Security of Cyber-Physical Systems Through the Lenses of the Dark Web

Ashwini Dalvi, Samata Salve, Gauri Zape, Faruk Kazi, and S. G. Bhirud

1 Introduction

Cyber-physical systems (CPSs) manifest primarily in critical infrastructures like a power grid, transportation, water network, etc., with Industrial Control System (ICS) as their integral part. Security for ICS is a daunting task because of the sheer difference between the shelf life of installed ICS system and the advancement of Information and Communication Technology (ICT).

The researchers have offered statistics on the number of incidents reported by US ICS-CERT during 2012–2016 [1], where they asserted that cyber attacks on critical infrastructures have multi-folded implications, including an economic loss to life-threatening conditions. Therefore, there is a growing need towards finding vulnerabilities, assessing possible attack vectors related to the cyber-physical systems in critical infrastructure. The approaches, like the Red Team exercise, are backed by the research community in tightening the infrastructure security from the past several years [2].

The mitigation solutions for secure CPSs are discussed in the literature for proactive, as well as reactive, security measures along with the forensic investigation. The present discussion aims to open the possibility of collecting intelligence from the dark web for investigating the cyber attacks on CPSs.

The term darknet originated in 1970 when part of the ARPANET network was isolated. This network was designed to receive the messages but not to acknowledge the received messages. In 1971, Students of MIT (Massachusetts Institute of Technology) and Stanford University exerted drug transactions through ARPANET.

Further, in pioneer work [3], the authors predicted that the Dark Net would be a future challenge. The authors in their work advocated that the peer-to-peer nature of

A. Dalvi (✉) · S. Salve · G. Zape · F. Kazi · S. G. Bhirud
Veermata Jijabai Technological Institute, Mumbai, India
e-mail: aadalvi_p19@ce.vjti.ac.in

B. Agarwal et al. (eds.), *Proceedings of International Conference on Intelligent Cyber-Physical Systems*, Algorithms for Intelligent Systems,
https://doi.org/10.1007/978-981-16-7136-4_4

the darknet would make copyright infringement of digital content unavoidable and become a deterrent in executing effective Digital Rights management.

The growth and scope of the dark web forced related stakeholders, including law enforcement agency (LEA) and the research community, to consider the illicit interactions on the dark web platform as a rich source of data, and hence various approaches are employed to access the dark web content. Since the crawling of onion sites is mostly an inexpensive task and doesn't require much manual intervention, a significant amount of unlabelled data could be collected. Hence, specific semi-supervised approaches were explored to learn from both the labelled examples and the unlabelled examples.

Researchers discussed the threat intelligence analysis framework that helps law enforcement agencies analyze crimes and criminals with the relevant information from the dark web [4]. The framework implemented to carry out this analysis is known as the Dark Web Threat Intelligence Analysis (DWTIA) Platform. The DWTIA framework implemented as the traditional network investigation method based on IP address found it challenging to trace the cybercriminal's identity, but provided access to a large volume of information, combining the surface web and dark web together. It did so by providing or using the OnionScan Dark web crawler.

Also, commercial industries offering paid services to monitor specific organization-related data on the dark web are on the rise [5–8]. One of the standard features of dark web monitoring services is searching the dark web and alerting the organization about the spread of data breaches or potential threats by curating intelligence collected from the dark web.

The presented work offers the novel objective of collecting data from the dark web to convert it into investigating leads. The outcome of the objective is presented with results obtained for the Florida water supply cyber attack.

The following sections of work are arranged as follows: section two discusses the literature review on addressing cyber attack challenges in CPS, typically in the water sector. The next part of the literature involves work on how the dark web is studied as a source of threat intelligence. Section three includes methodology followed by results and discussion.

2 Background Work

2.1 Securing Cyber-Physical Systems

The research on the cybersecurity aspect of CPSs is evolving. The work published in 2020 documented the vulnerabilities, threats, and attacks associated with CPSs [9]. The work also offered a review on measures on protection, limitation of proposed mechanisms, and future directions.

The authors mentioned the cyber attack on the Florida water supply; therefore, a brief review on cyber attacks on water management-related infrastructure is mentioned.

Researchers documented the fifteen cyber attacks on water supply infrastructure from the past few years [10]. The attack methodologies and learnt lessons from attacks were mentioned, along with mentioned trends of cyber attacks in the form of ransomware, crypto jacking, insider threats, etc., [11].

The CPSs security is discussed by referring to the scope of quantum computing, brain-like structure approach. The researchers recommend exploring the possibility of quantum cryptography to protect CPSs [12]. As with recent advancements of technologies, the CPSs will be dealing with emerging technologies: industry 4.0, Fog computing, etc., on various levels. Therefore, the trade-off between security and privacy handling through peripheral technologies and CPSs needs to be addressed diligently. The researchers proposed “Brain-Like Distributed Control Security” for the protection of CPSs [13] which was a self-autonomous protection mechanism to identify a flag raised by the intrusion detection mechanism in CPSs infrastructure [14].

The approaches mentioned in the preceding paragraph are the gist of the work research community to secure CPSs. Still, there are incidents of cyber attacks on CPSs involved in critical infrastructure. Therefore, it is always better to keep an open mind to achieve secure CPSs infrastructure.

2.2 *Dark ‘Web as an Investigative Mechanism*

The requirement of a data-driven mechanism and challenges associated with it to protect cyber-physical system is discussed in [15]. Thus, the authors of the presented work mentioned the need to extend the scope of dark web data consideration in the design of data-driven protection and mitigation mechanism.

The researcher represented an automatic crawling infrastructure termed as ZeroCrawler and a prototype called AMCL (Automated Multi-Categorization Labeling) over ZeroCrawler to identify the illicit web pages based on identified hidden themes in the ZeroNet [16]. ZeroNet is an emerging platform facilitating services to host illegal data. The online hacker forums for identifying the emerging threats in terms of popular trends and tool functionality were dissected with Diachronic Graph Embedding Framework (D-GEF) in [17]. Investigative research on image data from the dark web marketplace resulted in top vendor names, top markets, and top hash analysis results [18]. Further, a thorough evaluation of the emerging ransomware-as-a-service (RaaS) economy in the dark web, where cybercriminals or expert malicious users demand ransom or payment in return to release the infected digital assets, is presented in [19].

Thus, the above research attempts to highlight how the research community investigates the dark web data for threat intelligence or proactive measures.

Commercial solutions like DarkOwl Vision offer a proactive strategy of entity searching, monitoring, and tracking to identify and assess the threats present in the marketplace or environment to provide additional security measures if required for prevention and cybersecurity defences [8].

In summation, it is visible that CPSs driven infrastructure required considerate proactive and reactive security solutions, and dark web data is receiving interest from researchers for drawing meaningful insights. In the present work, the authors attempted to improvise an approach to gathering data from the dark web to collect all possible information related to respective cyber attacks.

3 Methodology

3.1 Crawling Mechanism

The in-house dark web crawler mechanism is depicted in Fig. 1. The dark web crawling can be initiated by providing a seed URL or Keyword(s) along with the depth of crawling. If the provided URL is already crawled and the results exist in the database, the results are extracted from the database and displayed on the web page. But, if the URL/ keyword has not been already crawled, it checks if it is the keyword or seed URL that has been provided. If it is a keyword, it uses Tordex to retrieve the results and store them in a database, whereas if a seed URL has been provided, it stores that URL in the database and takes this as the base URL.

The Tor (The Onion Router) browser is launched automatically to create a channel to crawl the dark web. The crawler checks if the number of pages visited is less than the depth provided. It then uses the Tor browser for IP rotation and pseudo-user agent generation to avoid tracking by the websites and then visits the page on the dark web while extracting the links from that page and storing them in the database. It proceeds further by crawling the stored links from the database and repeats the above process until the number of pages visited is greater than the depth provided.

Once the crawler runs until the pre-decided depth level, it proceeds to get extracted links from the database one by one to retrieve information such as title, page content, parent link, image URLs, link status, and stores them as a document in the database. Finally, these links are displayed on the crawler web page with a further option of iterative crawling. If it is not, the crawler terminates the crawling, but in case of a yes, the crawler scans whether the number of keywords crawled is less than five. If yes, it retrieves the five most occurring keywords during the previous crawl and provides the user with an option of choosing one. In case none is selected within a certain time, the crawler auto-selects the first keyword to initiate the next crawling process, and then the entire process repeats. When the number of crawls is greater than five, then it terminates the crawling. The results of the iterative crawls can be observed individually as well as a group on the crawler's dashboard. Overall, this is the crawling process for the dark web.

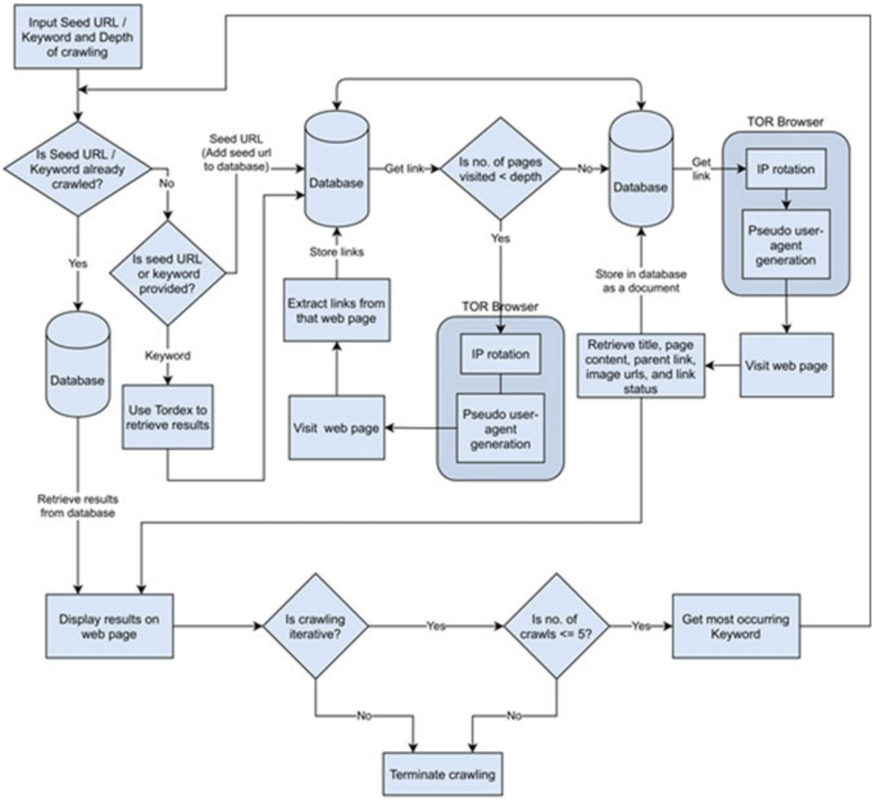

Input Seed URL / Keyword and Depth of crawling
Is Seed URL / Keyword already crawled?
No
Yes
Is seed URL or keyword provided?
Seed URL (Add seed url to database)
Keyword
Use Tordex to retrieve results
Database
Retrieve results from database
Database
Get link
Is no. of pages visited < depth
No
Yes
Store links
Extract links from that web page
Visit web page
TOR Browser
IP rotation
Pseudo user-agent generation
Database
Get link
TOR Browser
IP rotation
Pseudo user-agent generation
Store in database as a document
Retrieve title, page content, parent link, image urls, and link status
Visit web page
Display results on web page
Is crawling iterative?
Yes
Is no. of crawls <= 5?
Yes
Get most occurring Keyword
No
No
Terminate crawling

Fig. 1 Proposed dark web crawling mechanism

3.2 Discussion on Florida Water Supply Attack

The recent incident of cyber attack on the Florida water supply was reported on February 2021. In the particular attack, the attacker attempted to change the value of Sodium Hydroxide from 100 parts per million to 11,100 parts per million in the Oldsmar water treatment facility. Such increase value of Sodium Hydroxide in water is dangerous for human intake. Fortunately, due to alert human vigilance, the attack was averted.

The particular incidence motivated authors to investigate the Florida water supply cyber attack traces on the dark web. The authors ran the in-house dark web crawler with the keyword "Florida water supply," "cyber attack on water treatment plants" and "increased sodium hydroxide." The keyword "Increased sodium hydroxide" wasn't giving significant results, so the keyword was changed to "sodium hydroxide."

Table 1 presents link statistics covering the number of links crawled for entered keywords. Further, the links are compared among themselves to confirm how many links are common. For example, for the keyword Florida water supply cyber attack

Table 1 Link statistics with respect to the searched keywords

Keyword	Number of links crawled	Number of similar links w.r.t other related keywords	Number of active links during crawling	Number of inactive links during crawling	Percentage of active links
Florida water supply cyber attack	311	69 (w.r.t cyber attack on water treatment plants)	281	30	90.35%
		1 (w.r.t sodium hydroxide			
Cyber attack on water treatment plants	191	69 (w.r.t Florida water supply cyber attack)	165	26	86.39%
		2 (w.r.t sodium hydroxide)			
Sodium hydroxide	225	1 (w.r.t Florida water supply cyber attack)	183	42	81.33%
		2 (w.r.t cyber attack on water treatment plants)			

the total links collected are 311, and 69 links are common with the keyword "cyber attack on water treatment plants", and one link is common with respect to the keyword "sodium hydroxide".

The nature of the dark web is volatile. The hidden services that were active once could be inactive in the next instance. Therefore, the other two columns mention the number of active and inactive links during crawling observed over a period of one week, while the last column gives the value of active links in terms of percentage.

To comprehend the collected information without manual intervention, the Word Cloud visualization technique is employed. The dark web is famous for illicit activities; therefore, the proposed mechanism is designed so that without opening the links collected from the dark web, the essence of information is displayed with the help of Word Cloud. The Word Cloud is a popular visualization technique to represent words that frequently occur in the targeted text. The size of the word is an indicator of how many times the word appears in the text.

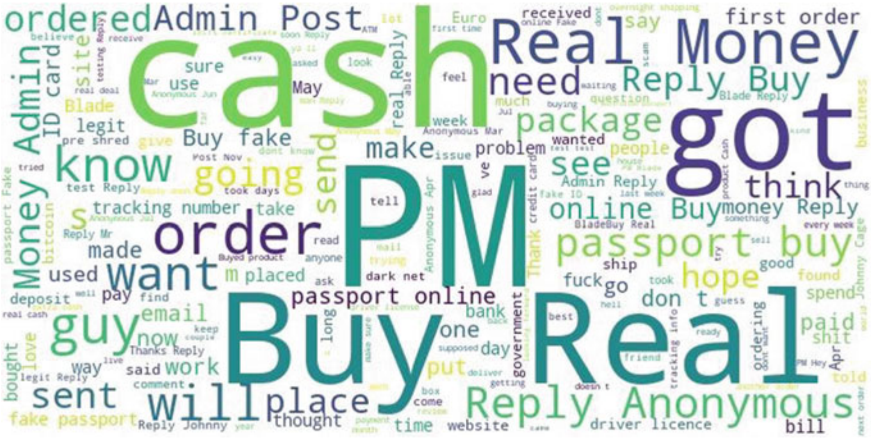


Fig. 2 Word cloud for “Florida Water Supply Cyber attack” (311 links)

Figure 2 presents the most occurring words in 311 links related to Florida water supply cyber attack. As depicted, the visibility of the words like “Buy,” “Cash,” “Money,” “order,” “get” infers that Florida water supply cyber attack is frequently discussed on the dark web marketplace. The word “PM” is also prominent on links crawled for the keyword “cyber attack on water treatment plants” as shown in Fig. 3.

The visual clues could be picked from Word Cloud to investigate ripples on the dark web regarding the cyber attacks on the surface web. On inspection of Word cloud generated for the keyword “sodium hydroxide” (225 links), depicted in Fig. 3, the word “Praveen” appears frequently but with low frequency on respective dark web pages (Fig. 4).

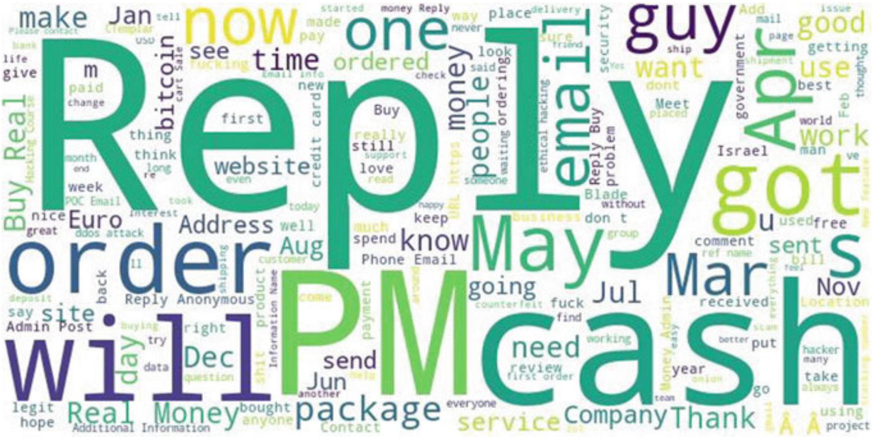


Fig. 3 Word cloud for “cyber attack on water treatment plants” (191 links)

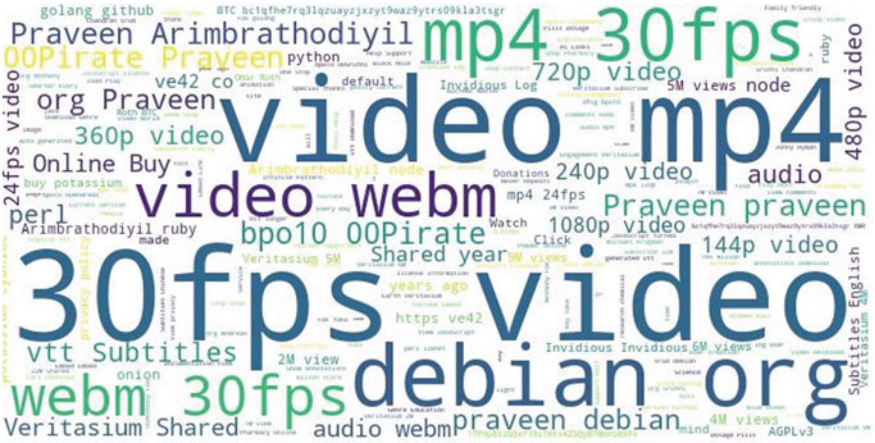


Fig. 4 Word cloud for "sodium hydroxide" (225 links)

Interestingly, the word Praveen appears in the search related to another keyword as well, "web wolf". The keyword 'web wolf 'is picked from the Word cloud of the keyword "covid-19 offers". Thus, it is apparent from the result that the Florida cyber attack is most likely discussed in the dark web forum/marketplace. Once it is confirmed that the cyber attack is discussed on hidden service, the particular links are monitored and inspected further.

3.3 Results of Some Other Malware-Related Keywords

3.3.1 DoppelPaymer in Critical Infrastructure

DoppelPaymer is a ransomware that has been evolving since 2017 and is considered similar to another ransomware, BitPaymer. In a blog post in June 2020, it was reported that DoppelPaymer ransomware groups had successfully breached the network of Digital Management Inc. (DMI), a Maryland-based company providing managed IT and cybersecurity services with NASA as one of their clients [20].

The group is alleged to operate via hacking forums where they release the compromised data while blackmailing them. In the dark web, word cloud as shown in Fig. 5 has the word "reply" occurring most times indicating that most of the pages crawled were part of a forum. The word "Drake" also occurs several times and could be a reference to a threat group called "Gold Drake" which is rumoured to be consist of operators from "Gold Heron", a group of financially motivated cybercriminals [21].

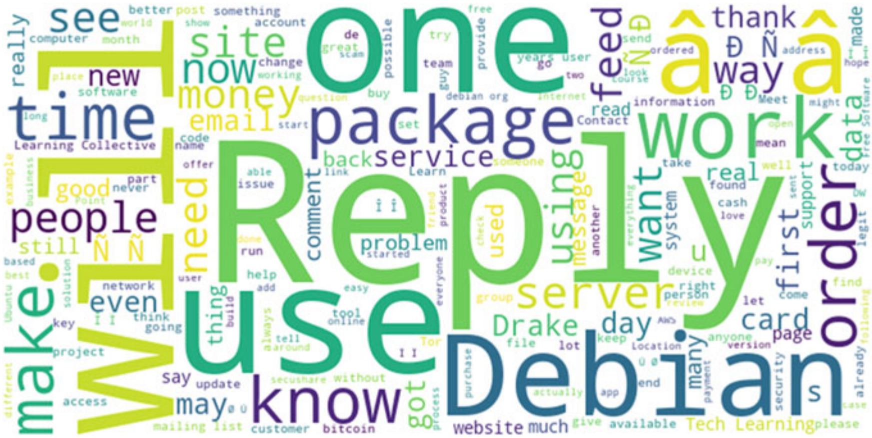


Fig. 5 Word cloud for “DopplerPaymer” (305 links)

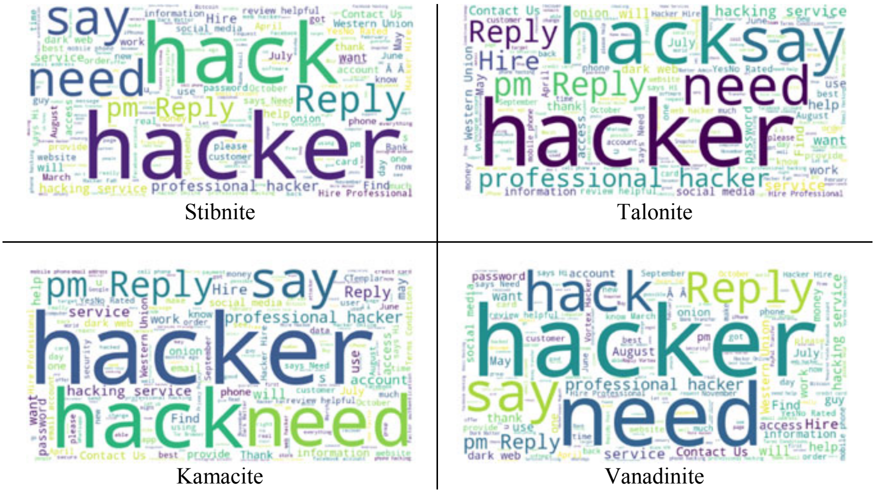
3.3.2 New Groups of Cyber Criminals

Dragos, a private security consulting group, analyzed the trends of the past fifteen years and came across four new hacking groups, namely: Stibnite, Talonite, Kamacite, and Vanadinite [22]. According to them, these groups targeted Operational Technology (OT) and Industrial Control Systems (ICS), with each of them having different target specifications.

In Table 2, the rows indicate the searched keyword with the number of links extracted mentioned in brackets, while the columns indicate the number of links that are common between the keyword in the row and the keyword in the column. The number in the bracket indicates the total percentage of the similar links in these keywords from all the links extracted from the keyword mentioned in the row. The high percentage indicates that the four groups work around the same marketplaces in the dark web, similar to their clients.

Table 2 Number and percentage of links common between each of the keywords

Keyword	Stibnite	Talonite	Kamacite	Vanadinite
Stibnite (172)	–	159 (92.44%)	159 (92.44%)	172 (100%)
Talonite (161)	159 (98.75%)	–	161 (100%)	161 (100%)
Kamacite (199)	159 (79.89%)	161 (80.90%)	–	161 (80.90%)
Vanadinite (195)	172 (88.20%)	161 (82.56%)	161 (82.56%)	–



Stibnite

Talonite

Kamacite

Vanadinite

Fig. 6 Word clouds of the four words: stibnite, Talonite, Kamacite, Vanadinite. These are four new hacker groups active on the dark web

The word cloud compilation of the four words in Fig. 6 shows that the word “hacker”, “need”, and “hire” occur quite often indicating that the groups are active on the dark web and have an array of clients that desire their services.

4 Conclusion and Future Scope

The cyber attacks on CPSs, in critical infrastructure, are severe and challenging to mitigate. The security analysts have attempted proactive and reactive security measures to secure cyber-physical infrastructure, but the attacks continue. The anatomy of cyber attacks shows that attackers get creative while crafting the cyber attacks on infrastructure. Similarly, the defence and investigation mechanism need to be customized to curb the potential cyberattack. The proposed approach has the potential to reach the attacker or get zero-day exploits if utilized constructively. The stakeholders of the proposed works are Law Enforcement Agencies and the security research community. The case study of Florida is intriguing enough to realize that engaging in dark web data result in a stream of threat intelligence.

In further development, the authors aim to map the timeline of link finding over the stipulated timeframes like immediately after mentioned attacks, consecutive one week after the attack. The objective of plotting the timeline is to comprehend the cyber attacks are discussed concerning attack methodology, exchange of exploits among dark web forums, and dark web marketplace. Also, the webpage classification will be employed to confirm the type of webpage, i.e., whether the web page is a forum, dark web marketplace, blogs, etc. Once the web page type is confirmed, then a customized investigation approach will be applied. For example, if a web page belongs to a forum, the forum thread will be investigated.

Thus, the proposed work discussed one module of planned automated and intelligent mechanism to investigate the landscape of the dark web.

References

1. Noguchi, M., Ueda, H.: An analysis of the actual status of recent cyberattacks on critical infrastructures. NEC Techn. J. Spec. Issue Cybersec. **12**(2), 19–24 (2019)
2. Brown, G., Carlyle, M., Salmeron, J., Wood, K.: Defending critical infrastructure. Interfaces **36**(6), 536–544 (2006)
3. Biddle, P., England, P., Peinado, M., Willman, B.: The darknet and the future of content distribution. In: ACM Workshop on digital rights management, vol. 6, p. 54. (2002)
4. Zhang, X., Chow, K.P.: A framework for dark Web threat intelligence analysis. In: Cyber Warfare and Terrorism: Concepts, Methodologies, Tools, and Applications, IGI Global, pp. 266–276. (2020)
5. Zerofox Dark Web Monitoring https://www.zerofox.com/products/dark-web-monitoring. Accessed 20 April 2021
6. Network Box Dark web monitoring https://www.network-box.com/nb5-darkWeb_monitoring Accessed 20 April 2021
7. Acid Cyberintelligence https://www.acid-tech.co/?page_id=66 Accessed 20 April 2021
8. Dark owl monitoring https://www.darkowl.com/ Accessed 20 April 2021
9. Yaacoub, J.P.A., Ola, S., Hassan, N.N., Nesrine, K., Ali, C., Mohamad, M.: Cyber-physical systems security: Limitations, issues and future trends. Microprocess. Microsyst. 77:103201 (2020)
10. Hassanzadeh, A., Rasekh, A., Galelli, S., Aghashahi, M., Taormina, R., Ostfeld, A., Banks, M.K.: A review of cybersecurity incidents in the water sector. J. Environ. Eng. **146**(5), 03120003 (2020)
11. Tuptuk, N., Hazell, P., Watson, J., Hailes, S.: A systematic review of the state of cyber-security in water systems. Water **13**(1), 81 (2021)
12. Tosh, D., Galindo, O., Kreinovich, V., Kosheleva, O.: Towards security of cyber-physical systems using quantum computing algorithms. In: IEEE 15th International Conference of System of Systems Engineering (SoSE) (2020)
13. Yang, H., Zhan, K., Kadoch, M., Liang, Y., Cheriet, M.: BLCS: brain-like distributed control security in cyber physical systems. IEEE Netw. **34**(3), 8–15 (2020)
14. Kholidy, H.A.: Autonomous mitigation of cyber risks in the cyber-physical systems. Futur. Gener. Comput. Syst. **115**, 171–187 (2021)
15. Jiang, Y., Yin, S., Kaynak, O.: Data-driven monitoring and safety control of industrial cyber-physical systems: basics and beyond. IEEE Access **6**, 47374–47384 (2018)
16. Ding, J., Guo, X., Chen, Z.: Big data analyses of zeronet sites for exploring the new generation darkweb. In: 3rd International Conference on Software Engineering and Information Management p. 46–52 (2020)
17. Samtani, S., Zhu, H., Chen, H.: Proactively identifying emerging hacker threats from the dark web: a diachronic graph embedding framework (D-GEF). ACM Trans. Priv. Sec. **23**(4), 1–33 (2020)
18. Jeziorowski, S., Ismail, M., Siraj, A.: Towards image-based dark vendor profiling (2020)
19. Meland, P.H., Bayoumy, Y.F.F., Sindre, G.: The ransomware-as-a-service economy within the darknet. Comp. Secur. **92**, 101762 (2020)
20. The Business of Federal Technology https://fcw.com/articles/2020/06/04/johnson-dmi-nasa-ransomware-attack.aspx Accessed 11 April 2021

21. SecureWorks Threat Profiles https://www.secureworks.com/research/threat-profiles/gold-heron%20%20 Accessed 11 April 2021
22. Dragos ICS cybersecurity year in review. Available via https://hub.dragos.com/hubfs/Year-in-Review/Dragos_2020_ICS_Cybersecurity_Year_In_Review.pdf (2020) Accessed 11 April 2021