A Review of Dark Web: Trends and Future Directions

Shahriar Sobhan*, Timothy Williams†, Md Jobair Hossain Faruk*, Juan Rodriguez*
Masrura Tasnim*, Edwin Mathew*, Jack Wright* & Hossain Shahriar‡

*Institute for Cybersecurity Workforce Development, Kennesaw State University, USA
†Department of Software Engineering and Game Development, Kennesaw State University, USA
‡Department of Computer Science, Kennesaw State University, USA

{ssobhan, twill690, mhossa21, jrodr225, mtasnim1, emathew9, jwrig272}@students.kennesaw.edu &
{hshahria}@kennesaw.edu

***Abstract*—The dark web is often discussed in taboo by many who are unfamiliar with the subject. However, this paper takes a dive into the skeleton of what constructs the dark web by compiling the research of published essays. The Onion Router (TOR) and other discussed browsers are specialized web browsers that provide anonymity by going through multiple servers and encrypted networks between the host and client, hiding the IP address of both ends. This provides difficulty in terms of controlling or monitoring the dark web, leading to its popularity in criminal underworlds. In this work, we provide an overview of data mining and penetration testing tools that are being widely used to crawl and collect data. We compare the tools to provide strengths and weaknesses of the tools while providing challenges of harnessing massive data from dark web using crawlers and penetration testing tools including machine learning (ML) techniques. Despite the effort to crawl dark web has progressed, there are still rooms to advance existing approaches to combat the ever-changing landscape of the dark web.**

***Index Terms*—Dark Web, TOR, Cybersecurity, Privacy, Web crawlers, Anonymity, Data Mining, Data Protection**

I. Introduction

In the rapid development of technological sophistication Dark Web architecture, security issues, and challenges are very sensitive issues. The dark web is any form of network or content that requires special software to access. This type of network is specifically created and managed to ensure the anonymity of the user. This is often confused with the term Deep web, which describes any website that cannot be directly accessed by a normal web browser. The deep web consists of more than 90 percent of the internet, while the dark web is an extremely small section of the deep web and only consists of less than five percent of the internet. One such software that is used to access much of the dark web is called The Onion Router (TOR). This is a free software and uses "onion routing" and is run by the TOR project nonprofit organization.

Currently, there are multiple research studies on Dark Web. Plenty of people use the Dark Web. To give a better picture, consider that the amount of bitcoins traded in the Dark Web was estimated at 600 million USD in 2017. [1]. In mid-2013, according to the statistics of the Federal Bureau of Investigation, the "Silk Road" has reached a transaction volume of 1.2 billion US dollars during the vigorous development stage, and has 150,000 anonymous users and more than 4,000 illegal merchants [2]. In this paper, we analyze the current state-of-the-art Dark Web technologies used in the virtual world. This survey is also important to help researchers to know the remaining challenges of the different dark web techniques used.

Contributions: Our work intends to utilize systematic survey of different papers of Dark Web and project analysis to participate in the network without fear of data theft or tampering. The implementation in this paper is transportation systematic review and evaluate the trend on dark web. The rest of this paper will be organized as follows: Section II describes research that has already been done in related fields, Section III provides background information including descriptions of related technologies and challenges that we overcome during the development of the framework described in this paper, Section IV provides detailed implementation of different research papers demonstrates a load evaluation of our implementation, and finally Section V concludes the paper and discusses future work.

II. Research Methodology

This section will describe anticipated challenges with this work and our motivation for creating and implementing this framework. A Systematic Literature Review on Dark Web technology and ride-sharing services is provided as well.

In order to carry out the study on Dark Web, we utilize a systematic literature review [3]. The main purpose of the systematic review is to identify, study, and investigate the suitable existing approaches. We first carried out a "Search Process" to identify potential research papers from the scientific databases using pre-selected search keywords or strings including "Dark Web" and "Cybersecurity". We had to identify these search strings to avoid findings from non-related research papers and those keywords are based on the Dark Web, TOR, Cybersecurity, Privacy, Web Crawlers, Anonymity, Data Mining, or Data Protection related terms and their derivatives,

978-1-6654-8810-5/22/$31.00 ©2022 IEEE
DOI 10.1109/COMPSAC54236.2022.00283

acronyms, and widely used synonyms. Besides, among various scientific databases, we used three digital database sources including (i) IEEE Xplore (ii) ScienceDirect and (iii) Springer Link. Choosing these three bibliographic databases, our goal is to identify research papers that were published in reputable conferences, journals, and books.

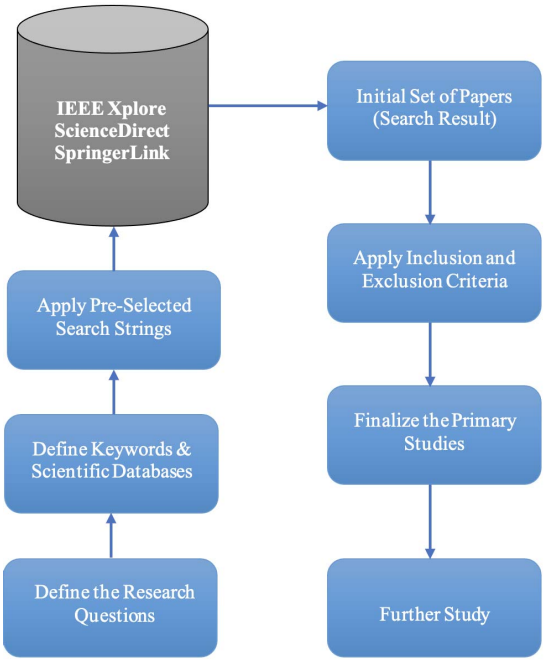

Fig. 1. Attrition of Systematic Literature thorough processing [3], [4]

We adopt the paper classification process from Kitchenham *et al.* [3] and M.J.H. Faruk *et al.* [4], [5] depicted in Fig. 1. We filtered based on time restrictions that we set between 2008 to 2022 to search studies published. We also filtered publication topics including *Systems and Data Security* for Springer Link and *Publication Topics* for IEEE Explore and *Computer Science and Security* for ScienceDirect. A total of 196 studies were found during the initial search (IEEE Xplore 49, ScienceDirect 112 and Springer Link 35). Once the search processes are completed, we have gone through a screening process for finding relevant papers based on the paper title at first followed by reading and understanding the abstract and conclusion from screened papers. In order to apply the inclusion and exclusion criteria, we set a number of exclusion criteria including (i) duplicate papers (ii) full-text availability, and (iii) papers that are not related to Dark Web shown in table I - II.

III. Literature Review

A group of researchers, mixture of Graduate and Undergraduate students from Kennesaw State University delicately analysed the different papers on Dark Web to have survey. A total of 32 papers related to dark web was surveyed. After reviewing the surveyed papers the researcher classified the papers as General Observation on Dark Web and Different

TABLE I
GENERALIZED TABLE FOR SEARCH CRITERIA

Scientific Database	Initial Search	Total Inclusion
IEEE Xplore	49	8
ScienceDirect	112	6
Springer Link	35	3
Total	196	32

TABLE II
OVERVIEW OF EXCLUSION AND INCLUSION

Condition of Exclusion and Inclusion		
Category	Condition (Inclusion)	Condition (Exclusion)
Type of Papers	Dark Web, TOR, Data mining	Other studies than aforementioned topics
Duplicate Papers	Papers are not duplicated in different databases	Similar papers in different databases
Relativity	Papers and proposed approaches are similar aspects	Studies that do not depict expected aspects
Text Availability	Studies that are available in the full format	Studies are not available fully

Technology used in this Dark Web. Fig. 2 below represents the statistics of the classified papers.

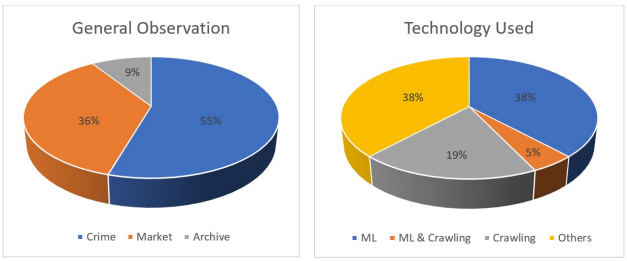

Fig. 2. Statistics of Dark Web Surveyed Papers

A. General Observation on Dark Web

In terms of all surveyed papers on dark web the papers related to general observation was 11 (34%) among 32 papers. In these observations 55% of the papers were related to crime and criminal activities in the dark web, 36% were related to dark web market where different illegal activities were highlighted and one of the paper was related to the web archiving.

The work of Basheer and Alkhatib (2021) [6] describes CTI (Cyber Threat Intelligence) and how it can be obtained on the dark web. The work contains a compilation of tools required to fully analyze and enumerate what threats exist on the dark web. Such tools have capabilities that can predict cyber threats, analyze malicious behavior and determine the main suspects or actors behind an attack, and optimize the efficiency of data collection and mining. By studying the goals, tools, and conclusions of many papers surrounding CTI, the authors were able to conclude that the dark web is one of the major sources of information and data when it comes to

analyzing cyber threats. Such analysis naturally comes with some gaps, as the work notes that existing struggles such as "analyzing the encrypted messages" that are shared between these dark web forums [6]. Until such tools are properly developed and tested, it remains excruciatingly difficult to identify malicious actors on the dark web. Fig. 3 shows the tiers of Cyber threat intelligence that are covered in the paper.

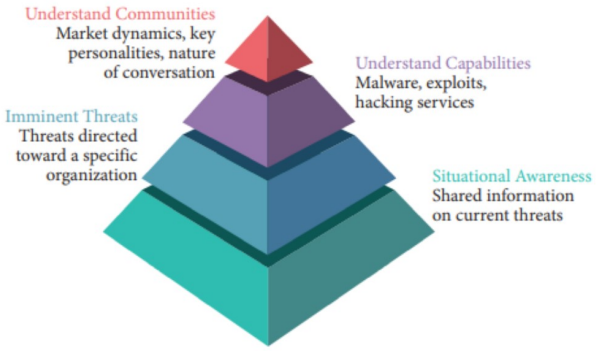

Fig. 3. The tiers of Cyber threat intelligence [6]

Godawatte *et al.* [7] mentioned that the dark web is a vast and confusing road of the internet and is host to some of the biggest illegal goodâs marketplaces that can be accessed via The Onion Router or TOR. Ranging from illegal weapon sales, child pornography, drug sales, and even hitmen, these marketplaces are widely used on the dark web and are constantly being tracked by law enforcement agencies. Tracking these vast marketplaces such as "The Silk Road" may seem easy to do on paper, its extremely difficult to keeps tabs on it because of the nature of the dark webâs URL addresses constantly changing, so the way you found to enter the site the first time might not work again. These dark web marketplaces are not limited to just individual users, as it has found that even multi-national companies use the dark web for surveillance to gain a business advantage over competitors. In addition, terrorist organizations such as ISIS use the dark web to spread their ideals and recruit new members, as wells as spread information to their followers. To keep tabs on all of these different threats across the dark web, law enforcement uses a botnet in order to categorize them into different groups. This categorization also included the IP addresses of all the people of Interest, and they were categorized as well and even led to arrest.

The work of Belshaw *et al.* [8] Dives into the education that surround the dark web, and the authors dive into the research of why the dark web should be taught about in criminal justice programs. Citing the attacks on Target, Equifax, and Capital One, the research shows an increasing frequency in cybercrimes. In 2017, a team of "cyber cops" shut down AlphaBay, the largest source of contraband on the dark web at the time. With more awareness to the dark web and how to use it, criminal justice systems can better help shut down illegal schemes on the dark web, many of which sell the data gained from large scale hacks like the three mentioned above. While only about 1.5% of users on TOR access hidden sites (where most of the illegal markets are located), that small percentage is still a large criminal network that is incredibly difficult to trace.

In the research of Koch (2019) [9] Data Leaks are becoming increasingly more frequent. These data leaks and information are being allowed to be purchased on the dark web. Credentials for illegal system access or installing malicious code in the target network, information that allows for identification, plans, strategies and more can all pose major threats too security of armed forces. The 3 Basic categories of Deanonymizing users.

- Cat 1: The first basic level that include attacks at a technical level, it is also the most dangerous but very rarely used. Attacks the flaws of the tor software and weaknesses in the design of the network protocol for tor.
- Cat 2: This attack is not based on technical characteristics but instead but exploiting indirect shortcomings, which are not based on technical vulnerabilities an example is the use of default configurations in apache servers; (Apache Java servers: Apache is responsible for accepting directory (HTTP) requests from Internet users and sending them their desired information in the form of files and Web pages.) it come with a feature called mod status which provides a website at /server-status, containing statistics like resource usage and virtual hosts. For security reasons, this page is by default only reachable from localhost. Yet if the Tor demon for onion services is running on a localhost which allows connections to the status page from external clients if the configuration is unchanged. Due to this sensitive information can be leaked. Fig. 4(uXDT attack) and its process.

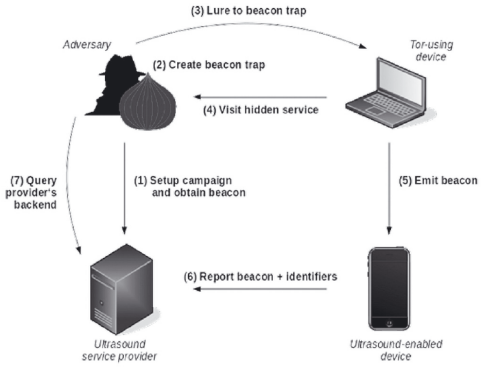

Fig. 4. Ultrasound Tracking Based Attack Scheme To Deanonymize TOR Users. [9]

- Cat 3: Based on Human error examples include providers forgetting to cover their tracks, or returning multiple times to a certain site.

B. Different Technology Used

In terms of all surveyed papers on dark web the papers related to different technology used was 21 (66%) among

32 papers. In these used technologies 38% of the papers were related to different Machine learning (ML) techniques, 19% were related to crawling techniques. The rest 38% other techniques where mentionable techniques are Vector Space Model (VSM) Technology, Risk-based security (RBS) concept, Annotated Probabilistic Temporal (APT) logic, Lexicon keys related to improvised explosive device (IED) used by the terrorist organizations, Semantic, Hyperlink, Database, and Vulnerability Assessment.

With the growing rate of online users and internet activity the concern for users' privacy and anonymity has grown exponentially. Montieri *et al.* [10] led to the creation and common use of (ATs) otherwise known as Anonymity Tools. These tools are capable of hiding a user's identity even towards the final stage, which is the web server, The traffic is forwarded and constantly encrypted and decrypted making it very difficult to trace back. A list of these Anonymity Tools (ATs) are; Tor, I2P [the Invisible Internet Project] and JonDonym also known as Java Anon Proxy. Traffic Classification (TC) is an important part of Internet traffic engineering and has applications in several fields such as network monitoring, and security it is able to classify the traffic flow according to its application. Some of the methods used in TC include Port based methods, and machine learning. In relation to this the different Classification of Algorithms are also highlighted.

- Anon 17: Anon 17 is a public dataset that specifically contains data collected from our three anonymity networks; Tor, JonDonym, and I2P, a down sample was used in order to filter out any non-informative flows. In order to view this traffic a flow exporting tool named "Tranalyzer 2" is used one of the main features and it lists the flow set as either client to server, or server to client. However it also lasts for as long as the connection is stable. Furthermore, the study also down sampled by 5% of the original listed dataset of each traffic type seat (D5) Fig. 5 below displays the results after down sampling the whole data set by 5% and shows the percentage of flows labeled with different traffic types. The first bar of the Anon17 Data set is at its full sample, the second (D5) after being down sampled by 5% and the final bar D/5) is the removal of all zero-payload flows.

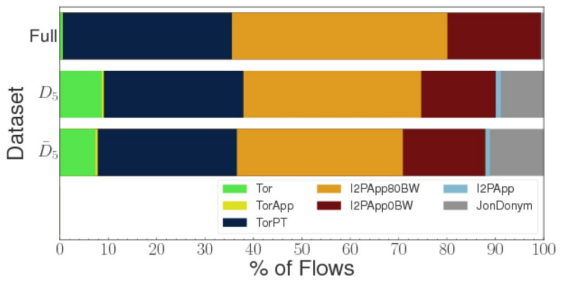

Fig. 5. Anon 17 [10]

- Naive Bayes (NB): This classifier is a simple one that assumes class conditional independence of the features and to reduce complexity. With this Multinomial Naive Bayes (MNB) adopts sample histograms as a different set of features and compares the sample histogram of each test instance with the aggregated histogram of all training instances per class.
- Bayesian Networks (BNs): BNs are graphical representations which model dependence relationships between features and classes, represented as a random set of variables and are NOT based on the conditional independence for the features. C4.5: C4.5 is an algorithm employed to generate a decision tree used (mainly) for classification purposes it is like the random forest (RF) but less complex.
- Random Forest (RF): This classification method is based on several decision trees. In the Table below shown as Fig. III we see the Best Overall Accuracy and Macro F-Measure for Dataset (D5) our down sample. The F-Measure is used to find overall accuracy, precision, and recall.

TABLE III
BEST OVERALL ACCURACY AND MACRO F-MEASURE FOR DATASET D5 OBTAINED WITH DIFFERENT (OPTIMAL NUMBER OF FEATURES) EMPLOYED

Flow-Based Classifier	Metric	L1	L2	L3
NB_SD	Accuracy	97.74%[70]	84.92%[15]	62.97%[35]
	F-measure	97.28%[70]	77.86%[20]	53.23%[30]
BN_TAN	Accuracy	99.83%[65]	95.18%[20]	71.39%[35]
	F-measure	99.81%[65]	91.57%[20]	61.31%[40]
C4.5	Accuracy	99.56%[65]	96.87%[70]	73.41%[30]
	F-measure	99.55%[65]	94.17%[65]	68.60%[74]
RF	Accuracy	99.87%[74]	96.87%[65]	73.99%[70]
	F-measure	99.87%[74]	94.06%[20]	69.05%[30]

In 2018, Marin *et al.* [11] Among the many services offered on the Dark Web, Hacking, malware and exploits that have surged in. These exploits have affected businesses and in 2015 FireEye a Cybersecurity firm reported that the same exploit was being used to steal credit card information. Identifying malicious hacking vendors and communities may reveal the patterns about the structure, organization, operation, and information flow of their corresponding networks, helping intelligence agencies target critical communities for removal or surveillance [11]. Other methods of exploits and malware being sold include Keyloggers, Phishing services, and Carding.

Nunes *et al.* [12] in the research explained that, in modern-day security policy there is a need for adequate assessment systems to predict threats. The National Institute of Science of Technology (NIST) monitors a comprehensive list of publicly disclosed vulnerabilities. If a system is at risk the National Vulnerability Database (NVD) see's a poor correlation between the Common Vulnerability Scoring System (CVSS) score. This shows that organizations are constantly looking for ways to identify vulnerabilities in systems. Common Vulnerability enumeration (CVE) is a unique identifier assigned to a system

vulnerability reported to NIST. Common platform enumeration (CPE) a list of software/hardware products that are vulnerable for a given CPE. NIST makes this data available for each vulnerability. Table. IV presents three system components of Common Platform Enumeration (CPE) [12]. Fig. 6 gives an overview of the reasoning system.

TABLE IV
SYSTEM COMPONENTS AND EXAMPLES

Components	Explanation and Examples
Platform	Can be either hardware (h), operating system (o), or application (a) based on what the vulnerability exploits.
Vendor	The owner of the vulnerable product. Examples include Google, Microsoft, The Mozilla Foundation, and the University of Oxford.
Product	The product that is vulnerable. Examples include Internet Explorer, Java Runtime Environment, Adobe Reader, and Windows 2000.

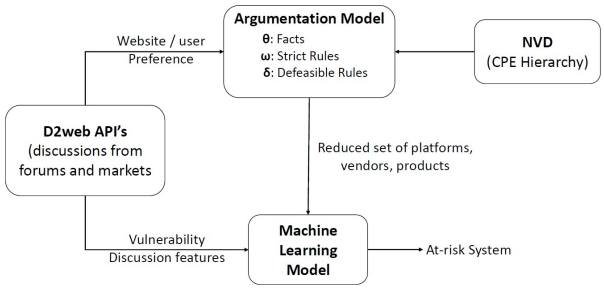

Fig. 6. Overview of the reasoning system. [12]

Chen (2008) [13] in the research explained that, When it comes to extremists terrorist groups the internet is their best friend when it comes to communicating and posting propaganda to the public. Even with a strong precense on the dark web, itâs hard to find content pertaining to these malicious groups' methods and training. However, one certain piece of content in which all security agencies are interested in IEDs which are improvised explosive content. Getting information on these said IEDs is an acquired desire challenge. The focus was on the crawling approach for the discovery collection of IED related content upon the dark web. In the work of Rahayuda & Santiari (2017) [14], The University California Berkley estimates that the dark web has been declared to have a size larger than the usual web in ML & crawling technique.

IV. DISCUSSION

This research project has allowed us to skim the only surface of the vast contents of the dark web. We want to continue to explore on how this data changed and evolved some questions are the current marketplaces in 2022 and how reliable they remain, and with new technology evolving the methods used to scrap and search these forums and archived pages can be improved with new forms of machine learning. Our next steps are to continue investigating and becoming hands on the Dark web and putting our research to the test. Developing a Web Crawler, or attending a conference to further our insight would be ideal. As mentioned in beginning of the paper for this research with all the surveyed paper related to the dark web we have classified the reviewed papers into two main categories: general observation on dark web and different technology used. In each category, we discuss the different aspects of the dark web.

In the reviewed papers as evaluated by the researchers it was found that the majority of the paper was related to different technology used to mitigate the dark web issues. The major portion of the technology used were the ML techniques and the crawling techniques. The other techniques can be mentioned as Vector Space Model (VSM) Technology [17], Risk-based security (RBS) concept [18], Annotated Probabilistic Temporal (APT) logic [19], Lexicon keys related to improvised explosive device (IED) used by the terrorist organizations [20], Semantic [21], Hyperlink [22], Database [23], and Vulnerability Assessment [24]. The other category of the reviewed paper was related to crime and criminal activities in the dark web and dark web market where different illegal activities were highlighted. One of the paper was focused on to the web archiving. To have a better idea on the overall pictorial view of the dark web reviewed papers a comparison of the surveyed paper is stated as Table. V where focus of the paper and the evaluation is mentioned.

The future directions for dark web research are varied due to the vast and complicated nature of the dark web. However, as researchers learn more and more was to crawl the dark web and extract its information from various websites, we begin to learn more and more about it. These research findings can be fundamental in the progression for future techniques and technology in order to understand the dark web even more than we already do, as well as shut down the illegal operations that inhabit it.

V. CONCLUSION

The purpose of Dark web analysis is to protect the system from various kinds or keeping internet activity anonymous and private, which can be helpful in both legal and illegal applications. The dark web is the hidden collective of internet sites only accessible by a specialized web browser. In this paper, we provided an overview of data mining and penetration testing tools that are being widely used to crawl and collect data by comparing the tools to highlight strengths and shortcomings while providing challenges of harnessing massive data from the dark web by utilizing crawlers and penetration testing tools including machine learning (ML) techniques. We also addressed what constructs the dark web by compiling the published research where we found browsers including Onion Router (TOR) are specialized web browsers that enable anonymity by routing traffic across numerous servers and encrypted networks between the host and client, hiding the IP address of both ends. Such findings indicate difficulty in terms of controlling or monitoring the dark web, leading to its popularity in criminal underworlds. Despite the effort to crawl the dark web has progressed, there are still rooms to advance

TABLE V
COMPARISON ON DARK WEB SURVEY PAPERS

Paper	Focus of Surveyed Papers	Evaluation	Year
[6]	Cyber Threat Intelligence (CTI) and detecting unethical hackers before they attack	General Observation on Crime	2021
[7]	Change in Market Places where users range form ordinary people to major co-operations and terrorist groups	General Observation on Crime	2019
[8]	Awareness on Cybersecurity	General Observation on Crime	2020
[9]	De-anonymization and Data Leaks and exposure of the methods used	General Observation on Market	2019
[15]	Expresses the layers of cybercrime undeground econmomy and gives examples to the hiearchy and what each layer means	General Observation on Market	2019
[10]	Alternatives to TOR Anonymity and the correct usage	Technology - Machine Learning	2017
[11]	Malware and Exploits being sold on the dark web and vendor similarity	Technology - Machine Learning	2018
[12]	Identification procedures based upon new dark web theories and incidents	Technology - Machine Learning	2018
[14]	Cybersecurity awareness in Onion crawlers and TOR browsers	Technology - Machine Learning and Crawling	2017
[16]	Web Crawlers and Data Mining for illicit content	Technology - Crawling	2016
[13]	IED devieces and the correlation of its impact it has on its world especially through it niche between the darkweb	Technology - Crawling	2008

existing approaches to combat the ever-changing landscape of the dark web. The new research or the future directions of dark web research are extremely varied as the technology to be able to collect data and conduct investigations are only in the early stages of development. More research can be completed into how to automate fingerprinting on the dark web, specifically for illicit activities under investigation. Another avenue of investigation would be to continually develop tools such as captcha breakers to widen the access of data mining and conduct wider searches into the dark web.

ACKNOWLEDGMENT

The work was partially supported by the U.S. National Science Foundation (NSF) award #1723578 and #2100115, and KSU Office of Undergraduate Research. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors only.

REFERENCES

[1] Ehsan Arabnezhad, Massimo La Morgia, Alessandro Mei, Eugenio Nerio Nemmi, and Julinda Stefa. A light in the dark web: Linking dark web aliases to real internet identities. In *2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2020.

[2] Hengrui Zhang and Futai Zou. A survey of the dark web and dark market research. In *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, pages 1694–1705. IEEE, 2020.

[3] Barbara Kitchenham and Stuart Charters. Guidelines for performing systematic literature reviews in software engineering. 2007.

[4] Hossain Faruk Md Jobair, Tahora Sharaban, Tasnim Masrura, and Shahriar Hossain. A systematic literature review of quantum cyber security: Opportunities and risks. 2022.

[5] Md Jobair Hossain Faruk, Santhiya Subramanian, Hossain Shahriar, Maria Valero, and Li Xia. Software engineering process and methodology in blockchain-oriented software development: A systematic study. 05 2022.

[6] Randa Basheer and Bassel Alkhatib. Threats from the dark: A review over dark web investigation research for cyber threat intelligence. *Journal of Computer Networks and Communications*, 2021, 2021.

[7] Kithmini Godawatte, Mansoor Raza, Mohsin Murtaza, and Ather Saeed. Dark web along with the dark web marketing and surveillance. In *2019 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, pages 483–485. IEEE, 2019.

[8] Scott H Belshaw, Brooke Nodeland, Lorrin Underwood, and Alexandrea Colaiuta. Teaching about the dark web in criminal justice or related programs at the community college and university levels. *Journal of Cybersecurity Education, Research and Practice*, 2019(2):5, 2020.

[10] Antonio Montieri, Domenico Ciuonzo, Giuseppe Aceto, and Antonio Pescapé. Anonymity services tor, i2p, jondonym: Classifying in the dark. In *2017 29th international teletraffic congress (ITC 29)*, volume 1, pages 81–89. IEEE, 2017.

[9] Robert Koch. Hidden in the shadow: The dark web-a growing risk for military operations? In *2019 11th International Conference on Cyber Conflict (CyCon)*, volume 900, pages 1–24. IEEE, 2019.

[11] Ericsson Marin, Mohammed Almukaynizi, Eric Nunes, and Paulo Shakarian. Community finding of malware and exploit vendors on darkweb marketplaces. In *2018 1st International Conference on Data Intelligence and Security (ICDIS)*. IEEE, 2018.

[12] Eric Nunes, Paulo Shakarian, and Gerardo I Simari. At-risk system identification via analysis of discussions on the darkweb. In *2018 APWG symposium on electronic crime research (eCrime)*. IEEE, 2018.

[13] Hsinchun Chen. Discovery of improvised explosive device content in the dark web. In *2008 IEEE International Conference on Intelligence and Security Informatics*, pages 88–93. IEEE, 2008.

[14] I Gede Surya Rahayuda and Ni Putu Linda Santiari. Crawling and cluster hidden web using crawler framework and fuzzy-knn. In *2017 5th International Conference on Cyber and IT Service Management (CITSM)*, pages 1–7. IEEE, 2017.

[15] Jonathan Lusthaus. Beneath the dark web: Excavating the layers of cybercrime's underground economy. In *2019 IEEE European symposium on security and privacy workshops (EuroS&PW)*. IEEE, 2019.

[16] Ahmed T Zulkarnine, Richard Frank, Bryan Monk, Julianna Mitchell, and Garth Davies. Surfacing collaborated networks in dark web to find illicit and criminal content. In *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, pages 109–114. IEEE, 2016.

[17] Hussein Alnabulsi and Rafiqul Islam. Identification of illegal forum activities inside the dark net. In *2018 international conference on machine learning and data engineering (iCMLDE)*. IEEE, 2018.

[18] Dimitris M Kyriazanos, Konstantinos Giorgos Thanos, and Stelios CA Thomopoulos. Automated decision making in airport checkpoints: Bias detection toward smarter security and fairness. *IEEE Security & Privacy*, 17(2):8–16, 2019.

[19] Ericsson Marin, Mohammed Almukaynizi, and Paulo Shakarian. Inductive and deductive reasoning to assist in cyber-attack prediction. In *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 0262–0268. IEEE, 2020.

[20] Hsinchun Chen. Ieds in the dark web: Lexicon expansion and genre classification. In *2009 IEEE International Conference on Intelligence and Security Informatics*, pages 173–175. IEEE, 2009.

[21] Nicolas Ferry, Thomas Hackenheimer, Francine Herrmann, and Alexandre Tourette. Methodology of dark web monitoring. In *2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pages 1–7. IEEE, 2019.

[22] Virgil Griffith, Yang Xu, and Carlo Ratti. Graph theoretic properties of the darkweb. *arXiv preprint arXiv:1704.07525*, 2017.

[23] Ying Yang, Huanhuan Yu, Lina Yang, Ming Yang, Lijuan Chen, Guichun Zhu, and Liqiang Wen. Hadoop-based dark web threat intelligence analysis framework. In *2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, pages 1088–1091. IEEE, 2019.

[24] Ch AS Murty, Harmesh Rana, Rachit Verma, Roshan Pathak, and Parag H Rughani. A review of web application security risks: Auditing and assessment of the dark web. In *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pages 1–7. IEEE, 2021.