Eagle-Eye: Open-Source Intelligence Tool for IoT Devices Detection

Yusuf Al mahmeed
College of Information Technology
University of Bahrain
Sakhir, Bahrain

Wael Elmedany
College of Information Technology
University of Bahrain
Sakhir, Bahrain

Mhd Saeed Sharif
School of Architecture, Computing and Engineering
University of East London
London, England

***Abstract*—The use of Internet of Things (IoT) devices has been growing over the last few years making these appliances available in every household and organization. This significant rise of usability led to misuse, especially by non-technical people, making it an easy target for attackers to intrude on these networks. Therefore, the conventional thinking of protecting the information technology devices needs to embrace these frequent changes. Open-Source Intelligence (OSINT) is one of the modern techniques that can be used to keep track of these new systems by harvesting publicly available information. Collecting the needed information can be challenging for the IoT devices manufacturing companies and clients. This paper proposes an Open-Source Intelligence tool for IoT devices detection called Eagle-Eye which is integrated with Shodan search engine to perform OSINT queries and display it in user-friendly format. With the use of this tool companies, clients and researchers can automate their task of identifying and searching for different IoT devices statics that can be utilized and analyzed to harden these devices.**

***Index Terms*—Open-Source Intelligence Internet of Things Shodan**

I. Introduction

Today's technology is in constant evolution. The advent of the Internet of Things is often considered as a key turn of the end-user experience same as when the internet first launched. This advancement leads to open debate, whether such technological advancement is secure enough or it's winding the attack space for the attackers to use against the end-users. This development is often referred to as the digital transformation of modern society. And there are hardly any aspects of modern lifestyles that are not affected by this change in technology.

In addition to the abstract and introduction this paper will be segmented into the following sections. The "background" section will outline the background information needed to understand the main terminologies used in later sections. the "related work" section will discuss the most recent related papers and researches done covering the same topic. The "proposed Eagle-Eye tool" section that will explain the proposed solution, design and implementation of the tool. The "results and discussion" section will discuss the findings and results of running the tool to collect information about five IoT devices and explain how it can help other researchers and cybersecurity professionals to collect and analyze OSINT data related to IoT devices. Lastly The Paper will be concluded in the "conclusion" section and the references used will be available in the reference section.

II. Background

This section will explore the background of the main technologies and terminologies used in this paper, it will explain the concept of Open-Source Intelligence, along with the terminology of Internet of Things and the background of one of the most used Open-Source Intelligence search engine.

A. Open-Source Intelligence

The concept of Open-Source Intelligence can be described as the collection, search, analysis, and the use of open sources techniques and tools to gather publicly available information. This method of collecting data has emerged out of military need to gather specific information that provides advantage or knowledge. Many studies have been done since this information method came to light, developing and proposing new techniques of utilizing OSINT in different areas [1].

B. Internet of Things Security Issues

The Internet of Things is a centric concept like smart environments, self-driven vehicles, augmented reality, etc. have had an inescapable presence in the last few years. The desired objective of these appliances is to introduce plug and play experience providing the end-user with smoother and easier operation in addition to remote access control to monitor and configure [2]. The plug and play experience provided by the IoT devices allow for a potential missuse by non-technical users by setting up the devices with the default settings which lack the security hardening. This issue allows these devices to be an easy target for attackers, especially if they are exposed publicly over the internet.

C. Shodan

Shodan is a web based tool acting as a search engine dedicated for IoT appliances, gadgets and machines that are publicly available on the internet, it offers a variety of details about these publicly accessible appliances such as IP addresses, location, operating systems and more but full access to this devices information requires a paid subscription. Shodan was launched by John Matterly in 2009 and it came to be marked amongst the most popular tools available over the last few years for OSINT information gathering. Shodan.io scans the internet looking for services and devices publicly accessible on the internet to collect and store these assets

information to present it visually on https://www.Shodan.io/ or customized via Shodan Application Programming Interface (API) [3].

III. Related Work

This segment of the paper will present the most recent proposed solutions for IoT OSINT gathering tools and methods. IoT devices which are exposed over the internet are facing a lot of vulnerabilities especially the devices which are exposed on the internet. Many researchers work on utilizing OSINT search engines and customized tools that can gather information that can be analyzed to further enhance these devices' security.

A paper published by Novianto et el proposed a bash script tool integrated with Shodan using API to sweep all networks for each Autonomous System Number (ASN) in Indonesia and store it into a Comma-separated values file. The aim behind developing this tool is to supply data that can be used to profile vulnerabilities affecting these exposed internet-enabled devices in Indonesia. In addition, the information gathered using this tool can be utilized to notify the companies that manage ASN within Indonesia to be alerted of the threats and vulnerabilities in their systems [4].

Al-Alami et al published a paper presenting how can Shodan be used as a comprehensive vulnerability scanning tool for IoT appliances in Jordan. The paper aims to raise the citizen's awareness about the security problems with the wrong implementations of IoT devices and how these appliances can become extremely vulnerable in the case of misuse. The authors presented data about vulnerable devices within Jordan and advice the IoT users to be very careful and to follow security expert's recommendations such as turning off unused services [5].

Zaidi et al presented a study about the exposure of IoT devices in India by using Shodan. The authors have given a rundown of Shodan in the point of view of India. Moreover, they have analyzed several IoT tools utilizing Shodan based on different parameters within India. Furthermore the authors provided their views on how these exposed devices can be exploited by utilizing Shodan. The results of the study showed that the most exposed devices are in Mumbai followed by Delhi then Bangalore which is considered absolutely astonishing in standpoint of Bangalore as it is considered as the hub of technologies in India. Regarding the most exposed protocols, Telnet is on top while Network Basic Input/Output System (NetBIOS) and File Transfer Protocol (FTP) are in second and third place [6].

Daskevics Nikiforova proposed ShoBeVODSDT which is a tool combining Binary Edge based vulnerable open data sources and Shodan. ShoBeVODSDT can be used for non-intimidating testing of publicly available data sources to find their vulnerabilities and their extent. ShoBeVODSDT inspects a list of predetermined eight data sources to search for vulnerabilities that affect these versions. The predefined list contains data sources including PostgreSQL, Memcached, Redis, MySQL, CouchDB, MongoDB, Elasticsearch and Cassandra. This proposed solution allows for extensive analysis of exposed data sources and another perspective for the organizations to view their data sources from [7].

Daskevics Nikiforova used ShoBeVODSDT in another paper to present the most vulnerable data sources from the previously mentioned eight data sources in three Baltic countries which are Estonia, Latvia, Lithuania. Furthermore, this paper categorized these countries by which have the upmost number of open data sources, and which have the upmost number of data being accessible to external actors. The results of the paper showed that weakest results are determined by Lithuania with 3.45 out of 5 points, second weakest results indicted by Estonia with 3.18 then Latvia with 3.02 points. For the services under question, the worst results are determined by MongoDB, then by PostgreSQL, followed by ElasticSearch and Memcached [8].

Arnaert et al developed an ontology to improve the results and lower the complexity of Shodan and Censys search engines to help information technology managers to find vulnerable IoT appliances in their organizations. The authors used Stanford's ontology editor 'Protégé 2000' to develop an ontology of diagnostic and research. The authors conduct a preliminary test on the proposed ontology in order to validate its results. The results showed that the data provided by the ontology is accurate [9].

Ko et al proposed a tool integrated with Shodan and Google Map API to provide IP exposure notification system for IoT devices. The tool first utilizes python to query Shodan by using Shodan API, then it outputs the queried data into Extensible Markup Language (XML) file. After that the XML file is used as an input to create a an array storing Information of IP address, longitude, country, latitude, port number. then, the tool uses this single array to categorize the exposed IoT appliances data. In the end the tool uses the longitude and latitude attributes of that IP address to mark the information of this IP address on Google Maps [10].

Ullah and Mahmoud have presented a machine learning technique that identifies IoT appliances by examining network traffic. The authors adapted an IoT-AD-20 dataset includes 17 flow-based features from the IoT-23 dataset's pcap. Furthermore, the test conducted on the proposed technique achieved 100% accuracy, precision, recall, and F score. One of the limitations that might impact this technique of device detection is that some of these devices communicate through an encrypted channel which will impact the visibility on the network traffic [11].

Yao et al presents a method of identifying smart appliances using neural networks and web fingerprints. The authors have used web crawlers and asynchronous stateless scanning to gather target's HTTP response, the text in the gathered response data will be extracted using the natural language processing technology, meanwhile the neural networks are used to create a classification model. After refining the data, each IP response data is converted into concise text to be used as a feature vector, finally these texts are used to instruct the neural network model to understand the identification of smart appliances. One of the test models is RCNN and it achieved

the shortest converge time with reaching a test set accuracy of 90.59% and training accuracy of 98.66% making the proposed algorithm practical and has a great capability to identifying smart appliances [12].

Fagroud et al presented a study where they explore different methods of feature selection and their effects in order to achieve great precision and performance of connected devices classifications. Moreover, the authors have assessed these different methods by applying a set of machine learning models. Furthermore, in order to extract the most representative features of the proposed dataset the authors employed Recursive Feature Elimination (RFE), univariate feature selection, Tree-based feature selection (Random Forest). In addition, Random Forest, Decision Tree and XGBoost have been applied for performance evaluation based on the extracted features. The assessment results show that the selection of critical features helps enhancing the accuracy of classifying the connected devices using machine learning classifiers [13].

Soltanizadeh and Falahati, have developed Interacting Multiple-Mode (IMM) tracker and Recursive Least Squares (RLS) tracker to track IoT devices within different types of wireless communication channels. Moreover, autoregressive (AR) coefficients have been gathered theoretically for all channels between IoT devices and Base Station (BC) antennas by taking in consideration an AR model as an estimation of the channel model between the BS antenna and the IoT device. The implementation of the proposed trackers is evaluated through simulations, and the reduced sum-rate of enormous Multiple-Input Multiple-Output (MIMO) systems are shown under the impact of time-varying channels [14].

Meidan et al presented a method of identifying IoT devices accurately using machine learning algorithms to classify the devices from the network traffic data. The authors have labeled and collected network traffic data from nine in order to evaluate and train evaluate the classifier.such as distinct smartphones, personal computers and IoT devices. In addition, they have trained multi-stage meta classifier by using supervised learning in two stages. In the first stage, the classifier will distinguish between traffic generated by non-IoT and IoT devices. Meanwhile in the second stage, each discovered IoT device will be associated to a specific IoT device class. the testing results shows an overall IoT classification accuracy of 99.281% [15].

IV. Proposed Eagle-Eye Tool

This section will be divided into three subsections covering the main contribution of this study which is developing a tool written in python and integrated with Shodan which can be used to query Shodan and display the data of a selected IoT devices in a user-friendly format.

A. Proposed Solution

The common issue with the related works is that each of these papers covers a certain country or region which limits the ability of the security professionals and organizations to customize the data search in order to conduct a data analysis of their choice. Eagle-Eye is the tool which has been created to solve this issue by providing a search for the top exposed IoT devices in the world. The user can select a certain IoT device from the prescript list and query Shodan to display the top five IP address, domains, organizations, ports and countries which have these devices exposed on. In addition to that the tool is publicly available which allows for further enhancements which allow the users to customize it based on the devices that they want to obtain information about.

B. Design

The tool overall design is simple and straight forward with plenty of comments within the script in order to allow for flexibility in terms of the user requirements. The tool is written in Python. The reason for choosing this programming language is because it is the most popular and user-friendly programming language available which allows for easier modification to be done by the users. Furthermore, the tool is integrated with Shodan by using the API feature provided by Shodan and this feature can be used with the basic (free) Shodan account which is free but since the tool uses an advance search filter function it will require the users to have a paid subscription in order to query Shodan as they do not provide the advance search filter function to the basic (free) Shodan account.

Figure 1 display the flow of the application when the user runs it. As can be seen it will start by checking if the user has input a value from 1 till 6 if not then it will display an error message. For any choice from 1 till 5 chosen by the user the tool will query Shodan based on the hard coded IoT devices query, the tool will store and display the output for the chosen device and stop. If the user choose option 6 then the tool will stop running.

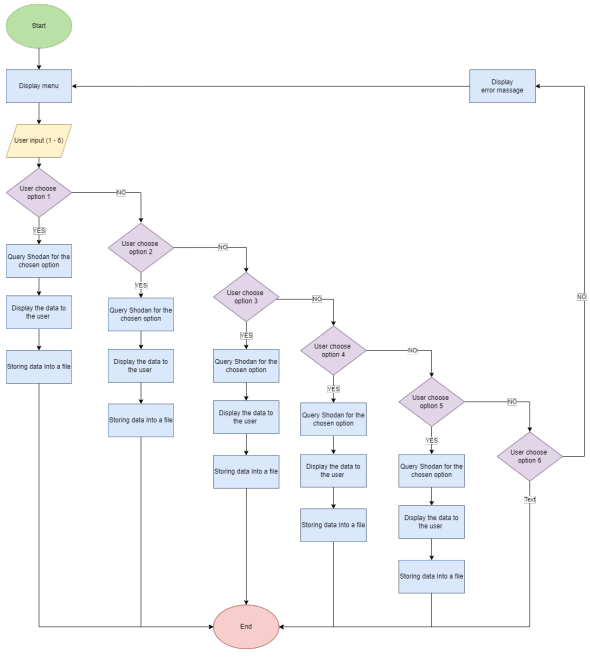

Fig. 1. Flow Chart Diagram

C. Implementation

The tool used python 3 and three modules which are sys, Shodan and pyfiglet. Sys modules were used to call two functions sys.stdout and sys.exit, the first function is used to open a file and write the query results in it. While the second function is used to stop the script when exception occurs within the try statement. Shodan module is used to call the Shodan function which is used to call the Shodan API key and allows for the integration with Shodan. Pyfiglet module is used to create the banner which displays the tool name and information when the user runs the tool.

The first section of the code consists of three parts, the first is identifying a variable which the user needs insert their Shodan API key in, second part is where the facets of properties we want summary information on are identified, the last part is providing titles for the previous listed properties.

The second section is creating the menu which provides the user with the list of devices to choose from in addition to the quit option which terminates the program. Within the second section there are subsections where the query happens for the chosen device.

Figure 2 shows case a demonstration of how to use the tool in addition to how the output will be displayed to the user in case of choosing the first option in the menu (yawcam)

Fig. 2. Tool Demonstration

V. Results and Discussion

The demonstration conducted illustrate the functionality of the tool. The demonstration provided the user with a menu containing five espoused IoT devices in Shodan, yawcam, HP Printers, Google Chromecast, Etherium Miners, and Apple AirPlay Receivers.

The results illustrated in table VI showcase the results obtained from using Eagle-Eye for five exposed IoT devices. in regard for yawcam China is leading the countries as the most country which have exposed yawcam devices within it network. Furthermore we Korea Telecom is the leading Organization in regard of exposed HP Printers. Furthermore, telecable.es is the top domain with exposed Chromecast devices. Lastly, the United States is the top country with the most exposed Etherium Miners and Apple AirPlay Receivers.

All of the above results shows the amount of devices exposed over the internet in general, this raise a sincere concern for the information security professionals as the attack surface keeps growing by the miss use of the end users.

VI. Conclusion

As of the result obtained by the demonstration we can confirm the usability and benefits of the tool where it provided accuracy and flexibility. By making the tool publicly available as an open-source project on Github

As for future work the tool will be enhanced to allow the users to insert the country that they would like to have the results about in addition to the current option which provides an overall result for the selected devices. Moreover, the tool output will be enhanced by outputting the results into an XML file in order to generate a more visually appealing and high-level report.

References

[1] J. R. G. Evangelista, R. J. Sassi, M. Romero, and D. Napolitano, "Systematic literature review to investigate the application of open source intelligence (osint) with artificial intelligence," *Journal of Applied Security Research*, vol. 16, no. 3, pp. 345–369, 2021.

[2] K. Shafique, B. A. Khawaja, F. Sabir, S. Qazi, and M. Mustaqim, "Internet of things (iot) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5g-iot scenarios," *IEEE Access*, vol. 8, pp. 23022–23040, 2020.

[3] F. Z. Fagroud, E. H. Ben Lahmar, H. Toumi, K. Achtaich, and S. El Filali, "Iot search engines: Study of data collection methods," in *Advances on Smart and Soft Computing* (F. Saeed, T. Al-Hadhrami, F. Mohammed, and E. Mohammed, eds.), (Singapore), pp. 261–272, Springer Singapore, 2021.

[4] B. Novianto, Y. Suryanto, and K. Ramli, "Vulnerability analysis of internet devices from indonesia based on exposure data in shodan," *IOP Conference Series: Materials Science and Engineering*, vol. 1115, p. 012045, mar 2021.

[5] H. Al-Alami, A. Hadi, and H. Al-Bahadili, "Vulnerability scanning of iot devices in jordan using shodan," 12 2017.

[6] N. Zaidi, H. Kaushik, D. Bablani, R. Bansal, and P. Kumar, "A study of exposure of iot devices in india: Using shodan search engine," in *Information Systems Design and Intelligent Applications* (V. Bhateja, B. L. Nguyen, N. G. Nguyen, S. C. Satapathy, and D.-N. Le, eds.), (Singapore), pp. 1044–1053, Springer Singapore, 2018.

[7] A. Daskevics and A. Nikiforova, "Shobevodsdt: Shodan and binary edge based vulnerable open data sources detection tool or what internet of things search engines know about you," in *2021 Second International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, pp. 38–45, 2021.

[8] A. Daskevics and A. Nikiforova, "Iotse-based open database vulnerability inspection in three baltic countries: Shobevodsdt sees you," in *2021 8th International Conference on Internet of Things: Systems, Management and Security (IOTSMS)*, pp. 1–8, 2021.

[9] M. Arnaert, Y. Bertrand, and K. Boudaoud, "Modeling vulnerable internet of things on shodan and censys: An ontology for cyber security," in *Proceedings of the Tenth International Conference on Emerging Security Information, Systems and Technologies (SECUREWARE 2016)*, pp. 299–302, 2016.

[10] Y.-S. Ko, I.-K. Ra, and C.-S. Kim, "A study on ip exposure notification system for iot devices using ip search engine shodan," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 10, no. 12, pp. 61–66, 2015.

[11] I. Ullah and Q. H. Mahmoud, "Network traffic flow based machine learning technique for iot device identification," in *2021 IEEE International Systems Conference (SysCon)*, pp. 1–8, 2021.

[12] L. Yao, H. Zhuang, Q. Su, Z. Lin, and J. Gu, "Automatic smart device identification based on web fingerprint and neural network," in *2021 3rd International Conference on Big-Data Service and Intelligent Computation*, BDSIC 2021, (New York, NY, USA), p. 33–41, Association for Computing Machinery, 2021.

[13] F. Z. Fagrou, H. Toumi, E. H. B. Lahmar, K. Achtaich, S. El Filali, and Y. Baddi, "Connected devices classification using feature selection with machine learning," *IAENG International Journal of Computer Science*, vol. 49, p. 445–452, May 2022.

[14] H. Soltanizadeh and A. Falahati, "On the channel tracking under uncertain state model for multiuser massive mimo in high-rate internet-of-things," *Physical Communication*, vol. 48, p. 101434, 2021.

[15] Y. Meidan, M. Bohadana, A. Shabtai, J. D. Guarnizo, M. Ochoa, N. O. Tippenhauer, and Y. Elovici, "Profiliot: A machine learning approach for iot device identification based on network traffic analysis," in *Proceedings of the Symposium on Applied Computing*, SAC '17, (New York, NY, USA), p. 506–509, Association for Computing Machinery, 2017.

TABLE I
RESULTS OBTAINED FROM EAGLE-EYE

Device	IP address	Ports	Domains	Countries	Organizations
yawcam	47.88.6.186	8888	linode.com	China	Linode
	8.216.32.94	8081	hwclouds-dns.com	United States	Aliyun Computing Co., LTD
	8.209.127.181	80	linodeusercontent.com	Germany	Huawei Public Cloud Service (Huawei Software Technologies Ltd.Co)
	8.210.59.86	52869	comcast.net	Singapore	Alibaba.com Singapore E-Commerce Private Limited
	45.33.12.251	7657	rr.com	United Kingdom	Aliyun Computing Co.LTD
HP Printers	1.212.171.229	80	sbcglobal.net	Ukraine	Korea Telecom
	1.221.51.54	443	rr.com	United States	SK Broadband Co Ltd
	12.2.188.156	8080	comcast.net	Germany	LG DACOM Corporation
	12.182.34.91	631	t-ipconnect.de	Italy	LG POWERCOMM
	14.33.139.108	8081	telenet.be	Canada	AT&T Corp
Google Chromecast	1.36.190.141	8008	telecable.es	Korea	LG POWERCOMM
	1.36.191.70		seed.net.tw	United States	Korea Telecom
	1.36.191.92		elisa-laajakaista.fi	Spain	SK Broadband Co Ltd
	1.36.230.124		tbcnet.net.tw	Taiwan	TeleCable
	1.55.174.28		netvigator.com	Sweden	DACOM-PUBNETPLUS
Etherium Miners	3.1.98.36	8545	amazonaws.com	United States	Hetzner Online GmbH
	3.1.130.126		your-server.de	Germany	DigitalOcean, LLC
	3.6.110.51		contaboserver.net	Korea	Korea Telecom
	3.13.160.121		vultrusercontent.com	France	OVH Hosting, Inc
	3.14.222.200		ovh.ca	Canada	Amazon Technologies Inc.
Apple AirPlay Receivers	2.125.4.217	5353	berkeley.edu	United States	University of California at Berkeley
	2.207.57.171		yorku.ca	Canada	York University
	5.79.176.35		uri.edu	Korea	Korea Telecom
	12.3.105.41		utk.edu	Sweden	University of Rhode Island
	12.35.42.190		wayne.edu	China	University of Tennessee