Combining Threat Intelligence with IoT Scanning to Predict Cyber Attacks

Jubin Abhishek Soni
Researcher
Senior Member, IEEE
San Francisco, CA, USA
jubin.soni@ieee.org

Amit Anand
Researcher
Senior Member, IEEE
Austin, TX, USA
amit.anand@ieee.org

Rajesh Kumar Pandey
Researcher
Senior Member, IEEE
Seattle, WA, USA
rajesh.pandey@ieee.org

Aniket Abhishek Soni
Researcher
Senior Member, IEEE
Brooklyn, NY, USA
aniketsoni@ieee.org

***Abstract*—While the Web has become a global platform for communication; malicious actors, including hackers and hacktivist groups, often disseminate ideological content and coordinate activities through the "Dark Web"—an obscure counterpart of the conventional web. Presently, challenges such as information overload and the fragmented nature of cyber threat data impede comprehensive profiling of these actors, thereby limiting the efficacy of predictive analyses of their online activities. Concurrently, the proliferation of internet-connected devices has surpassed the global human population, with this disparity projected to widen as the Internet of Things (IoT) expands. Technical communities are actively advancing IoT-related research to address its growing societal integration. This paper proposes a novel predictive threat intelligence framework designed to systematically collect, analyze, and visualize Dark Web data to identify malicious websites and correlate this information with potential IoT vulnerabilities. The methodology integrates automated data harvesting, analytical techniques, and visual mapping tools, while also examining vulnerabilities in IoT devices to assess exploitability. By bridging gaps in cybersecurity research, this study aims to enhance predictive threat modeling and inform policy development, thereby contributing to intelligence research initiatives focused on mitigating cyber risks in an increasingly interconnected digital ecosystem.**

***Index Terms*—Internet of Things (IoT), Dark Web, IoT Vulnerability, Machine Learning in Cybersecurity, Threat Intelligence.**

I. Introduction

The Internet has evolved into a worldwide platform that allows individuals to easily distribute, exchange, and communicate ideas. While it offers numerous benefits, the misuse of the Internet has grown increasingly problematic. Malicious entities, including hackers, extremist organizations, hate groups, and racial supremacists, are leveraging the web more than ever to carry out illegal activities, facilitate internal communications, orchestrate attacks on institutions and governments, and spread harmful ideologies. It has become commonplace to hear about cyber breaches in the news, where cybercriminals post messages targeting large corporations and organizations, demanding weapons, financial backing, and volunteers [4]. Consequently, there is a pressing need to gather intelligence that enhances our understanding of the tactics employed by hackers and criminal organizations.

In this paper, we focus on the Internet of Things (IoT) as a vulnerability due to its expanding presence and susceptibility to exploitation [2]. By employing the structured methodology outlined in this study, we evaluate the degree to which cybercriminal groups target IoT systems. Through our analysis of 23 criminal websites on the Dark Web, we discovered that even seemingly harmless devices, such as small sensors, can cause substantial damage when compromised. IoT opens up an entirely new realm for cybercriminals to exploit. These hacker groups generally do not discriminate in their targets—they exploit any weakness or vulnerability they can find.

The remainder of the paper is organized as follows: Section II provides an overview of the concepts introduced in this paper; Section III outlines a methodology for collecting and analyzing Dark Web data, and identifying vulnerabilities; Section IV details how this methodology was applied in a case study of IoT using Shodan, the IoT search engine; and the final section concludes the study and explores potential solutions.

II. Literature Review

A. Dark Web Utilization

Hackers, criminals, and terrorists leverage the web to facilitate their illicit activities. Since the emergence of the Dark Web, numerous new marketplaces have surfaced, offering a wide and often alarming array of products, including prescription pills, meth, heroin, speed, crack, guns, stolen identities, gold, and erotica. The Dark Web also serves as a platform for spreading extremism, facilitating communication among terrorists and criminals, and sharing knowledge about illegal hacks and attacks, thereby enabling harmful activities that threaten society. Most of these markets operate using Bitcoin and other cryptocurrencies, which allow for anonymous online transactions. According to one study, criminals and terrorists use Bitcoin to sell drugs and subsequently purchase guns and ammunition. During our exploration of the dark market known as Hansa Market, we identified a website called Arms and Ammunition, which poses significant threats to nations both internally and externally. This underscores the importance of developing robust threat intelligence.

B. Discovering IoT Devices Through Shodan

We focused on IoT devices due to their growing integration into daily life and critical infrastructure. Like mobile phones, these devices are increasingly capable of learning

and predicting user behavior. Despite their utility, many IoT devices operate on low-power processors and custom operating systems that often lack robust encryption capabilities, making them inherently vulnerable.

For example, a smart thermostat not only regulates temperature but also stores behavioral data on centralized servers. While a compromised thermostat in a home may seem low-risk, similar vulnerabilities in industrial control systems can lead to serious disruptions. Attackers exploiting such weaknesses could manipulate systems or target individuals based on behavioral patterns, posing threats to both personal safety and public infrastructure.

To identify exposed IoT devices, we used Shodan, a search engine that scans the internet for connected devices with open ports. Shodan can discover a wide range of systems, from webcams and servers to industrial control devices and power grids. Its effectiveness has been demonstrated in prior research, such as a study in which newly connected devices were indexed by Shodan within 14 days, revealing their IPs, ports, and system metadata [1].

Shodan gathers information through service banner interrogation—extracting data provided by devices in response to connection requests. With powerful filters such as IP address, port, location, and operating system, Shodan offers both attackers and defenders a valuable tool for identifying vulnerable systems and enhancing threat intelligence.

C. Developing Cyber Threat Intelligence

Our research focused on determining whether the systematic collection and processing of Dark Web content could support the development of actionable threat intelligence to protect vulnerable IoT devices. We examined how insights from underground hacker forums, marketplaces, and other illicit platforms could enhance understanding of emerging cyber threats targeting IoT ecosystems. The remainder of this paper details the methodologies we employed for data collection, the techniques used to filter and refine relevant information, and the analytical approaches applied to transform raw data into meaningful threat intelligence. By outlining these processes, we demonstrate how Dark Web intelligence can be leveraged to strengthen IoT security measures and mitigate cyber risks before they materialize into active threats.

III. A Methodology for Collecting and Analyzing Dark Web Information

A. The Methodology

To address the diverse threats posed by the information sources used by hackers, criminals, and terrorists, we propose a comprehensive threat intelligence framework. This framework integrates Dark Web data aggregation, machine learning-driven anomaly detection of IoT vulnerability patterns, and human-AI collaboration for real-time cyber-attack prediction.

Unlike prior approaches [25], our method correlates hacker discussions (e.g., IoT exploitation trends) with Shodan-derived device vulnerabilities to prioritize high-risk targets. Figure 1 illustrates this methodology, designed to assist investigators in obtaining Dark Web intelligence through diverse sources, collection techniques, filtering, and analysis [14].

- *Information Sources* cover a wide range of hacker-related platforms, from public search engines to sites maintained by hacker groups, hacktivists, terrorists, or supporters—often requiring domain expertise for access.
- *Collection* employs systematic crawling starting from seed URLs or keywords, leveraging various web scraping tools to gather relevant IoT threat data.
- *Back-link Search* uses search engines like Google to trace web pages linking to hacker domains, helping uncover supporters who glorify or coordinate attacks.
- *Personal/People Data Search* through portals such as Yahoo! and MSN reveals profiles with links that expose hidden connections among hacker communities.
- *Meta-searching* aggregates results from multiple search engines using related keywords to refine and prioritize relevant information about groups like "Anonymous."
- *Filtering* requires domain and linguistic expertise to remove irrelevant data, considering the history, language, and symbolism of cyber-attackers.
- *Analysis* aids investigators by profiling IoT exploit motives, clustering hacker entities, mapping risk topographies, scoring threats based on hacker interest and exposure, and discovering vulnerabilities linked to specific targets.

B. Application of the Methodology

Although the Internet has been publicly accessible since the 1990s, the Dark Web has only emerged as a significant domain in recent years. Its anonymous and decentralized nature has facilitated a range of illicit activities, yet the development of effective methodologies for identifying vulnerabilities and tracking cybercriminal activities within this hidden space has remained a persistent challenge [23]. The absence of robust and scalable techniques has hindered law enforcement agencies, cybersecurity researchers, and policymakers in their efforts to combat cybercrimes effectively.

As discussed earlier, our proposed methodology integrates a combination of advanced data mining, web mining, and machine learning techniques while leveraging human domain expertise to refine their application. This hybrid approach enhances the ability to systematically identify and monitor criminal groups, analyze their discussions, and extract intelligence relevant to emerging cyber threats. By incorporating both automated data processing and human analytical oversight, our methodology balances computational efficiency with contextual accuracy, enabling more precise intelligence gathering. Ultimately, this semi-automated framework presents a scalable solution for uncovering hidden threat actors and understanding the evolving landscape of cybercrime on the Dark Web.

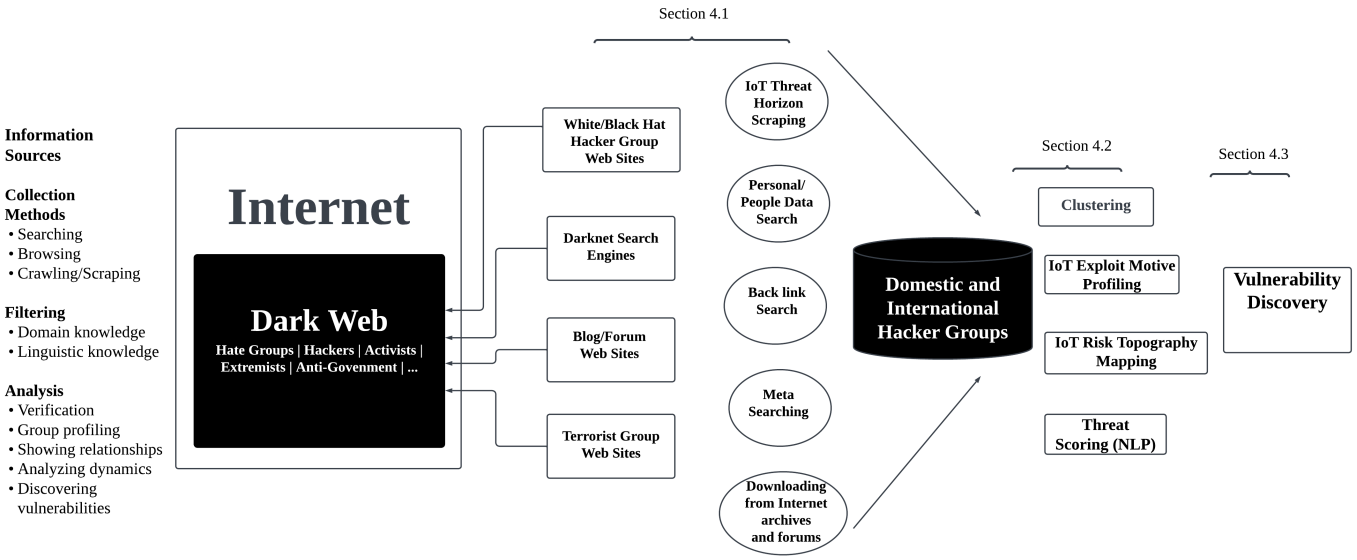


Fig. 1: A methodology for collecting and analyzing Dark Web information.

IV. Case Study: IoT Scanning to Predict Cyber Attacks

To illustrate the effectiveness and applicability of the methodology, we applied it to the collection and analysis of web-based activities by malicious actors targeting IoT devices. These devices are especially vulnerable and projected to number 26 billion by 2030. We describe our approach in Figure 1.

A. Collection

To gather data, we first identified several suspicious URLs using web searches, referencing dark web markets. We conducted web spidering using the 'Open Web Spidering' tool. For instance, starting with keywords like "freedom fighters" and "digital Robin Hood," we found four intriguing URLs: HackHound (a forum for hackers), Du Bist Anonymous (a website supporting the Anonymous group), CiHad.net (a suspicious website associated with a terror group), and Intel Exchange (a political ideology website). Using these categories as references, we uncovered additional websites during the data collection process.

We then performed the same searches on dark web search engines such as Torch, Grams, and Ahmia. Although the results were scattered, we stored them nonetheless, as they primarily pointed to dark web markets. Manually browsing these markets revealed offers of both hacking services and services required for hacking. We selected keywords such as "hacking services," "free world," "personal army," and "Internet of Things," which yielded a variety of relevant results. Overall, we collected information from hundreds of sources, and a promising future direction would be to study how different data collection techniques affect the quality and depth of subsequent analysis.

B. Filtering

We conducted two rounds of filtering to refine the dataset. In the first round, we manually excluded websites that were clearly unrelated to cybersecurity or hacking, such as news portals, gaming platforms, and adult content sites. We retained websites hosting forums, technical blogs, ideological discussions, and dark market listings offering hacking tools, data leaks, or illicit services such as ammunition. This process reduced the dataset to thirty-two websites.

In the second round, we refined the selection further. Given limitations in language coverage and domain-specific nuance, we manually removed sites that appeared redundant, linguistically inaccessible, or thematically distant from our focus. To support this process, we applied natural language processing (NLP) and basic machine learning (ML) techniques to tag security-relevant keywords and compute a threat relevance score for each website. This score was generated using TF-IDF-based feature extraction and a logistic regression model trained on a small labeled dataset comprising known malicious and benign cybersecurity forums. Websites with higher scores—indicating stronger alignment with cyber threat indicators—were prioritized for deeper analysis. This hybrid manual-automated filtering approach helped reduce subjectivity and improved consistency in identifying high-risk content.

C. Analysis

1) Classification and Usage of Websites: We conducted two rounds of filtering. First, we manually removed unrelated sites, such as news, games, and porn sites, which had no reference

to hacking. We retained websites of forums, blogs, ideological discussions, and dark markets for hacking and ammunition services. After this round, thirty-two sites remained.

Second, given our limited domain expertise and knowledge of languages, we initially manually removed some websites that felt less relevant and were similar to English forum websites in our results. Subsequently, we used natural language processing (NLP) and machine learning (ML) for keyword tagging and threat scoring.

We performed clustering and classification of the remaining 23 websites. These websites provided a diverse range of information necessary for our analysis. Table I lists these websites with hyperlinked URLs and descriptions.

Website	Description
Hansa Market	Dark market - Hacking Services
Arms and Ammunition	Guns & Ammunition
Du Bist Anonymous	Hacking services (Anonymous Hackers Site)
Rent A Hacker	Hacking services
Alpha7-Bravo- Blog	Political Hate Ideology
Thoman Paine Common Sense	Political Hate Ideology
Random United Kingdom	Hate group forum
Intel Exchange	Hacking Forum
Bugged Planet	Hate group forum
Jaggers Blog	Hacking Blog and Forum
TorChan	Hacking Discussions Forum
HackerWeb	Cyber Security Hacking Discussion Forum
HackHound	Cyber Security Hacking Discussion Forum
HackersTribe	Hacking Discussions Forum
School-of-HackNet	Hacking Discussions Forum
iC0de	Cyber Security Hacking Discussion Forum
Valhalla	Dark market - Hacking Services
Rhyliv	Cyber Security Hacking Discussion Forum
Alokbab	Suspicious Terror Group Site
Fadaian	Suspicious Terror Group Site
CiHad.net	Bogus Site – Suspicious Terror Group
Oasis	Dark market - Hacking Services

TABLE I: Websites collected with the implementation of the methodology.

We followed the following steps to analyze “IoT Hacking” or “hacking IoT devices” discussions on some of these websites:

- We filtered cybersecurity hacking discussion forums from the list above, ignoring the rest of the websites.
- We manually browsed through these forums and searched for the keywords listed.
- We gathered statistics as recent as possible (up to 2024) and generated a graph (Figure 2) and a table (Table II).

The graph in Figure 2 displays the percentage of IoT hacking-related discussions occurring in the seven forums we selected for analysis. The x-axis shows the names of these forums, while the y-axis shows the percentage of IoT hacking discussions. The ”Hacker Web” forum, which aggregates 18 major forums, has 3.3% of its posts discussing IoT hacking. Although this percentage is lower, it does not imply fewer discussions, given the forum’s size. The other forums show 12% to 30% of their topics discussing IoT hacking.

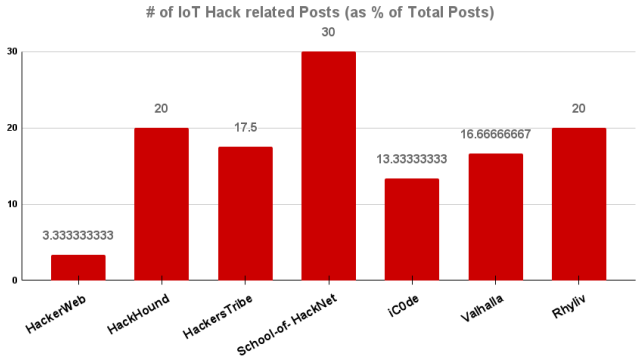


Fig. 2: Statistics of IoT device hacking discussions within December 2023 - July 2024.

Website	Searched	Found	# of Posts
HackHound	Internet of Things	Botnet, Malware, Rats for Android devices	4
Hackers Tribe	Hack Internet of Things	Sensors, File Sharing in devices	5
School-of-HackNet	Hack devices	Phones Listening to Network	1
HackerWeb	Hack Internet of Things	Knife for IoT: XMPP	1

TABLE II: This table shows the keywords searched and data found on the mentioned websites.

By manually browsing through these forums, we analyzed the most frequently discussed topics related to IoT hacking. We selected four websites listed in Table II, searched for the keywords in the ’Searched’ column, and recorded the results and total post counts in the ’Found’ and ’# of Posts’ columns, respectively. Our analysis revealed that botnets, malware, and sensors are the most discussed methods for exploiting IoT

devices. To determine how easily these vulnerabilities can be exploited, we used the IoT search engine, Shodan.

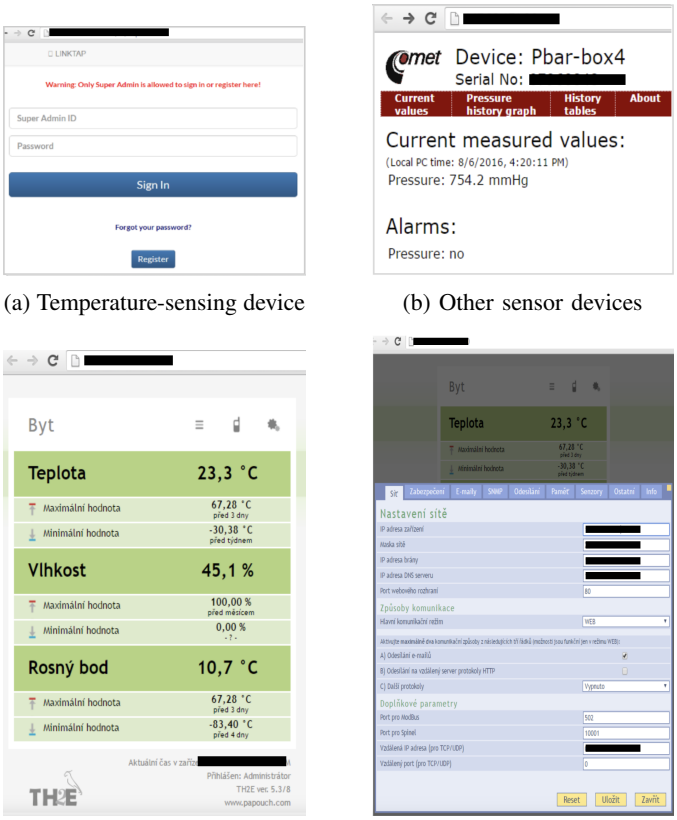
2) Shodan: Discovering Vulnerable IoT Devices: Using Shodan, we followed these steps for our analysis:

Upon executing the Python script detailed in Listing 1, we obtained 582 results for sensors. In the try block of the code, we captured the number of results and filtered information such as company name, IP address, port, operating system, and data details, including content type, authentication protocol, and location specifics like country and city. The results from our Shodan search provided comprehensive information about the devices, including company name, location, IP address, and port. This level of detail facilitates various forms of analysis.

1) Configured Shodan's REST API with Python.
2) Searched for the keyword `sensor`.
3) Exported the results in `.csv` format.
4) Used Excel for further analysis.

```python
# Importing the necessary libraries
import shodan
import csv

# Define your Shodan API key
SHODAN_API_KEY = "your_api_key"

# Initialize the Shodan API client
api = shodan.Shodan(SHODAN_API_KEY)

# Open a CSV file to store the results
with open('shodan_results.csv', 'w', newline='') as
    csvfile:
    fieldnames = ['IP', 'Data']
    writer = csv.DictWriter(csvfile, fieldnames=
        fieldnames)

    # Write the header to the CSV file
    writer.writeheader()

    try:
        # Perform a search query using the Shodan
            API
        results = api.search('sensor')

        # Print the total number of results found
        print('Results found: {}'.format(results['
            total']))

        # Loop through each result and save it to
            the CSV
        for result in results['matches']:
            print('IP: {}'.format(result['ip_str']))
            print(result['data'])

            # Write each result to the CSV file
            writer.writerow({'IP': result['ip_str'],
                'Data': result['data']})

    # Handle API-related errors
    except shodan.APIError as e:
        print('Error: {}'.format(e))
```

Listing 1: Python code used for searching in Shodan and saving results to CSV



(a) Temperature-sensing device

(b) Other sensor devices

(c) Pressure-sensor device

(d) Admin control panel

Fig. 3: Examples of vulnerable IoT devices discovered via Shodan

First, we filtered the results to include only those with port 8080 and directly accessed some of the IP addresses. As shown in Figure 3a, we encountered a page requesting login credentials. This is concerning because it exposes the device to a range of security vulnerabilities, such as cross-site scripting, denial-of-service attacks, SQL injection, unexpected application termination, and arbitrary code execution. Additionally, the connection was not secured over SSL.

Second, we manually examined the IP addresses providing direct access. As illustrated in Figures 3b and 3c, we gained access to a pressure-sensing device named 'Pbar-box4' and a temperature-sensing device named 'Byt.' We had full control over these devices, allowing us to view past temperature readings and manipulate administrative settings, including changing email addresses where data was being sent. Figure 3d demonstrates an example of the control panel for the 'Byt' device.

V. Conclusions

Analyzing and collecting Dark Web information is exceedingly challenging due to cyber-attackers' ability to conceal their identities and erase traces of their activities online. The sheer volume of web information complicates the task of obtaining a comprehensive understanding of their operations, making it even more difficult to predict their next targets. In this article, we propose a methodology to address these

challenges. By employing web mining, content analysis, and visualization techniques, the methodology leverages various information sources and analyzes 23 hacker websites for IoT vulnerabilities. Information clustering is used to differentiate between novice and experienced hackers on the web. Our detailed evaluation indicates that the methodology produced promising results, which could significantly aid in discovering new vulnerabilities and their sources.

For IoT, we propose the following solutions:

1) Strict security standards must be established for the manufacturing of IoT hardware and software.
2) IoT software should be developed by dedicated companies that rigorously adhere to IoT security protocols.
3) Best practices must be followed by users and those configuring IoT devices.

This research was conducted from multiple angles, and we believe it will contribute to the existing literature on cyber-security, potentially guiding policy-making and threat intelligence research.

VI. Declaration

Data Availability

The datasets generated and analyzed during this study are available from the authors upon reasonable request. Figures 1 (methodology flowchart), 2 (forum discussion statistics), and 4 (anonymized IoT device screenshots) are included in the manuscript, with all sensitive information redacted to prevent misuse.

Funding

The authors received no specific funding for this work.

References

[1] Bodenheim, Roland C. "Impact of the Shodan computer search engine on internet-facing industrial control system devices." No. AFIT-ENG-14-M-14. AIR FORCE INSTITUTE OF TECHNOLOGY WRIGHT-PATTERSON AFB OH GRADUATE SCHOOL OF ENGINEERING AND MANAGEMENT, 2014.

[2] Bloomberg, Jason. "7 Vulnerabilities of IoT." http://www.wired.com/insights/2014/07/7-reasons-internet-things-doomed/, 2014.

[3] Ren, Liwei. "IoT Security Problems and Challenges." http://www.snia.org/sites/default/files/DSS-Summit-2015/presentations/Liwei-Ren_Iot_Security_Problems_Challenges_revision.pdf, 2015.

[4] Grau, Alan. "Securing embedded Internet of Things." http://www.iconlabs.com/prod/internet-secure-things%E2%80%93-what-really-needed-secure-internet-things., 2016.

[5] Greenberg, Andy. "After Jeep Hack Chrysler recalls 1.4 M vehicles for bug fix." https://www.wired.com/2015/07/jeep-hack-chrysler-recalls-1-4m-vehicles-bug-fix/, 2015.

[6] Hewlett Packard. "Internet of Things Research Study." http://www8.hp.com/h20195/V2/GetPDF.aspx/4AA5-4759ENW.pdf, 2014.

[7] Bit Defender. "IoT Risks in a connected home." http://download.bitdefender.com/resources/files/News/CaseStudies/study/87/Bitdefender-2016-IoT-A4-en-EN-web.pdf, 2016.

[8] Khandelwal, Swati. "100,000 IoT devices hacked to perform cyber-attack." http://thehackernews.com/2014/01/100000-refrigerators-and-other-home.html, 2014.

[9] Zunnurhain, Kazi. "Vulnerabilities with Internet of Things." Proceedings of the International Conference on Security and Management (SAM). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2016.

[10] Barnum, Sean. "Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX™)." MITRE Corporation 11 (2012).

[11] Choo, Kim-Kwang Raymond. "The cyber threat landscape: Challenges and future research directions." Computers and Security 30.8 (2011): 719-731.

[12] Burger, Eric W., et al. "Taxonomy model for cyber threat intelligence information exchange technologies." Proceedings of the 2014 ACM Workshop on Information Sharing and Collaborative Security. ACM, 2014.

[13] Chen, Hsinchun, et al. "Uncovering the dark Web: A case study of Jihad on the Web." Journal of the American Society for Information Science and Technology 59.8 (2008): 1347-1359.

[14] Xu, Jennifer, and Hsinchun Chen. "The topology of dark networks." Communications of the ACM 51.10 (2008): 58-65.

[15] Chen, Hsinchun, and Jie Xu. "Intelligence and security informatics." Annual review of information science and technology 40 (2006): 229.

[16] Covington, Michael J., and Rush Carskadden. "Threat implications of the internet of things." Cyber Conflict (CyCon), 2013 5th International Conference on IEEE, 2013.

[17] Roman, Rodrigo, Jianying Zhou, and Javier Lopez. "On the features and challenges of security and privacy in distributed internet of things." Computer Networks 57.10 (2013): 2266-2279.

[18] Roman, Rodrigo, Jianying Zhou, and Javier Lopez. "On the features and challenges of security and privacy in distributed internet of things." Computer Networks 57.10 (2013): 2266-2279.

[19] Hayden, Michael V. "The future of things cyber." Conflict and Cooperation in Cyberspace: The Challenge to National Security (2016): 1.

[20] Chen, Hsinchun. "Exploring extremism and terrorism on the web: the dark web project." Pacific-Asia Workshop on Intelligence and Security Informatics. Springer Berlin Heidelberg, 2007.

[21] Fachkha, Claude, et al. "Investigating the dark cyberspace: Profiling, threat-based analysis and correlation." 2012 7th International Conference on Risks and Security of Internet and Systems (CRiSIS). IEEE, 2012.

[22] Provos, Niels, Moheeb Abu Rajab, and Panayiotis Mavrommatis. "Cybercrime 2.0: when the cloud turns dark." Communications of the ACM 52.4 (2009): 42-47.

[23] Abbasi, Ahmed, and Hsinchun Chen. "Affect intensity analysis of dark web forums." Intelligence and Security Informatics, 2007 IEEE. IEEE, 2007.

[24] Chen, Hsinchun. "Dark Web: Exploring and Data Mining the Dark Side of the Web." https://www.semanticscholar.org/paper/Dark-Web%3A-Exploring-and-Data-Mining-the-Dark-Side-Chen/bc0a3f270fd55ff76b8a01000b28aa5f615026d1, 2011.

[25] Palo Alto Networks. "IoT Threat Report." https://iotbusinessnews.com/download/white-papers/UNIT42-IoT-Threat-Report.pdf, 2020.