amphetamines—ecstasy (MDMA), and psychedelics— such as NPS and LSD—are the main drugs offered on these platforms.

Dark Net Volume Measurement and Content Analysis: According to (39), a significant amount of in-depth study has been carried out over the past ten years to improve the actual measures from various online criminal ecosystems in order to develop effective defenses (11),(12), (13), (14), (15), (16), (73), (74), (75), (76). The dark web becomes a safe place to conduct illegal activities anonymously due to its prerogative nature. Darknet websites are known as ”Hidden Services” (HS) in the TOR community and can be accessed via a unique browser called the TOR Browser. (77) discussed the exponential growth in the number of Darknet domains, which increased from 30K to 60K between August 2015 and 2016. Due to the unclear nature of the Dark net, there are only 6 to 7 thousand publicly available domains (78). (77) concentrated on establishing and developing a system that categorises the unlawful actions on the Darknet in response to the significant concealed materials and excessive misuse on the Darknet. where they used a publickly available dataset, Dark net Usage Text Addresses (DUTA) genarated from TOR Darknet. By combining three classifiers—SVM, LR, and NB—with two text representation techniques, TF-IDF and BOW, their proposed system can classify the unlawful behaviours of TOR HS with 96.6% accuracy. Details of their finding have shown in Figure 14. However, the main drawback of their system is that it can handle a small amount of dataset compared to the volume of the dark web resource.

In measuring the long-term growth of the online anonymous marketplace ecology, (39) made a substantial contribution. In order to have a thorough picture of the development of the online anonymous marketplace ecosystem, they conducted a long-term measurement (four years) on dark web resources. Additionally, they tracked changes in the products being sold and evaluated how antagonistic occurrences like large-scale fraud or law enforcement activities affected the growth of the economy as a whole. This study’s main contribution is to provide an assessment of how traditional and offline criminal activity is evolving online, much like how traditional commerce diversified online in the 1990s.

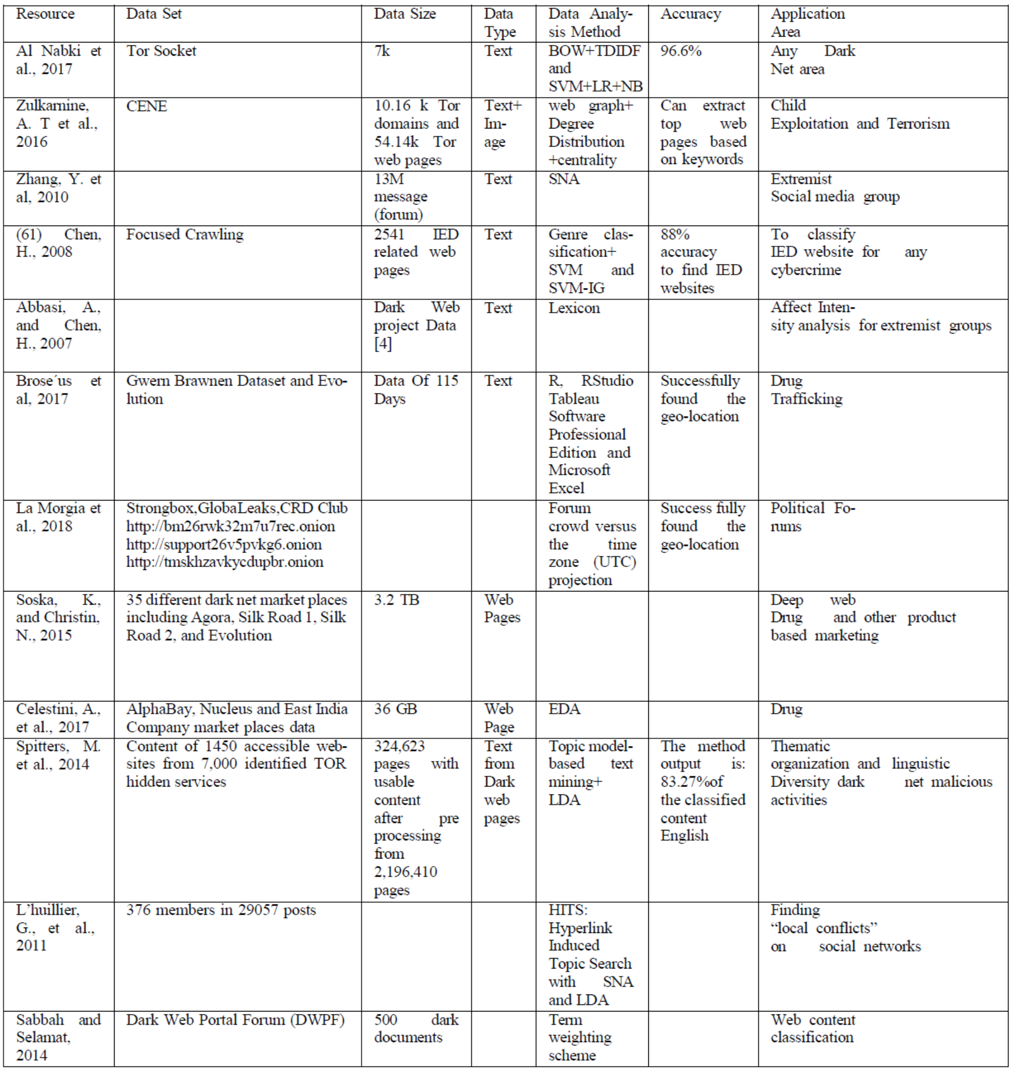
More than a thousand TOR hidden services’ contents were mined using categorisation and topic model-based text mining algorithms in (79) to model their linguistic diversity and theme arrangement. The results are presented in Figure. 15 and Figure 16. However, this system lacks detailed topic level analysis. To develop their system, they used topic-model based text mining algorithms and text classifiers.

Thematic Analysis of Dark Net Data: A good number of research work has been performed on Dark Web content analysis. Below is a review which is listed in Table IV and Table **??**.Topic modelling and social network analysis have been conducted in the work of (80), although this primarily focuses on information produced by extremists or terrorists, most of which is related to Islamic doctrine and religion. Using the incremental spidering method, they gathered 3,504 threads and 29,016 postings for Social Network Analysis (SNA). These threads come from 29 forums, out of which 17 are in Arabic, 7 are in English, 3 are in French, and the remaining 2 are, respectively, in German and Russian. Almost the same approach (HITS -Hyperlink-Induced Topic Search- with SNA and LDA) has been applied on (80) for finding “local conflicts” on social networks. This research claims their application on discovering Dark web potential groups of topics (e.g. potential homelandsecurity threats), and key-members that potentiate thesetopics. Their total dataset was on 376 members in 29057 posts.

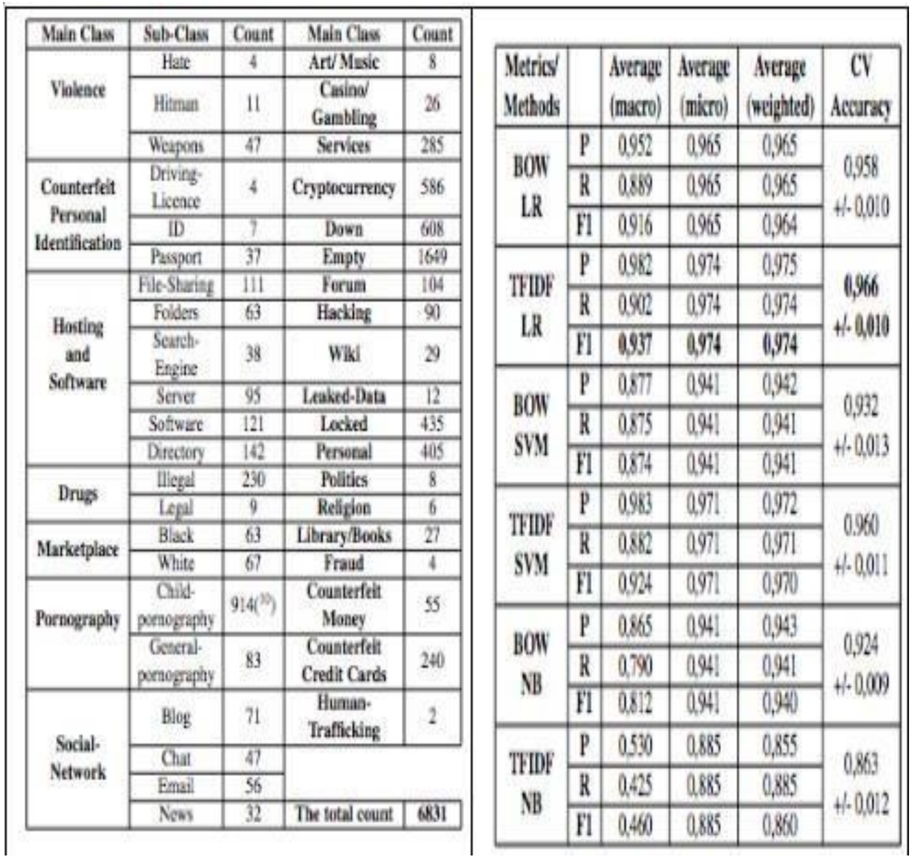
6. DARK NET CHALLENGES

The analysis of dark web is becoming a prominent research area day by day. Researchers and law enforcement people are taking a deeper investigation to properly understand the depth, content, malicious trades happening on the dark web. At the same TABLE 4: Analysis of the Dark Net Data

TABLE 4: Analysis of the Dark Net Data



Resource	Data Set	Data Size	Data Type	Data Analysis Method	Accuracy	Application Area
Al Nabki et al., 2017	Tor Socket	7k	Text	BOW+TDIDF and SVM+LR+NB	96.6%	Any Dark Net area
Zulkarnine, A. T et al., 2016	CENE	10.16 k Tor domains and 54.14k Tor web pages	Text+ Image	web graph+ Degree Distribution +centrality	Can extract top web pages based on keywords	Child Exploitation and Terrorism
Zhang, Y. et al, 2010		13M message (forum)	Text	SNA		Extremist Social media group
(61) Chen, H., 2008	Focused Crawling	2541 IED related web pages	Text	Genre classification+ SVM and SVM-IG	88% accuracy to find IED websites	To classify IED website for any cybercrime
Abbasi, A., and Chen, H., 2007		Dark Web project Data [4]	Text	Lexicon		Affect Intensity analysis for extremist groups
Brose´us et al, 2017	Gwern Brawnen Dataset and Evolution	Data Of 115 Days	Text	R, RStudio Tableau Software Professional Edition and Microsoft Excel	Successfully found the geo-location	Drug Trafficking
La Morgia et al., 2018	Strongbox,GlobaLeaks,CRD Club http://bm26rwk32m7u7rec.onion http://support26v5pvkg6.onion http://tmskhzavkycdupbr.onion			Forum crowd versus the time zone (UTC) projection	Success fully found the geo-location	Political Forums
Soska, K., and Christin, N., 2015	35 different dark net market places including Agora, Silk Road 1, Silk Road 2, and Evolution	3.2 TB	Web Pages			Deep web Drug and other product based marketing
Celestini, A., et al., 2017	AlphaBay, Nucleus and East India Company market places data	36 GB	Web Page	EDA		Drug
Spitters, M. et al., 2014	Content of 1450 accessible websites from 7,000 identified TOR hidden services	324,623 pages with usable content after pre processing from 2,196,410 pages	Text from Dark web pages	Topic model-based text mining+ LDA	The method output is: 83.27%of the classified content English	Thematic organization and linguistic Diversity dark net malicious activities
L'huillier, G., et al., 2011	376 members in 29057 posts			HITS: Hyperlink Induced Topic Search with SNA and LDA		Finding "local conflicts" on social networks
Sabbah and Selamat, 2014	Dark Web Portal Forum (DWPF)	500 dark documents		Term weighting scheme		Web content classification



Main Class	Sub-Class	Count	Main Class	Count
Violence	Hate	4	Art/ Music	8
	Hitman	11	Casino/ Gambling	26
	Weapons	47	Services	285
Counterfeit Personal Identification	Driving-Licence	4	Cryptocurrency	586
	ID	7	Down	608
	Passport	37	Empty	1649
Hosting and Software	File-Sharing	111	Forum	104
	Folders	63	Hacking	90
	Search-Engine	38	Wiki	29
	Server	95	Leaked-Data	12
	Software	121	Locked	435
	Directory	142	Personal	405
Drugs	Illegal	230	Politics	8
	Legal	9	Religion	6
Marketplace	Black	63	Library/Books	27
	White	67	Fraud	4
Pornography	Child-pornography	914(^{30})	Counterfeit Money	55
	General-pornography	83	Counterfeit Credit Cards	240
Social-Network	Blog	71	Human-Trafficking	2
	Chat	47		
	Email	56		
	News	32	The total count	6831

Metrics/ Methods		Average (macro)	Average (micro)	Average (weighted)	CV Accuracy
BOW LR	P	0,952	0,965	0,965	0,958 +/- 0,010
	R	0,889	0,965	0,965	
	F1	0,916	0,965	0,964	
TFIDF LR	P	0,982	0,974	0,975	**0,966 +/- 0,010**
	R	0,902	0,974	0,974	
	F1	**0,937**	**0,974**	**0,974**	
BOW SVM	P	0,877	0,941	0,942	0,932 +/- 0,013
	R	0,875	0,941	0,941	
	F1	0,874	0,941	0,941	
TFIDF SVM	P	0,983	0,971	0,972	0,960 +/- 0,011
	R	0,882	0,971	0,971	
	F1	0,924	0,971	0,970	
BOW NB	P	0,865	0,941	0,943	0,924 +/- 0,009
	R	0,790	0,941	0,941	
	F1	0,812	0,941	0,940	
TFIDF NB	P	0,530	0,885	0,855	0,863 +/- 0,012
	R	0,425	0,885	0,885	
	F1	0,460	0,885	0,860	

Figure 14: Left Figure: DUTA Dataset Classees from [Al Nabkiet al., 2017]. Right figure: A comparison between the classification pipelines with respect to 10 folds cross-validation accuracy (CV), precision (P), recall (R) and F1 score metrics for micro, macro and weighted averaging. [Al Nabki et al., 2017]

time, the black hackers are creating new mechanisms to hide them more perfectly. Illegally placed dark webs and the detection of illegal activities are still an open research area.

The dark web is purposefully kept out of search results using anonymity techniques, and is the storage of a huge number of illicit activities and products. As per some studies (81), (77), the dark net consists of contents ranges from illegal pornography to drugs and weapons. It is stated in some works (77),(82), that the dark web also provides the information on botnets that are offered for rent, stolen datasets with login information, or zero-day flaws that can be used to predict, find, or preferably prevent the cyberattacks.

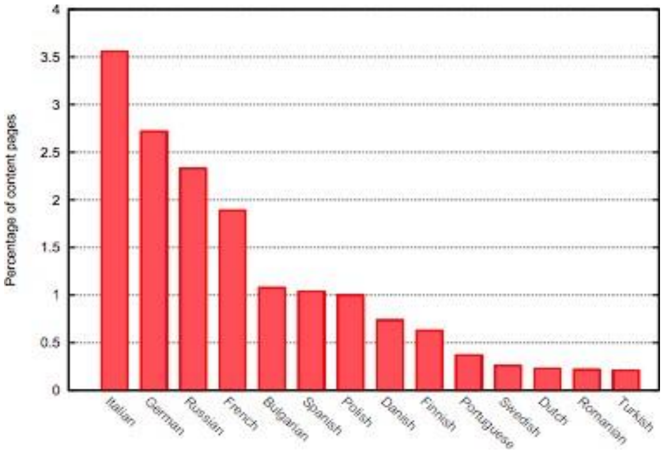
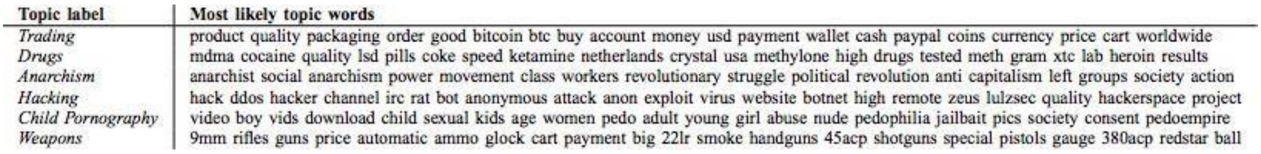


Figure 15: Percentages of languages other than English, computed on our collection of TOR hidden service content pages. On 83.27% of the analyzed content pages, the main written language was English. Languages with a share less than 0.1% are not shown. (79)



Topic label	Most likely topic words
Trading	product quality packaging order good bitcoin btc buy account money usd payment wallet cash paypal coins currency price cart worldwide
Drugs	mdma cocaine quality lsd pills coke speed ketamine netherlands crystal usa methylone high drugs tested meth gram xtc lab heroin results
Anarchism	anarchist social anarchism power movement class workers revolutionary struggle political revolution anti capitalism left groups society action
Hacking	hack ddos hacker channel irc rat bot anonymous attack anon exploit virus website botnet high remote zeus lulzsec quality hackerspace project
Child Pornography	video boy vids download child sexual kids age women pedo adult young girl abuse nude pedophilia jailbait pics society consent pedoempire
Weapons	9mm rifles guns price automatic ammo glock cart payment big 22lr smoke handguns 45acp shotguns special pistols gauge 380acp redstar ball

Figure 16: The top words for some sampled topics of 250 topic LDA model.(79)

6.1. No Inbound and Out-Bound Links

Regarding the actual volume of the Dark Net site there are several works done and it is still an open research to measure the volume of the dark web. This work (83) mentioned that they crawled more than 6,600 website main pages related to Bitcoin scams and bank card frauds where another study (84) revealed that 87% of these sites did not have any inbound or outbound links among them. This is completely different from the websites on the surface web which gives us a conclusion that the dark web sites operate individually and the reason they do this is a mystery.

6.2. Access to Dark Net Data

The first challenge is to identify the target forums of specific topics. There is no online website link on Dark net market places and the illicit activities. A cyber threat intelligence business called Intelliagg has made an attempt to access Dark Web websites via the Onion Router in order to map the Dark Web. Initially, they discovered about 30,000 websites; however, during their investigation, more than half of them vanished (81). From their report (81), we can conclude that there is not enough research done to keep the information about target forums up to date.

6.3. Dearth of Ground Truth

It is the norm that the scientific measurements should be compliant for testing and validation. In the area of the Dark Net, the ground truth is not available due to the limited number of research work that has been done in this area, and there are many technical challenges here. This area definitely needs more research to overcome the technical barriers and, thus, to collect more Dark Net data.

6.4. Static Vs Dynamic Challenges

Most of our collected research works performed strenuous efforts on the static environment for Dark Net data capture and analysis. From our initial analysis, we can say that the real-time analysis system is a vital requirement due to the short lifespan of some Dark Net sites and forums. Within our resources, we found only one research work (85) that implemented the real time Dark Net data analysis option.

7. CONCLUSIONS

Machine learning, data mining and big data analytic techniques have been used in most of the existing tools to collect important information from social media sites and sites from the Dark Net; for examples, hackers' forums, the forums of terrorists and stolen storage devices sold on the dark net marketplaces to identify threads of data breaches and cybercrimes. Further analysis can be done by ranking the malicious activities, their profits, their presence to help the law enforcement agencies. Future research directions are:

- Monitoring of Social media sites to observe message exchanges which leads to creation of new Dark Web domains;
- Monitoring of customer data for connections to non-standard domains to find new hidden services;
- Observing the hidden services for navigating the Dark Nets to find out the new sites for later analysis;
- Tracing future illegal activities for which thematic analysis can be performed;
- Analysing the content(sellers, users and the kinds of good) and the rising of the Dark Net business exchange to identify and monitor malicious and illegal activities.
- Geo-location specific research should be performed in collaboration with law enforcement agencies.
- Most of the techniques used supervised techniques, since, due to the extremely rapid growth of the Dark Net, criminals are always engaging in new ways of obfuscation of their motives. The sheer size of the Dark Net demands newer techniques such as automated tools with the ability of integrating new knowledge to the supervised system. This could be an important research direction for future.

REFERENCES

[1] *Darknet Vs Dark Web Vs Deep Web Vs Surface Web*, 2019. [Online]. Available: https://techlog360.com/darknet-vs-dark- web-vs-deep-web-vs-surface-web/

[2] R. Broadhurst, D. Lord, D. Maxim, H. Woodford-Smith, C. Johnston, S. Carroll, H. Chung, H. Trivedi, and B. Sabol, "Malware trends on 'darknet' crypto-markets: Research review," 07 2018.

[3] Cybercrime will cost businesses over 2 trillion dollar by 2019. [Online]. Available: https://www.juniperresearch.com/press/press-releases/cybercrime-cost-businesses-over2trillion

[4] D. Liben-Nowel, *The Link-prediction Problem for Social Networks*. JASIST, 58(7), pp. 1019-1031, 2007.

[5] Y. Zhang, S. Zeng, C. N. Huang, L. Fan, Y. Yu, X.and Dang, and H. Chen, *Developing a dark web collection and infrastructure for computational and social sciences*. In 2010 IEEE International Conference on Intelligence and Security Informatics (pp. 59-64). IEEE, 2010, May.

[6] A. T. Gustavson and D. E. Sherkat, *The Ideological Structuring of White Supremacy on the Internet: Analyzing Network Size*. Density, and Asymmetry, 2004.

[7] Y. Zhou, J. Qin, G. Lai, E. Reid, and H. Chen, *Exploring the dark side of the web: collection and analysis of us extremist online forums*. In International Conference on Intelligence and Security Informatics (pp. 621-626). Springer, Berlin, Heidelberg., 2006.

[8] J. Xu, H. Chen, Y. Zhou, and J. Qin, *On the topology of the dark web of terrorist groups*. In International Conference on Intelligence and Security Informatics (pp. 367-376). Springer, Berlin, Heidelberg, 2006.

[9] A. T. Zulkarnine, R. Frank, B. Monk, J. Mitchell, and G. Davies, *Surfacing collaborated networks in dark web to find illicit and criminal content*. In 2016 IEEE Conference on Intelligence and Security Informatics (ISI) (pp. 109-114). IEEE, 2016.

[10] M. La Morgia, A. Mei, S. Raponi, and J. Stefa, *Time-Zone Geolocation of Crowds in the Dark Web*. In 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS) (pp. 445-455). IEEE, 2018.

[11] D. WANG, G. VOELKER, and S. SAVAGE, *Juice: A longitudinal study of an SEO botnet*. In Proceedings of ACM CCS 2011 (Chicago, IL, Oct. 2011), 2013).

[12] D. WANG, M. DER, M. KARAMI, L. SAUL, D. MCCOY, S. SAVAGE, and G. VOELKER, *Search + seizure: The effectiveness of interventions on seo campaigns*. In Proceedings of ACM IMC'14 (Vancouver, BC, Canada, Nov. 2014), 2014).

[13] T. MOORE, N. LEONTIADIS, and N. CHRISTIN, *Fashion crimes: Trending term exploitation on the web*. In Proceedings of ACM CCS 2011 (Chicago, IL, Oct. 2011), 2011).

[14] D. MCCOY, A. PITSILLIDIS, G. JORDAN, N. WEAVER, C. KREIBICH, B. KREBS, G. VOELKER, S. SAVAGE, and K. LEVCHENKO, *Pharmaleaks: Understanding the business of online pharmaceutical affiliate programs*. In Proceedings of USENIX Security 2012 (Bellevue, WA, Aug. 2012), 2012).

[15] K. LEVCHENKO, N. CHACHRA, B. ENRIGHT, C. FELEGYHAZI, M.and GRIER, T. HALVORSON, C. KANICH, C. KREIBICH, H. LIU, D. MCCOY, A. PITSILLIDIS, N. WEAVER, V. PAX-SON, G. VOELKER, and S. SAVAGE, *Click trajectories: End-to-end analysis of the spam value chain*. In Proceedings of IEEE Security and Privacy (Oakland, CA, May 2011), 2011).

[16] J. JOHN, F. YU, Y. XIE, M. ABADI, and A. KRISHNA-MURTHY, *deSEO: Combating search-result poisoning*. In Proceedings of USENIX Security 2011 (San Francisco, CA, Aug. 2011), 2011).

[17] B. Kitchenham and S. Charters, *Guidelines for performing systematic literature reviews in software engineering*. Technical Report, Ver. 2.3 EBSE Technical Report. EBSEsn, 2007.

[18] Cecez-kecmanovic and S. K. B. Dubravka, *Systematic review and the hermeneutic circle of literature review*. University of New South Wales, Bibliometric Informetric Research Group, 2010.

[19] P. Boote, D. N. &Beile, *Scholars Before Researchers: On the Centrality of the Dissertation Literature Review in Research Preparation*. Educational Researcher, 2005, vol. 36(4).

[20] C. Hart, *Doing a Literature Review. Releasing the Social Science Research Imagination*. Thousand Oaks: SAGE Publications, 1998.

[21] W. W. Schwarz A. & Mehta M. & Johnson N.& Chin, *Doing a Literature Review. Releasing the Social Science Research Imagination*. DATA BASE for Advances in Information System, 2007, vol. 38(3).

[22] D. Boell, S. K. &Cecez-Kecmanovic, *A hermeneutic approach for conducting literature reviews and literature searches*. CAIS, 2014, vol. 34,12.

[23] T. S. Kuhn, *The structure of scientific revolutions*. Chicago and London, 1962.

[24] C. Wohlin, *Guidelines for snowballing in systematic literature studies and a replication in software engineering*. Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering. ACM, 2014.

[25] Y. Zhou, E. Reid, J. Qin, H. Chen, and G. Lai, *U.S. extremist groups on the web: Link and content analysis*. IEEE Intelligent Systems, 20(5), 44-51, 2005.

[26] G. Weimann,*www.terror.net:How modern terrorism uses the Internet*. Special Report, US Institute of Peace. Retrieved from http://www.usip.org/pubs/specialreports/sr116.pdf,2004.

[27] W. Tsfati and G. Weimann,*www.terrorism.com:Terror on the Internet*. Studies in Conflict Terrorism, 25(3), 317-332., 2004.

[28] Chen, H. and M. Sageman, *The dark web portal: Collecting and analyzing the presence of domestic and international terrorist groups on the web*. Proceedings of the seventh Annual IEEE Conference on Intelligent Transportation Systems, 2004.

[29] H. Chen, W. Chung, and J. Qin, *Uncovering the DarkWeb: A Case Study of Jihad on the Web*. Journal of the American Society for Information Science and Technology (JASIST), vol. 59, no. 8, 2008.

[30] C. Yang and T. D. Ng, *Analyzing Content Development and Visualizing Social Interactions in Web Forum*. Conference on Intelligence and Security Informatics (ISI'2008), pp. 25-30., 2008.

[31] E. Reid, J. Qin, and W. Chung, *Terrorism Knowledge Discovery Project: A Knowledge Discovery Approach to Addressing the Threats of Terrorism*. pp. 125-145, 2004.

[32] Y. Zhou, J. Qin, and G. Lai, *Collection of U.S. Extremist Online Forums: A Web Mining Approach*. in Annual Hawaii International Conference on System Science, 2007.

[33] H. Chen, *Intelligence and Security Informatics for International Security: Information Sharing and Data Mining*. London: Springer Press, 2006.

[34] ——, *Exploring Extremism and Terrorism on the Web: The Dark Web Project*. Lecture Notes in Computer Science, 2007.

[35] J. Martin, *Lost on the Silk Road: Online drug distribution and the "cryptomarket*. Criminol. Crim. Justice, 14, 351–367, 2014.

[36] J. Van-Buskirk, S. Naicker, A. Roxburgh, R. Bruno, and L. Burns, *Who sells what? Country specific differences in substance availability on the Agora cryptomarket*. Int. Journal of Drug Policy, 35, 16–23, 2016.

[37] D. S. Dolliver, S. P. Ericson, and K. L. Love, *A Geographic Analysis of Drug Trafficking 572 Patterns on the TOR Network*. Geogr. Rev., pp: 1–24, 2016.

[38] N. Christin, *Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace*. Proc. 22nd Int. World Wide Web Conf., Rio de Janeiro, Brazi, 2012.

[39] K. Soska and N. Christin, *Measuring the Longitudinal Evolution of the Online Anonymous Marketplace Ecosystem*. in Proc. 22nd USENIX Secur. Symp. (USENIX Secur. 2015), Washington, DC., pp. 33–48, 2015.

[40] M. C. Van-Hout and T. Bingham, *Surfing the Silk Road": A study of users' experiences*. Int. J. Drug Policy, 24, 524–529., 2013.

[41] ——, *Silk Road", the virtual drug 581 marketplace: A single case study of user experiences*. Int. J. Drug Policy, 24, 385–391., 2013.

[42] ——, *Responsible vendors, intelligent consumers: Silk Road, the online revolution in drug trading*. Int. J. Drug Policy, 2014, 25, 183–189, 2013.

[43] M. J. Barratt, J. A. Ferris, and A. R. Winstock, *Use of Silk Road, the online drug marketplace, in the United Kingdom, Australia and the United States*. Addiction, 2014, 109, 774–783, 2014.

[44] ——, *Safer scoring? Cryptomarkets, social supply and drug market violence*. Int. J. Drug Policy, 35, 24–31, 2016.

[45] J. J. Brose´us, D. Rhumorbarbe, C. C. Mireault, V. Ouellette, F. Crispino, and D. De´cary-He´tu, *Studying illicit drug trafficking on Darknet markets: Structure and organisation from a 591 Canadian perspective*. Forensic Sci. Int., 2016, 264, DOI 10.1016/j.forsciint.2016.02.045, 2016.

[46] J. Murdock, *Silk Road 2.0 staffer sentenced to 8 years in prison for running dark web drug market*. Retrieved from http://www.ibtimes.co.uk/silk-road-2-0-staffer- sentenced-8-years-prison-running-dark-web-drug-market-1563964, 2016.

[47] L. Greenemeier, *Human Traffickers Caught on Hidden Internet*. Retrieved from https://www.scientificamerican.com/article/human-traffickers-caught-on-hidden-internet/,2015, FEB 8.

[48] C. Reilly, *Human Trafficking: A Crime Hard to Track Proves Harder to Fight*. Retrieved from http://www.pbs.org/wgbh/frontline/article/what-is-human-trafficking-and-why-is-it-so-hard-to-combat/,2015, July 29.

[49] J. Satterfield, *FBI tactic in child porn sting under attack in Rockwood case*. Retrieved from http://archive.knoxnews.com/news/crime-courts/fbi-tactic-in-child-porn-sting-under-attack-in-rockwood-tenn-case-3b688c59- e467-5bae-e053-0100007f68-392361101.html/, 2016, September 06.

[50] FBI.gov., *'Playpen' Creator Sentenced to 30 Years*. Retrieved from https://www.fbi.gov/news/stories/playpen-creator- sentenced-to-30-years, 2017, May 5.

[51] C. Farivar, *Creator of infamous Playpen website sentenced to 30 years in prison*. Retrieved from https://arstechnica.com/tech- policy/2017/05/creator-of-infamous-playpen-website-sentenced-to-30-years-in-prison/s, 2017, May 5.

[52] M. Jerry, *Dark Web: Problems Law Enforcement Investigations Face On The Dark Web*. A Capstone Project Submitted to the Faculty of Utica College, 2017.

[53] B. Luscome, *Inside the Scarily Lucrative Business Model of Human Trafficking*. http://time.com/105360/inside-the-scarily- lucrative-businessmodel-of-human-trafficking, 2014.

[54] M. Morris, *The Impact of Advancing Technologies Upon Global Human Trafficking and Sexual Exploitation in Society Today*. Doctoral dissertation, Utica College, 2017.

[55] CybersafetyCop, *Dangerous Apps on Your Teen's Mobile Device*. http://cybersafetycop.com/Blog/Cyber-Safety- Cop/DangerousAppsonYourTeensMobileDevice, 2015.

[56] A. Estes, *Tor: The Anonymous Internet, and If It's Right for You*. Retrieved from https://gizmodo.com/Tor-the-anonymous- internet-and-if-its-right-for-you1222400823, 2013.

[57] S. Volpenhein, *Grand Forks Herald*. etrieved fromwww.govtech.com/internet/Dark-Web-Poses-Challenges-for-Law- Enforcement.html, 2017.

[58] J. M. Sears, *A reputation for the good stuff: user feedback signaling and the deep web market silk road*. Doctoral dissertation, Montana State University-Bozeman, College of Agriculture, 2016.

[59] B. Weiser, *Retrieved from https://www.nytimes.com/2015/05/30/nyregions/ross-ulbricht-creator-or-silk-roadwebsite-is- sentenced-to-life-in-prison*. The New York Times, 2017.

[60] J. Raab and H. B. Milwar, *Dark Networks as Problems*. Journal of Public Administration Research and Theory, Vol. 13, pp. 413-439, 2003.

[61] H. Chen, "Ieds in the dark web: Genre classification of improvised explosive device web pages," *In 2008 IEEE International Conference on Intelligence and Security Informatics. IEEE.*, pp. 94–97, 2008.

[62] A. Abbasi, H. Chen, and A. Salem, *Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums*. ACM Transactions on Information Systems (TOIS), 26(3), 1-34., 2008.

[63] Y. Zhou, J. Qin, G. Lai, and H. Chen, *Collection of us extremist online forums: A web mining approach*. In 2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07) (pp. 70-70). IEEEl, 2007.

[64] Y. Zhou, E. Reid, J. Qin, H. Chen, and G. Lai, *US domestic extremist groups on the Web: link and content analysis*. IEEE intelligent systems, 20(5), 44-51, 2005.

[65] D. S. Dolliver and J. L. Kenney, *Characteristics of drug vendors on the tor network: A cryptomarket comparison*. Victims and Offenders, pages 1–21, 2016.

[66] EUROPOL, *The internet organised crime threat assessment (iocta). Technical report*. EUROPOL, 2015.

[67] K. Thomas, D. Huang, D. Wang, E. Bursztein, C. Grier, and G. Holt, T. J.andVigna, *Framing dependencies introduced by underground commoditization*. In Workshop on the Economics of Information Security, 2015.

[68] M. Mignone and E. Bosio, *Criminological analysis of the nps market*. Technical report, RISSC, 2016.

[69] L. Laura and G. Me, *Searching the web for illegal content: the anatomy of a semantic search engine*. In International Conference on Global Security, Safety, and Sustainability, pages 113–122. Springer, 2015.

[70] J. Buxton and T. Bingham, *The rise and challenge of dark net drug markets*. Policy Brief, 7, 1-24, 2015.

[71] M. G.Celestini, A. and M. Mignone, *Tor marketplaces exploratory data analysis: the drugs case*. In International Conference on Global Security, Safety, and Sustainability (pp. 218-229). Springer, Cham., 2017.

[72] J. Brose´us, D. Rhumorbarbe, M. Morelato, L. Staehli, and Q. Rossy, *A geographical analysis of trafficking on a popular darknet market*. Forensic science international, 277, 88-102, 2016.

[73] L. LU, R. PERDISCI, and W. LEE, *SURF: Detecting and measuring search poisoning*. In Proceedings of ACM CCS 2011 (Chicago, IL, Oct. 2011), 2011).

[74] Z. LI, S. ALRWAIS, X. WANG, and E. ALOWAISHEQ, *Hunting the red fox online: Understanding and detection of mass redirect-script injections*. In Proceedings of the 2014 IEEE Symposium on Security and Privacy (Oakland'14) (San Jose, CA, May 2014), 2014).

[75] N. CHRISTIN, S. YANAGIHARA, and K. KAMATAKI, *Dissecting one click frauds*. In Proc. ACM CCS'10 (Chicago, IL, Oct. 2010), Oct,2010).

[76] N. CHRISTIN, *Traveling the Silk Road:A measurement analysis of a large anonymous online marketplace*. In Proceedings of the 22nd World Wide Web Conference (WWW'13) (Rio de Janeiro, Brazil), pp. 213–224, May, 2013.

[77] M. W. Al-Nabki, E. Fidalgo, E. Alegre, and I. de Paz, *Classifying illegal activities on TOR network based on web textual contents*. in Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, 2017.

[78] V. Ciancaglini, M. Balduzzi, R. McArdle, and M. Rosler, *Below the surface: Exploring the deep web*. Trend Micro Incorporated, 2016.

[79] M. Spitters and V. S. M. Verbruggen, S., "Towards a comprehensive insight into the thematic organization of the tor hidden services," *In 2014 IEEE Joint Intelligence and Security Informatics Conference*, pp. 220–223, 2014.

[80] G. L'huillier and R. S. A.-A. F. Alvarez, H., *Topic-based social network analysis for virtual communities of interests in the dark web*, 2011, vol. 12(2).

[81] Intelliagg, *Deeplight: Shining a Light on the Dark Web. An Intelliagg Report*, 2016.

[82] A. Biryukov, I. Pustogarov, and R. Weinmann, *Trawling for tor Hidden Services: Detection, measurement, deanonymization*. in IEEE Symposium on Security and Privacy (S and P), 2013.

[83] HyperionGray, *Hyperion Gray, "Dark Web Map,"*. Available:https://www.hyperiongray.com/dark-web-map/.[Accessed 7 1 2019]., 2019.

[84] X. Y. Griffith, V. and C. Ratti, *Graph Theoretic Properties of the Darkweb*. arXiv preprint arXiv:1704.07525, 2017, 2017.

[85] F. M. S.-M. E. M. L. M. Scha¨fer, M. and V. Lenders, *BlackWidow: Monitoring the Dark Web for Cyber Security Information*. In 2019 11th International Conference on Cyber Conflict (CyCon) (Vol. 900, pp. 1-21). IEEE, 2019.

AUTHORS

HumayraBinte Ali is presently a sessional academic in ATMC, Melbourne and teaching postgraduate IT courses to Federation University and Western Sydney University. She obtained her PhD from Flinders University of South Australia in 2017 in Artificial Intelligence and Machine Learning.

Mohamed Abdelrazekis a Professor of Software Engineering and IoT in A2I2, Deakin university. Mohamed has more than 15 years of software industry, research and teaching experience. Before joining Deakin University in 2015, Mohamed worked as a senior research fellow at Swinburne University of Technology and Swinburne-NICTA software innovation lab (SSIL). Mohamed has deep experience in designing, developing, integrating, and managing large-scale software systems including military process automation, ERP, and CRM systems. Mohamed is an active researcher in automated software engineering, cloud computing security, formal methods, highperformance computing, IoT, and data science. He has numerous research articles published in top-ranked international journals and conferences.

Shamsul Huda is a Senior Lecturer in School of Information Technology, Deakin University, Australia. Prior to join Deakin, he worked as an academic in Federation University and as an Assistant Professor in Khulna University of Engineering and Technology (KUET), Bangladesh. Dr Huda is a Certified Information System Security Professional (CISSP) by The International Information System Security Certification Consortium, (ISC)² . His main research areas are Communication and network security, secure operations for Industrial Control systems (SCADA) and Critical infrastructure, Intelligent counter measure against Mobile malware, detection of data breaches, IoT security, Malware analysis and detection, reverse engineering for endpoint security. He has published more than 70 journal and conference papers in well reputed journals.

MdMamunur Rashid received his Ph.D. degree in Computer Science from Monash University. Currently, he is working as a Senior Lecturer at the School of Engineering and Technology, CQUniversity, Australia. Prior to join CQUniversity, he worked as an academic at Monash University, Australia. His research interest includes cybersecurity, big data analytics, machine learning and distributed computing.

Md. **Toufiqur Rahman** has Bachelor of Science in Honors and Master of Science (M.Sc.) in Mathematics from the Department of Mathematics, Khulna University of Engineering & Technology (KUET), Bangladesh. Currently he is working as s research assistant in the Department of Mathematics, Khulna University of Engineering & Technology (KUET). His main research area is supply chain and Transportation problems, and optimization.

Amani Ibrahim is currently working at Deloitte Australia. Previously, she worked as a Senior Lecturer in cyber security with the School of Information Technology, Deakin University. He is a cybersecurity professional with over a decade of experience across academia and industry. He is also the Cybersecurity Research Discipline Lead with the Deakin Software and Technology Innovation Laboratory.