Shodan Search Engine

Paula A. Nevárez Román
Computer Science
Jeffrey Duffany, Ph.D.
Electrical & Computer Engineering and Computer Science
Polytechnic University of Puerto Rico

Abstract *— Search engines have become a crucial component of the Internet. Shodan is a unique type of search engine. Unlike conventional search engines that index the World Wide Web, Shodan offers users the ability to explore all publicly accessible devices on the Internet. Shodan explores the realm of IoT devices, servers, webcams, and more, providing valuable insights into the digital landscape. This paper delves into Shodan's functionalities, including IP scanning and other services, highlighting its applications in network security, market research, and cyber risk assessment. However, alongside its benefits, Shodan also poses significant risks due to its accessibility to anyone with internet access and basic search query knowledge. This paper emphasizes the responsible and ethical use of Shodan to prevent security vulnerabilities and privacy breaches, providing a comprehensive analysis of its features, implications, and ethical considerations in today's interconnected digital landscape.*

Key Terms *— features, risks, search engine, and Shodan*

INTRODUCTION

In the constantly evolving world of the internet, search engines have played a crucial role in the life of individuals. Used daily by individuals for personal and work-related tasks, search engines have become one of the most important tools. A search engine is a software program that allows users to search and retrieve data from the World Wide Web (WWW) using a user's query. [1] They can be used to find specific information, websites, images, videos, addresses, businesses, and other type of data. Most search engines work in three phases. Crawling consists of scanning the web for public information and creating a list of them. In this phase, the contents get analyzed. Indexing involves organizing, sorting and storing information that may include the title and description of a page. The last phase is ranking. This phase consists of analyzing a user's query, finding matches and presenting the results obtained. [2]

A year after the World Wide Web was invented, the first search engine was launched in 1990. Archie used downloadable files, making it's used limited to listings and not the content. Then, in 1994 launched Yahoo! who provided the first collection of web pages and WebCrawler the first search engine to index entire pages. AskJeeves was launched in 1996. It is the first search engine to have human editors that responded to user queries. [3] All of these search engines where just the beginning of what we have nowadays. Then, other common search engines were launched and improved over the years. This includes Google, Bing, MSN, and many others.

Shodan is a search engine that collects information on all the devices that are directly connected to the Internet. This search engine queries devices for different types of publicly available information if they are directly linked to the Internet [4]. The kind of devices that are indexed can range from small desktops to nuclear power plants, including everything in between them. There are many differences between common search engines like Google and Shodan. The main difference is that Shodan crawls the Internet, while this common search engines crawl the World Wide Web. The devices that support the World Wide Web are just a small part of the amount of devices that are connected to the Internet. This is why Shodan's main purpose is to present a complete view of the Internet. Another main difference is that Shodan requires the

understanding of the syntax for search queries, while this common search engines do not. This means that if you entered camera on Shodan you will not receive the expected results. [5]

BACKGROUND

Shodan was created by the computer scientist John Matherly as part of a hobby. In 2003, Matherly conceived the idea of tracking the devices connected to the Internet. Shodan was officially launched in 2009. The name of Shodan is a reference to a character from the video game series System Shock. The character SHODAN stands for Sentient Hyper-Optimized Data Access Network [6]. Shodan collects data on all the devices that are directly connected to the Internet and queries them for specific publicly-available information. This information is mostly recovered from banners, which can be defined as metadata about a software that's active on a device. It may include simple information such as a welcome message or more detailed such as what options are supported by the service, the server software, or any other type of information that the user would like to know before interacting with the server. Shodan uses specific ports to collect data from different devices: including the web servers or HTTP/HTTPS (ports 80, 8080, 443, 8443), as well as FTP (port 21), SSH (port 22), Telnet (port 23), SNMP (port 161), IMAP (ports 143, 993), SMTP (port 25), SIP (port 5060), and Real Time Streaming Protocol or RTSP (port 554).[6] Some examples of devices that use these protocols are routers, webcams, servers, and other systems Shodan has the ability to identify services running within these devices, open ports, and unsecured devices. This will help the user identify possible vulnerabilities on a system.

Shodan's Search Query Fundamentals

Shodan requires that a user understands their search query syntax to be able to use it. This includes basic guidelines such as their banners and search filters.

Banners: Shodan stores the information for each service in banners. This is what a user will be searching for when using this search engine. It is classified by properties that store different kinds of information in relation to a service. The data property is the only one that Shodan will search for. The other properties can be searched for by using filters. Banners may vary depending on what information is being searched for. [7] An example of a banner is shown in Figure 1.

Figure 1
Example of a Banner

Search Filters: Shodan implements the use of search filters by using unique keywords to search for specific properties. These filters take the format of filtername:value, without using any spaces. If the value contains any space, the value must be wrapped in quotes. There are no limitations to the amount of filters applied, meaning you can use multiple filters in one search. [7] In Figure 2, the highlighted section of the search query are filters. The city filter and the organization filter are applied to the search for camera.

Figure 2
Example of Search Filters

Search Syntax: The syntax is based on the banners and the search filters. As mentioned before Shodan only searches the data property by default. This means that any added details should be searched using filters. [7] In Figure 3, the composition of the search syntax is shown. The *camera* is the data property that is being searched for. There are two filters being applied. The *city: "Mayagüez*" section indicates the the search should be filtered to those cameras in Mayagüez city. At last, *org:Liberty* indicates that the organization that owns the data should be Liberty.

Figure 3
Example of a Search Query

Shodan's Features

Shodan provides many features depending on the type of account the user has. There is a free account that has very limited resources, but they do have many memberships. This project has been executed using the Academic Upgrade. [8] The features are the following:

- *Query credits* refer to the amount of search requests within Shodan that are allowed monthly. Depending on the plan they can range from 100 to Unlimited.
- *Scan credits* refer to the amount of IP scans allowed monthly. Depending on the plan they can range from 100 to Unlimited.
- *Monitored IPs* refers to the amount of IPs that can be monitored. Depending on the plan they can range from 16 to Unlimited.
- *Search Filters* refers to which filters are allowed to be used, the only ones limited are **vuln** and **tag**. Depending on the plan they are limited or unlimited.
- *Number of Users* refer to the amount of users that are allowed access to an account. Depending on the plan they are 1 or a custom amount.
- *Shodan Search pages* is the amount of search pages allowed for an account. Depending on the plan they can range from 20 to 200.
- *Shodan Monitor* refer to a tool that can monitor a specific network range. They are included if you have any type of membership
- *Shodan Trends* provides historical data for users on the system. They are included if you have any type of membership.
- *Private Firehose* provides users with access to a continuous stream of real-time data from Shodan's global network scans. They are included if you have any type of membership.
- *IP lookups* allows users to gather detailed information about a specific IP address including the owner of the address and any open ports. They are included if you have any type of membership.
- *Batch IP lookups* allows users to efficiently gather information about multiple IP addresses simultaneously. Depending on the plan they are included or not.
- *Bulk Data* allows users to access bulk data collected by Shodan's scanning activities. Depending on the plan they are included or not.
- *InternetDB* provides users with a comprehensive database about devices connected to the internet. Depending on the plan they are included or not.
- *Full firehose* provides users with real-time access to all the data indexed by Shodan. Depending on the plan they are included or not.
- *Internet scanning API* allows users to programmatically access Shodan's vast database of IoT device information. Depending on the plan they are included or not.
- *Hostnames scan* allows users to discover and enumerate hostnames associated with IP addresses. Depending on the plan they are included or not. [9]

PROBLEM

The services provided by Shodan can applied to multiple areas. In Network Security, it will help the user keep an eye on all devices at your company that are facing the Internet. It can also result beneficial in Market Research by helping the user find out which products people are being used in the real-world. Another area is Cyber Risk, that includes the online exposure of your vendors as a risk metric. On the Internet of Things, Shodan can help track the growing usage of smart devices. In Tracking Ransomware, it can help measure how many devices have been impacted by ransomware. [10] Even if Shodan presents many benefits it can also result in a great risk to the security of companies and individuals. This is because it is publicly available to anyone who has all the resources to use it. In some states in the US the use of Shodan with respect to devices that the user does not own is a felony crime. [11] But, even with having this in mind it does not mean that a user will not use Shodan for a malicious act.

EQUIPMENT AND MATERIALS

The equipment and materials used in this paper are a computer, access to the internet, and a windows system. In this project I used Windows 10. This project will also require access to the Shodan Software, focusing on the free features that are available to users with an account.

METHODOLOGY

The methodology for this paper is divided in three phases. Starting with an extensive literature review and background research phase. This entailed delving into academic journals, industry reports, technical documentation, and trusted internet sources to properly establish a strong foundational understanding of Shodan. Through the examination of the existing literature on historical development, technical work, and practical applications of Shodan, I was able to gain valuable insights into its functionalities, and potential implications and risks it may have.

In order to apply the information learned during the literature review, a hands-on approach was employed to gain firsthand knowledge into Shodan's capabilities. This involved the creation of a Shodan account and then, engaging actively by using the platform to conduct searches, analyze search results, and explore various filters and functionalities available within the tool. Through this technical search and testing, we were able to gain a deeper understanding of Shodan's user interface, search syntax, and their CLI, enhancing the accuracy and reliability of the analysis provided.

Finally, the last phase consists of a critical analysis of the ethical and legal frameworks concerning the use of Shodan. This analysis includes examining privacy laws, data protection regulations, ethical guidelines for cybersecurity research, and industry standards related to the responsible use of information gathering tools like Shodan. By considering the implications that may result of its improper use, this paper highlights the importance of ethical conduct, data privacy protection, and legal compliance in the context of using tools that have potential privacy and security implications to individuals and companies.

RESULTS AND DISCUSSION

The results obtained from using Shodan's search engine provide valuable insights into the digital landscape and the state of internet-connected devices. Through targeted queries and analysis of banners, Shodan offers a comprehensive view of various services, devices, and vulnerabilities present on the internet. One notable result is the ability to identify exposed IoT devices, such as webcams, routers, and servers, along with their respective configurations and potential security risks.

In this project, the search feature of Shodan was tested. I wanted to search if any of the systems in my household were compromised which is why

the search was based on the cameras exposed on the Internet that were located in Puerto Rico and the organization was Liberty. To perform this search within the Internet, I started by making a comparison between the results obtained of two simple searches that will look for banners that have “camera” and “webcam” in their data property. Both of these searches were done without using any specific filters. In Figure 4 you will be able to see the results obtained from the search for “webcam”.

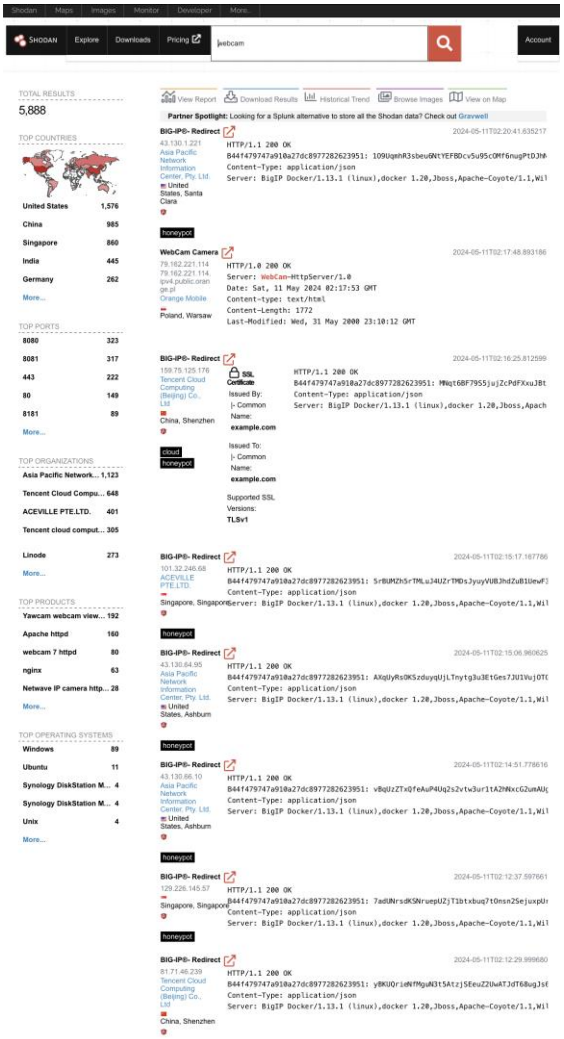

Figure 4
Search Results for “Webcam”

In Figure 5 you can see the results obtained from the search for “camera”. Both of these searches include information like the amount of results obtained, the amount on each of the top countries, ports, organizations, products, and operating systems. Also, they provide access to reports, results, historical trends, images and maps. All the results are not shown because of the amount of results obtained. After the initial searches, I opted to focus on the camera search because of the difference in the amount of results obtained.

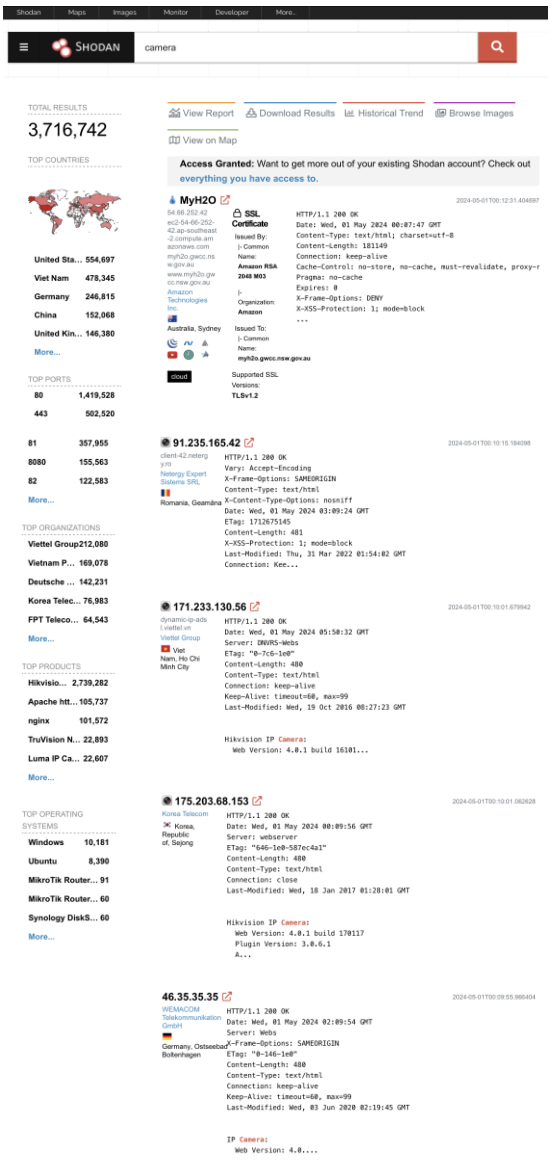

Figure 5
Search Results for “Camera”

In Figure 6, the report is shown. This report includes the information including the name and amount of identified devices of each item in each category. These categories are: country, open ports fount, organization vulnerabilities, products, tags, operating systems. It also provides HTTP Insights

including the website titles, the web technologies and the protocol versions. Additionally, it includes SSL Insights like SSL/TLS Versions, JARM Fingerprints, and JA3S Fingerprints.

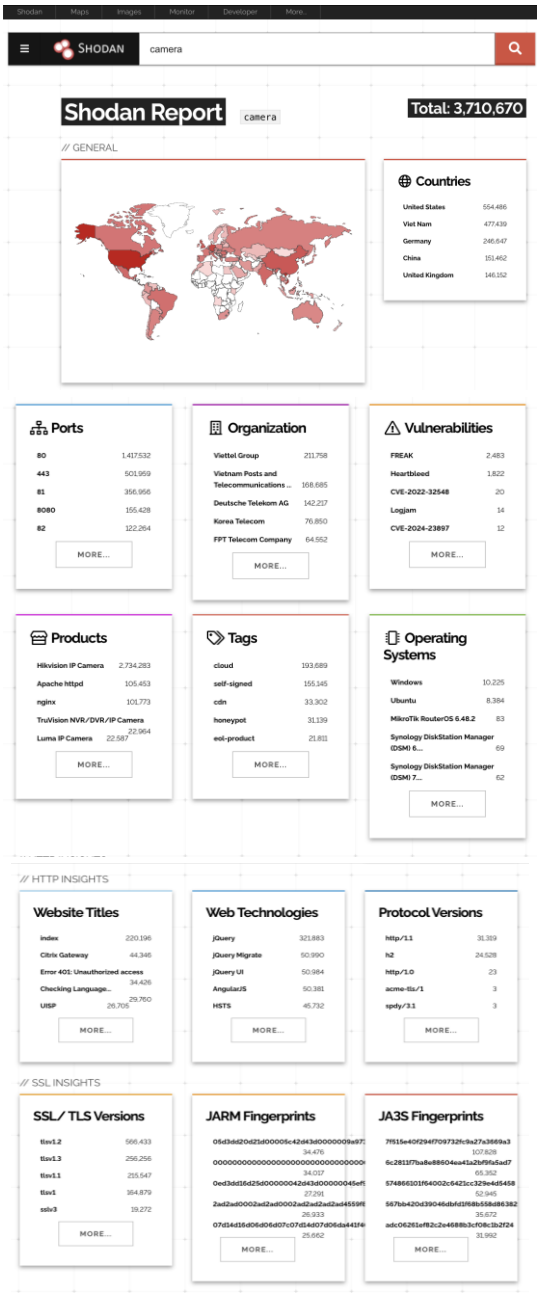

Figure 6
Shodan Report for "Camera" Search

In Figure 7, the historical trends provided by Shodan is shown. This section shows a time line going back to 2017. It includes an analysis of the amount of results obtained during that period of time. This analysis shows how much the internet has grown and how security has increased. The world map section provides a similar analysis but instead of using graphics it is more focused on the areas of the world where numbers have increased or decreased.

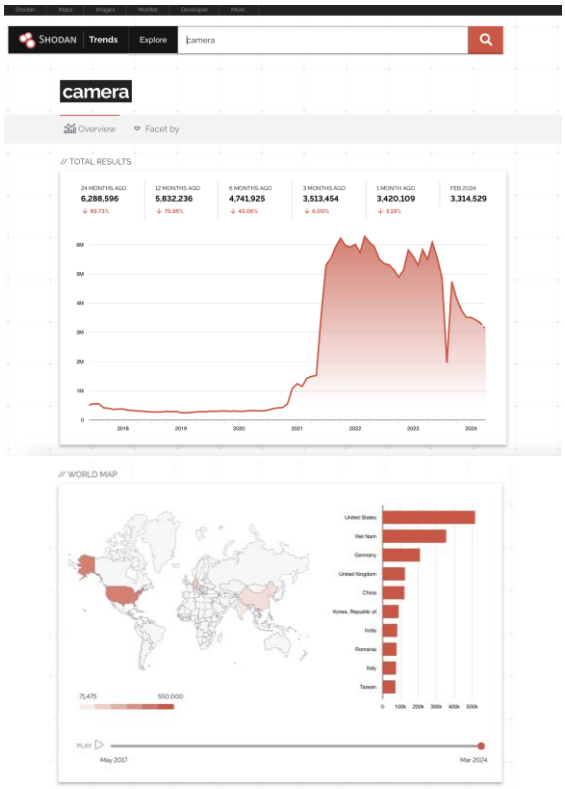

Figure 7
Historical Trends for "Camera" Search

In Figure 8, the results of opening the browse images tab are shown. When you click on image you can access all the data of the camera feed that is in fact an image of the camera feed.

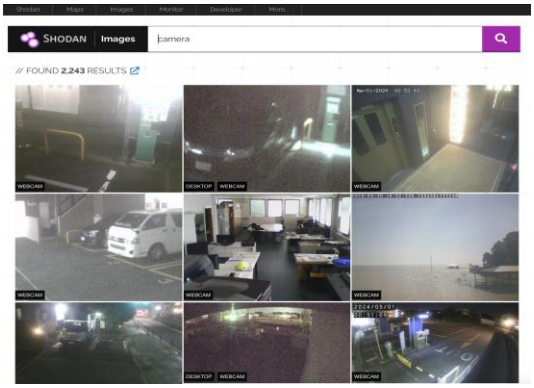

Figure 8
Images for "Camera" Search

In Figure 9, are the screenshots of the results including a regular view and the raw data which is no other than the general information written in code. This includes hostnames, domains, country, city, organization, ISP, and ASN. It also provides all the open ports and detailed information on them.

In Figure 10, the map of the search results are shown. This includes the total amount of results obtained, and top services, countries, and organizations. The red dots show where each result is located.

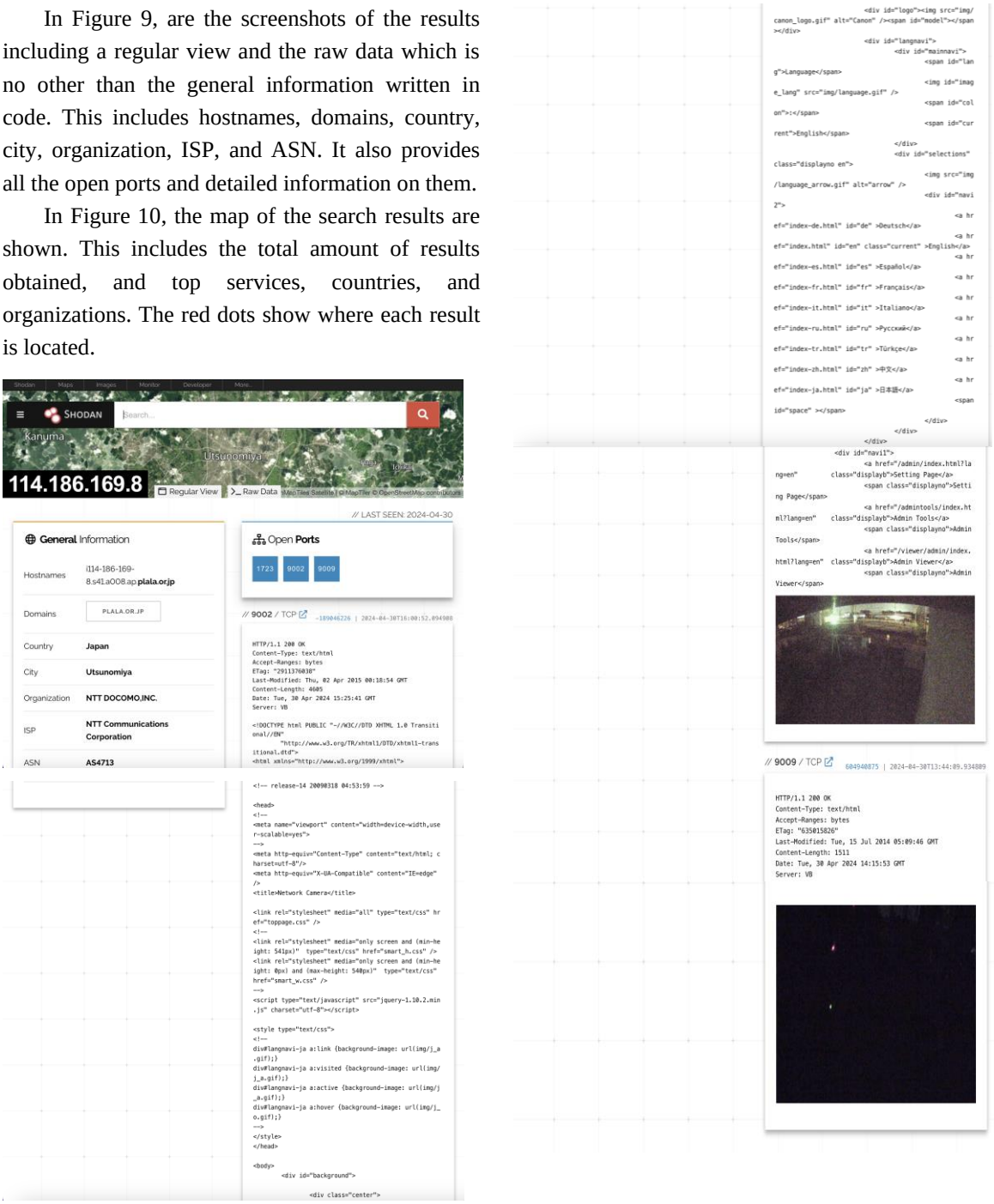

Figure 9
Image Data for "Camera" Search

Once the first search was done and I had analyzed all the results. The main search was performed using filters to see if any of the devices of my household were found. The search was based on the cameras exposed in the Internet that were located in Puerto Rico and the owner was Liberty. In this case the filters used were country and organization. Figure 11 shows the results obtained. In this specific search there were no images obtained which is why the browse images section does not appear.

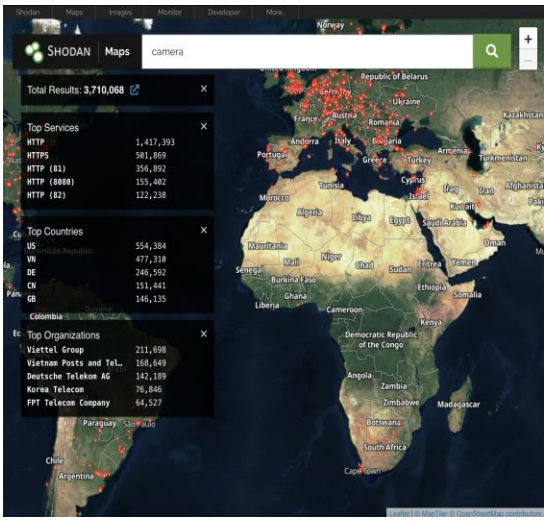

Figure 10
Map for "Camera" Search

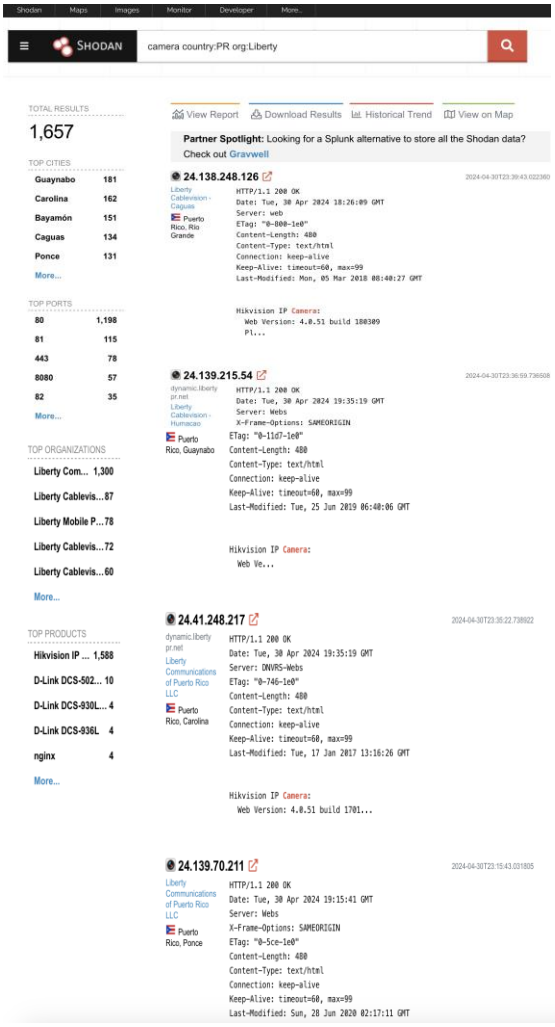

Figure 11
Search Results for "Camera" with Filters

In Figure 12, the results of the report are shown with detailed information on the device. Figure 13 shows the historical trends for the search and how the security has increased.

In Figure 14, the map of the search results are shown. These results include the total amount of results, the top services and to organizations with their details.

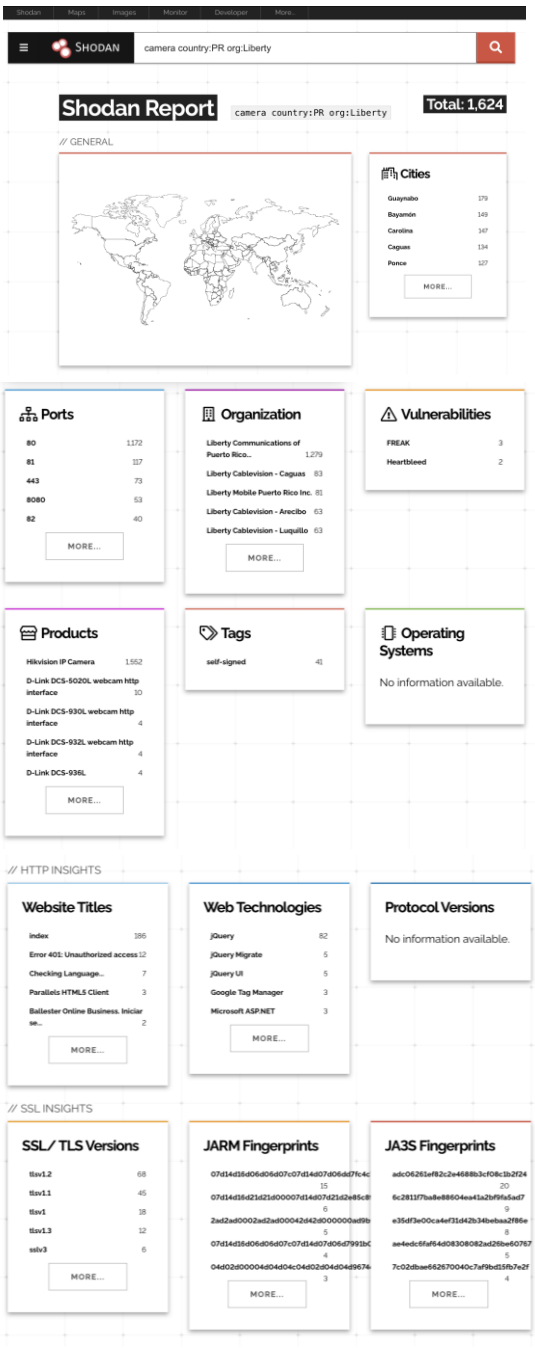

Figure 12
Report for "Camera" Search with Filters

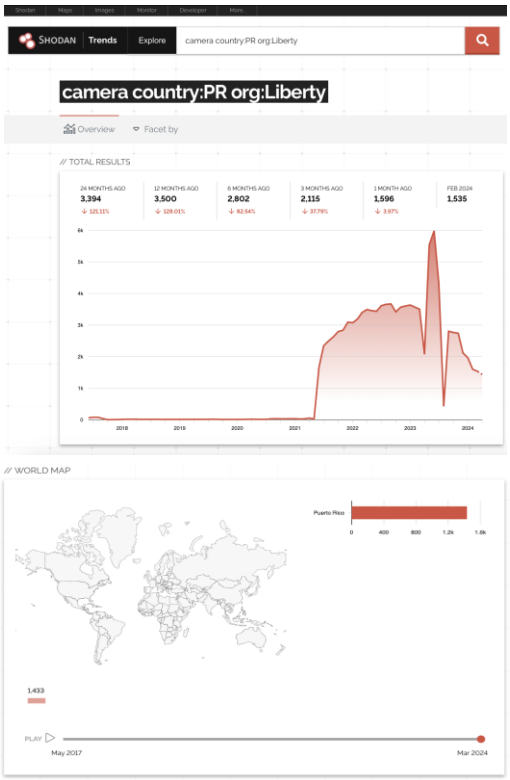

Figure 13
Historical Trends for "Camera" Search with Filters

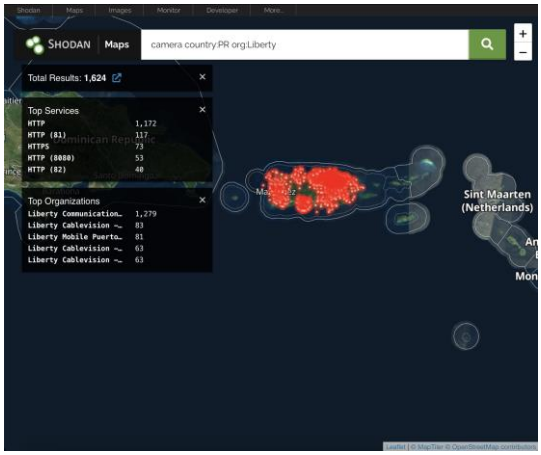

Figure 14
Map for "camera" Search with Filters

Additionally, when I applied the filters, I noticed that the Browse Images tab did not appear. Which means that according to the search there were no camera feeds being shown. After searching through the results obtained and analyzing all the data, I was able to arrive to the conclusion that the camera's in my household were not exposed in the Shodan search engine.

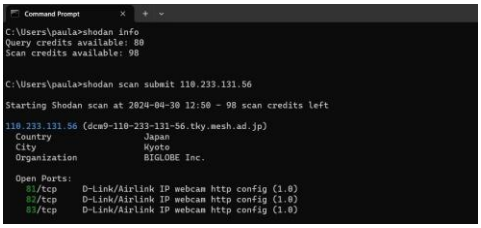

Figure 15
IP Lookup Results for 110.233.131.56

Then, I decided to test Shodan's feature of searching IP addresses. This feature allows the user to access information like the country, city and organization of that IP address. It also allows the user to see any open ports it has. I started by searching for IP's that I had already seen in the searches performed before. In figure 16, the IP scan feature is tested using the following IP address "110.233.131.56". Figure 16 shows the results of the IP address lookup found in figure 9. The results obtained from this search do not match the results obtained before. This is because figure 9 shows that the IP address "114.186.169.8" has three open ports, those being "1723", "9002", and "9009". While figure 16 shows only one port which is "9002". The last search I performed using this tool was for my IP address. This search was not successful because they were not able to search for a private IP. The results of this search are not be provided for privacy reasons.

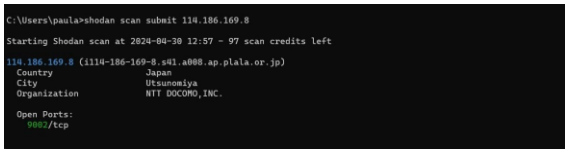

Figure 16
IP Lookup Results for 114.186.169.8

CONCLUSION

In conclusion, Shodan provides their users a glimpse into the interconnected world of Internet of Things (IoT) devices by offering unique insights and capabilities. Its distinctive features, such as the ability to search for specific devices, ports, vulnerabilities, and IP addresses makes it an extremely powerful tool for cybersecurity professionals, researchers, and individuals alike.

Shodan contributes to a better understanding of the complex nature and potential risks associated with our interconnected devices by allowing users access to information that may not be easily accessed through traditional search engines.

However, in addition to its amazing features, Shodan also presents major concerns surrounding security, privacy, and ethical standards. Shodan's IoT device indexing approach allows their users to uncover sensitive information about individuals, companies, and infrastructure. This presents a serious risk, especially when it is being misused or exploited by malicious actors for unauthorized access, data breaches, or surveillance purposes. For this reason, the responsible use of Shodan and similar tools requires strict adherence to ethical guidelines, data protection measures, and cybersecurity best practices.

The constantly evolving world of IoT and interconnected devices emphasize the importance of creating a balance between innovation and security. While Shodan offers extremely valuable insights into the state of our interconnected world, its potential risks cannot be ignored. Therefore, as we continue to leverage advanced technologies like Shodan, it is imperative to prioritize cybersecurity awareness, education, and proactive measures to safeguard individuals, companies, and critical infrastructure from potential vulnerabilities and threats in the digital age.

FUTURE WORK

While this paper on Shodan provides valuable insight on how the search engine works and some basic features, there are still multiple features that may be evaluated in future work. This future work could focus on an extensive analysis by introducing all other features that are not available for users without a cost or license. These features may include: Batch IP lookups, Bulk Data, InternetDB, Full firehose, Internet scanning API and Hostname scans. This analysis could involve evaluating the effectiveness and practical application of these features in cybersecurity, researching the quality and scope of bulk data sets for various industries, exploring historical data trends in IoT devices, studying real-time data streams for emerging threats, and assessing the capabilities and integration potential of scanning APIs and hostname scans with existing cybersecurity tools, aiming to enhance situational awareness, proactive measures, and data protection in the world of interconnected devices.

REFERENCES

[1] (2024, February 5). *What are Search Engines*? GeeksforGeeks. [Online]. Available: https://www.geeksforgeeks.org/what-are-search-engines-and-how-do-they-work/

[2] (2024, March 28). *An Introduction to Search Engines and How They Work.* WebAlive. [Online]. Available: https://www.webalive.com.au/an-introduction-to-search-engines/

[3] Rose, C. (2023, July 16). *The Complete History of Search Engines. SEO Mechanic*. [Online]. Available: https://www.seomechanic.com/complete-history-search-engines/

[4] Murillo, K. (2022, October 18). *Get to know Shodan, the scariest Search Engine on the Internet*. MasterDC. [Online]. Available: https://www.masterdc.com/blog/what-is-shodan-search-engine/

[5] *What is Shodan*? Shodan Help Center. [Online]. Available: https://help.shodan.io/the-basics/what-is-shodan

[6] (2024, April 14). *Shodan* (website). Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Shodan_(website)

[7] *Search Query Fundamentals*. Shodan Help Center. [Online]. Available: https://help.shodan.io/the-basics/search-query-fundamentals

[8] *Academic Upgrade*. Shodan Help Center. [Online]. Available: https://help.shodan.io/the-basics/academic-upgrade

[9] *Choose Your Plan*. Shodan Account. [Online]. Available: https://account.shodan.io/billing

[10] Ho, R. (2024, April 23). *What Is Shodan? How to Use It & How to Stay Protected* [2024]. SafetyDetectives. [Online]. Available: https://www.safetydetectives.com/blog/what-is-shodan-and-how-to-use-it-most-effectively/

[11] Porup, J. (2022, March 29). *What is Shodan?* The search engine for everything on the internet. CSO [Online]. https://www.csoonline.com/article/565528/what-is-shodan-the-search-engine-for-everything-on-the-internet.html