CRITICAL INFRASTRUCTURE SECURITY

Locating and Securing SCADA Devices on the Internet of Things

Eric Gross

MS MIS Candidate (2015)

Eller College of Management

Adviser: Hsinchun Chen

Contents

List of Tables

List of Figures

1 Abstract

Placing devices on the Internet of Things (IoT) has become commonplace, where everything from refrigerators to solar panels can be connected to increase the usability and accessibility of different devices. When devices are connected to the Internet of Things they become easily locatable using search tools such as Shodan, an online database of visible internet devices, which may create a potential security concern. Because these devices can control critical infrastructure, such as with Supervisory Control and Data Acquisition (SCADA) devices, these should be located and tested for potential vulnerabilities in an automated fashion. Ensuring the confidentiality, integrity, and availability of these devices can be done through the use of Shodan and custom made vulnerability assessment tools.

2 Introduction

As more devices are integrated into daily life, a paradigm shift occurs on the internet. In the past many devices were not connected to the internet, but were controlled individually. Now, devices needed for our daily lives are being connected to the internet, creating the Internet of Things. Forecasted growth suggests that 50 billion devices will be on the Internet of Things by 2020, which can be seen in Figure 1 (Evans, 2011).

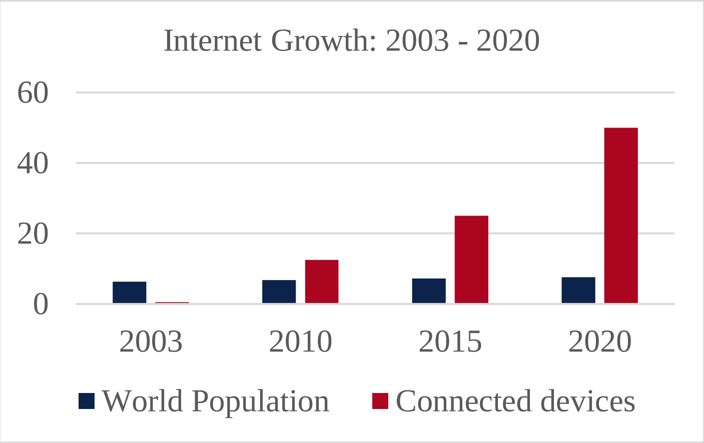


Figure 1: Internet of Things growth

This change of the types of devices which are on the internet has both positive and negative aspects. Many of these items allow for remote access, making administration and control of these devices easier for end users.

Unfortunately, this accessibility also creates the potential for malicious users who want to control many of these devices remotely. People believe that with the expansion of the Internet of Things, they can rely on security through obscurity, and thus do not protect their devices against attacks. This is a concern because it is now very easy to quickly scan the entire internet for devices using different tools, such as Zmap (Zmap, 2015). For the most part, these devices will not be severely impacted if they are exploited because they do not control any critical devices. These non-critical devices can include anything from refrigerators to home routers. However, there are also many devices that are insecure which maintain and run critical infrastructure, and a malicious attack on them could cause serious consequences. Given the rapid growth of devices on the Internet of Things, a proactive approach to securing these devices must be taken by first identifying vulnerable devices to describe the current problem.

Attack Category	**Total #**
Other	311
Buffer Overflow	201
Denial of service	193
Code Execution	68
XSS	38
Arbitrary File	33
Info Disclosure	25
SQL Injection	24
Privilege escalation	22
Memory Corruption	13

Table 1: (OSVDB, 2015)

Critical infrastructure includes devices that are a necessity for the security and safety of the users of these devices. These devices are called Industrial Control Systems (ICS) and Supervisory Control and Data Acquisition (SCADA) devices. Many of these devices on the Internet of Things perform critical tasks essential to the daily lives of everyone involved with those devices. These devices can include everything from nuclear plants, water treatment centers, pipelines, and solar plants to medical devices and more. Having these devices connected to the Internet of Things allows for the remote management of the device. Depending on the function of the ICS device, dangerous changes could be made through remote access. The exploitation of these devices can harm the critical infrastructure of companies and countries in serious ways. Many vulnerabilities for both infrastructure control devices and SCADA devices can be found in databases such as the National Vulnerability Database, Metasploit, or Exploit DB. There are several different attack vectors for SCADA devices. About 197 SCADA attack disclosures were made in 2014 (OSVDB, 2015). Of these attacks, some of the most used attack vectors included the use of unpatched and published vulnerabilities, buffer overflow attacks, and denial of service attacks (Idaho National Laboratory, 2011) (OSVDB, 2015). A list of the types of attacks on SCADA systems can be seen in Table 1. In order to secure these devices we must be able to locate all the SCADA devices with a simple and automated method.

One of the most prevalent tools used to search the Internet of Things is Shodan. Shodan is a tool which allows exploration of the Internet of Things in a quick and anonymous manner. Shodan allows for the searching of devices using both a web interface and an application programming interface (API). An example of the web interface can be seen in Figure 2. It is simple to find SCADA and ICS devices using both of the aforementioned methods. These tools scan many common ports including HTTP (80), FTP (21), SSH (22), and Telnet (23) (Shodan, 2015). Table 2 shows a sample port list that Shodan reports to collect, though not all ports are listed as new ports are continually being added. A full port list can be found in the Full Port List from Collected Results in Appendix A. Banner information, which is basic information that a device reports about itself when queried on a specific port, can be found from these ports, and other information such

Important Services/Ports	**Important Features**
HTTP: Port 80	Ip_str: The IP address of the host as a string
FTP: Port 21	Port: The port number that the service is operating on
SSH: Port 22	Data: Contains the banner information for the service
Telnet: Port 23	Location: An object containing all of the location information for the device

Table 2: (Shodan, 2015)

as HTML can be collected from the different ports that Shodan scans. Some of the most important information that is returned from the device includes the Internet Protocol (IP), device type, operating system, organization, and actual data of the device. Explanations for these fields can be found in Table 2. A full list of features can be found in the

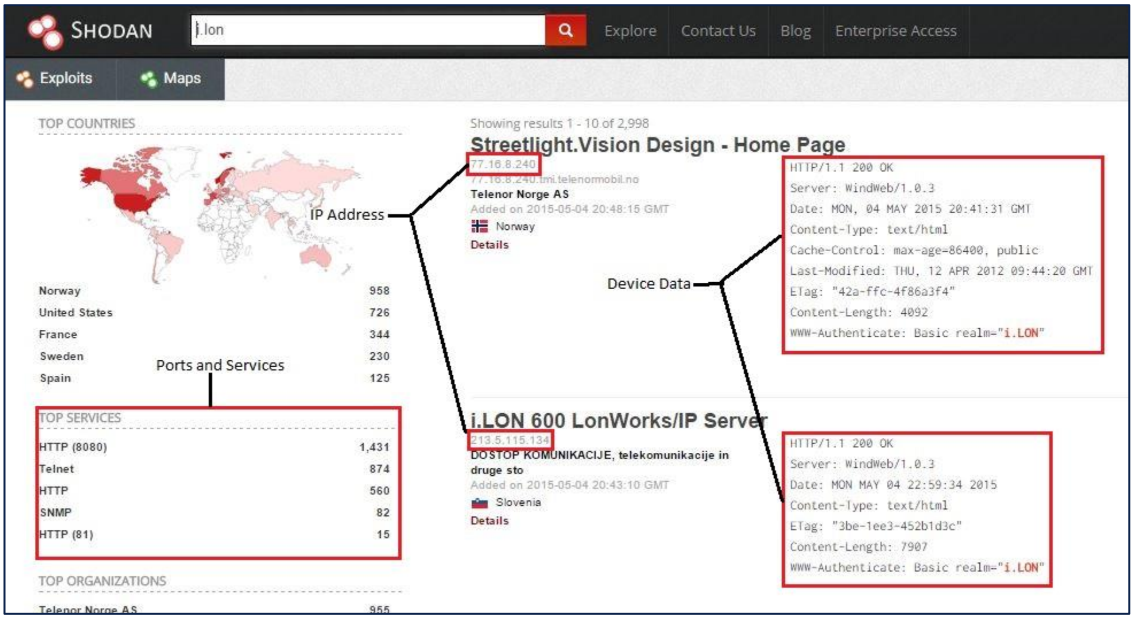
Port Number	Number of devices	Port Number	Number of devices	Port Number	Number of devices	Port Number	Number of devices
7	397035	623	577628	3749	25330	8090	475059
11	6610	626	37334	3784	4181	8098	66188
13	167087	631	1354506	3790	351	8139	426
15	4905	666	2979	4022	30615	8140	398
17	57615	771	20681	4040	81528	8181	981
21	18020071	789	10903	4369	295204	8333	7970
22	35306131	992	35408	4443	11143	8443	837421
23	23255072	993	2307122	4444	211040	8649	100900
25	19592417	995	2239856	4500	8465839	8834	311
26	1483579	1023	31706	4949	36533	8888	13795
37	270430	1177	231	5000	1363482	9051	22998
53	11006474	1200	55914	5001	12735	9100	409060
67	19223	1234	956517	5008	12367	9151	22424
79	108768	1434	336854	5060	30807264	9160	6493
80	188845478	1471	26496	5094	19435	9200	140749
81	6170994	1604	15405	5222	482436	9943	3134
82	720356	1723	8960570	5353	19185	9944	25263
83	33679	1900	13242527	5357	2840891	9981	49903
84	18883	1911	48394	5432	1185609	9999	121990
88	17709	2067	14365	5560	41180	10000	40336
102	6211	2082	3097853	5632	212505	10001	8733
110	12778071	2083	9922	5900	1509139	10243	1162308
111	4822490	2086	2689854	5901	42012	11211	311720
119	55597	2087	704	5985	688847	16010	19353
123	3912207	2123	714867	5986	11249	20000	582
137	6613066	2181	48172	6000	166802	25565	211627
143	11663809	2222	1205281	6379	180113	27017	102211
161	12353878	2323	438714	6666	15730	28017	97155
195	7621	2375	18443	7071	639	32764	14070
389	981846	2376	360	7547	88749014	44818	7993
443	15333882	2404	409	7657	19618	47808	34350
444	42229	2455	19887	7777	1277459	49152	5804356
445	1203714	2628	32076	8000	243948	50100	4098
465	1200827	3000	252528	8069	110031	55553	749
500	11545875	3128	1041221	8080	25953862	55554	370
502	29026	3306	12502920	8081	4763	62078	861336
515	727045	3388	14805	8087	941	64738	97138
523	24049	3389	9132392	8089	27736		

Table 3: Full port list

* Ports with less than 50 devices associated with them were not included

Appendix B

Full List of Services Available in Shodan in Appendix B. Though the information obtained is unstructured, and though some data in specific fields is not available, Shodan remains a tool with high utility for locating these devices. The ability to easily search using Shodan makes it a valuable tool, especially because it is more anonymous and less risky than scanning IP ranges with other tools. Furthermore, the Shodan database is always scanning and being updated with new devices and ports. The scanning function of this search engine allows for the entire internet to be collected about every two weeks (Matherly, 2015). This allows for the ability to complete an up-to-date analysis of the Internet of Things. With the use of the Shodan API, new automated methods of vulnerability assessment on the Internet of Things can be conducted.



TOP COUNTRIES

Norway	958
United States	726
France	344
Sweden	230
Spain	125

TOP SERVICES

HTTP (8080)	1,431
Telnet	874
HTTP	560
SNMP	82
HTTP (81)	15

TOP ORGANIZATIONS

Telenor Norge AS	955

Showing results 1 - 10 of 2,998

Streetlight.Vision Design - Home Page
77.16.8.240
77.16.8.240.tmi.telenormobil.no
Telenor Norge AS
Added on 2015-05-04 20:48:15 GMT
Norway
Details

```
HTTP/1.1 200 OK
Server: WindWeb/1.0.3
Date: MON, 04 MAY 2015 20:41:31 GMT
Content-Type: text/html
Cache-Control: max-age=86400, public
Last-Modified: THU, 12 APR 2012 09:44:20 GMT
ETag: "42a-ffc-4f86a3f4"
Content-Length: 4092
WWW-Authenticate: Basic realm="i.LON"
```

i.LON 600 LonWorks/IP Server
213.5.115.134
DOSTOP KOMUNIKACIJE, telekomunikacije in druge sto
Added on 2015-05-04 20:43:10 GMT
Slovenia
Details

```
HTTP/1.1 200 OK
Server: WindWeb/1.0.3
Date: MON MAY 04 22:59:34 2015
Content-Type: text/html
ETag: "3be-1ee3-452b1d3c"
Content-Length: 7907
WWW-Authenticate: Basic realm="i.LON"
```

Figure 2: Shodan Interface

3 Background and Problem Definition

3.1 Previous work

There have been multiple studies exploring the Internet of Things, security of the Internet of Things and SCADA devices, and studies of data mining on the Internet of Things. In particular the three main streams of research being reviewed include:

- SCADA device security
- Use of Shodan for device location and vulnerability testing
- Data mining on the Internet of Things

Ensuring a strong grasp on the topics in these past papers is important for identifying current research gaps in this field. This research will give us a better idea of how we can improve the security of devices on the Internet of Things and how to automate security on a large-scale basis.

3.1.1 SCADA device security

Prior research in SCADA security has qualitatively analyzed many potential vulnerabilities and attack vectors in SCADA systems. These studies do not attempt to test or identify current vulnerabilities of SCADA devices in an automated or scalable manner. The review for most of these vulnerabilities is a qualitative overview.

Assessment of SCADA vulnerabilities

Many SCADA devices do not have good patch management control, which leads to the increased number of vulnerabilities on these types of devices (Nicholson, Webber, Dyer, Patel, & Janicke, 2012). Previous work looked at the identification of multiple aspects of SCADA devices and different attack vectors, including the most common and dangerous attack vectors and system security. System security for SCADA devices includes ensuring that a password is set in order to control access, using antivirus to prevent system exploits, having the device behind a firewall, and patching systems (Robles & Choi, 2009). One of the next most important protection methods for the SCADA systems is the defense of the network. SCADA systems can be attacked in a number of ways from an external network. These exploits can allow the user to control the affected SCADA devices or interfere with the SCADA device functionality. Many SCADA devices use common platforms such as UNIX or Windows as their operating systems, which may allow for more attack vectors. There may be more custom software or shells depending on the device, which will require a more detailed knowledge about that specific SCADA device. Overall this past research looks at the current vulnerability climate for SCADA devices which needs to be protected to ensure the security of the critical infrastructure. (Robles & Choi, 2009)

SCADA devices usually have a software architecture which is based on commercial, off-the-shelf products. Because of this, many of these SCADA devices have very similar vulnerabilities to other devices which may be used in a company. (Hentea, 2008) (Igure, Laughter, & Williams, 2006) The SCADA devices usually do not get the same visibility as other devices so they may be more vulnerable to exploits (Hentea, 2008). Many people believed that SCADA devices were electronically segregated from other networks. (Igure, Laughter, & Williams, 2006). Many of these off-the-shelf products support services such as UNIX or DOS, which increases the number of potential vulnerabilities that can be located in these devices. SCADA device vulnerabilities are often overlooked because there is not a common framework which can help mitigate the risk of attacks on SCADA devices (Hentea, 2008). This dependence on off-the-shelf software can also lead to security issues which deal with the SCADA device vendor. A vendor may not publish patches and people may not install a patch on a SCADA device even if a vendor publishes one

(Nicholson, Webber, Dyer, Patel, & Janicke, 2012). Other types of attacks specifically target the availability of SCADA devices, such as DDOS (distributed denial of service) attacks carried out to stop the functionality of SCADA devices (Nicholson, Webber, Dyer, Patel, & Janicke, 2012). New security strategies and technologies are currently being developed for owners of SCADA systems. Multiple security fixes are available to mitigate the risk of the different attacks on SCADA devices. Some of these include sensor networks SCADA, creating SCADA devices which run on a microkernel architecture, developing with security in mind at the beginning of projects, integrating new technologies to increase security, and conducting vulnerability analysis based on discovery and adaptation solutions (Hentea, 2008). Other issues with SCADA networks include the confusion that can come up when creating these networks.

Multiple SCADA attacks have been carried out in malicious ways which is why these devices need to be secured. (Nicholson, Webber, Dyer, Patel, & Janicke, 2012). These malicious parties can be anyone from disgruntled employees to “hactivists”. Because of this, these devices need to be protected on all fronts. The risks of having a SCADA device compromised can be very impactful to the organizations and people who are affected. Companies can suffer from brand damage which can lead to monetary loss, and even the loss of life. (Nicholson, Webber, Dyer, Patel, & Janicke, 2012)

3.1.2 Shodan exploration and testing

Past work using Shodan mostly contains methods for using the API or web interface in a manual manner. Exploration has been done looking at using Shodan as a device discovery platform for the Internet of Things.

Discovering deliberately exposed verses obfuscated SCADA devices

Prior research has configured non-obfuscated and obfuscated devices with public internet address exposure to determine Shodan’s ability to discover them. Researchers set up four different Allen-Bradley PLCs (Programmable Logic Controllers) in order to examine the indexing and searching functionality in locating these devices. Two of these devices were set up with default banner information and two of these devices were set up with an obfuscated banner using Raspberry Pi proxies. These devices were set up with two specific ports, port 80, and port 44818, which are ports typically used for these types of PLC. These devices were then attached to external IP addresses in order to be indexed by Shodan. Once the PLCs were scanned and indexed by Shodan, the devices were then tested for ease of identification. This was done by tasking a researcher to manually create Shodan queries with a basic knowledge of PLCs. The results of this experiment exhibited the ability to locate the non-obfuscated devices with just general knowledge of these devices. The non-obfuscated devices were found multiple times with different queries created by the researcher. The devices with the obfuscated banners were also found, but they were only found in general searches where they were mixed with millions of other devices. This exercise showed the potential to sanitize banner information from devices in order to limit the ease of location on the Internet of Things. (Bodenheim, Butts, Dunlap, & Mullins, 2014)

Discovering Existing SCADA devices

Researchers have analyzed the effects of locating ICS devices using information from PLCs via common industry protocols (Williams, 2014). This research involved the collection of SCADA devices, in particular Allen-Bradley PLCs, using Shodan. The devices were collected using a query in Shodan relating to the CompactLogix and ControlLogix Allen-Bradly systems. A total of 493 unique IP addresses were collected from Shodan, with 167 unique IP addresses being used for the

CIP testing. The devices were verified with the help of Industrial Control Systems Engineers in order to ensure the device was acting as an ICS device. Before the ICS devices were tested, verification procedures to ensure that the device would not be impacted negatively were conducted. These devices were then tested using common industry protocol requests sent to different devices located with Shodan. These requests contained specific PLC programming information which incited a response from the device. The responses from these devices were then collected and analyzed to find out the function of the devices and in what industry sector they were used. Out of all the devices collected, the largest groups were unknown (27%), wastewater (32%), manufacturing (5%), oil and gas (5%), and alarm notification (5%). The results of the research demonstrate that it is possible to easily collect small subsets of specific devices from Shodan in order to do further and more specific analysis on them. (Williams, 2014)

Testing the vulnerability of Exposed SCADA devices

Ensuring multiple devices, including SCADA and ICS devices, are secure on the Internet of Things using Shodan is the main focus of the next work (Patton, et al., 2014). Research was conducted looking at the ability to easily locate devices on the Internet of Things using Shodan to determine the security of these devices. Work was done collecting devices from the Shodan API using different keywords that were used to classify the devices. The devices were saved to a database to allow for faster and easier access. About 250,000 of these devices were classified as SCADA devices from keyword analysis. A new database was then created in order to store default credentials to many different devices. Vulnerability assessment was conducted on a subset of 35,737 of the devices collected from Shodan. The vulnerability assessment included the testing of different default passwords on devices with open HTTP (80) and Telnet (23) protocols. The testing was done in an automated manner using a custom python script. The default passwords were used to assess the vulnerability of these devices by reporting how many of these devices were vulnerable. Three types of devices were tested which included ILON SCADA devices, Niagara SCADA devices, and traffic control systems. 1258 ILON devices were tested on port 25 of which 3.5% were using default credentials Figure 3. 34,248 Niagara devices were tested on port 80 of which .44% were using default credentials, and 231 traffic control systems were scanned, of which 40% were using a default username and password. The results of this study show that the ability to easily collect and test the vulnerability of devices in an automated manner exists, and precautions need to be followed in order to mitigate these threats. (Patton, et al., 2014)

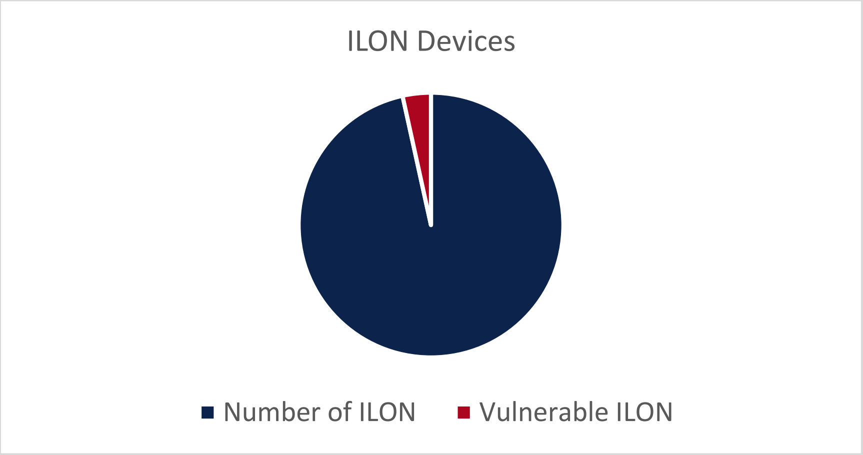


Figure 3: Vulnerable ILON devices on Telnet

Testing the vulnerability of all devices on the Internet of Things

Being able to test the vulnerability of all devices on Shodan is an important feature which can be used to help secure SCADA devices. The next line of research deals with the vulnerability assessment of all devices on Shodan, and used entries from the National Vulnerability Database to test devices and their security. The tool created, called ShoVat allowed for the quick discovery and analysis of devices which are collected from Shodan. This tool used the National Vulnerability

Database and their collection common platform enumerations (CPE). The CPE is the aggregate of a vendor, product, and version number. For example a CPE for a Linux Kernel version 3.4.54 would be: cpe:/o:linux:linux_kernel:3.4.54. The program looked at the device data returned by Shodan queries to determine if the devices had a vulnerable product and version. The program ranked the vulnerabilities based on how strong of a correlation the device data has to the vulnerabilities through a multi tree data structure where the CPE name is reconstructed. This tool allows for fast analysis of devices using their device information which can be located in Shodan (Genge & Enăchescu, 2014).

3.1.3 Data mining on the Internet of Things

Research on data mining on the Internet of Things has been done using structured data from devices on the Internet of Things. Past work does help in the methodology and collection of information from a large set of devices.

Data mining unstructured data from multiple sources

When working with unstructured data from devices, it is possible that the data will not always be classified correctly. In past work, classification was used to identify the differences between workers and machinery on work sites (Chi, 2011). This research explores the ability to classify different work-related machines and workers into groups in order to be able to automatically recognize them. Looking into creating the ability to have real time object identification using distributed cameras is similar to looking at a distributed network of other devices, such as the Internet of Things. In order to identify these objects classification of 750 images was done to train a classifier to differentiate between images of objects and images without objects. This methodology can also be applied to other data, such as device data on the Internet of Things. The ability to use classification algorithms, such as Naïve Bayes and Neural Networks, can allow for the successful identification of unknown data in real time (Chi, 2011). With the use of these different technologies, the Internet of Things can be made more intelligent which will allow for better services (Tsai, Lai, Chiang, & Laurence, 2014). Past research also shows that sensors can also be used to classify different activities from multiple data sources (Fleury, Noury, & Vacher, 2009). Multiple sensors can be trained by a classification algorithm to determine activities and habits of people at home. Once the classifier is trained, a high accuracy can be expected from looking at this distributed data (Fleury, Noury, & Vacher, 2009).

Data mining models for the Internet of Things

Past research looks at the complexity of different data mining techniques and models on the Internet of Things (Bin, Yuan, & Xiaoyi, 2010). Models such as a multi-layered model, distributed data mining model, grid based mining model, and a multi-technology integration model can be used to successfully mine data on the Internet of Things. These different models look at mining data from specific devices which will have structured data (Bin, Yuan, & Xiaoyi, 2010). A qualitative study of the different models helps to define what the best data mining structures are best for the large amount of data that will be collected from the Internet of Things.

3.2 Research Gaps

The review of related literature leads to the discovery of several different gaps which have yet to be researched. In past work looking at SCADA security, little work has been done looking at actual exploration of the current security on SCADA devices. Past research mostly looks at the qualitative aspects of SCADA security. Past work of Shodan consists of manual exploration of several different device or vulnerability types. No work has been done looking at subsets of devices taken from Shodan. Finally, past research about data mining on the Internet of Things consist of working with non-SCADA devices and semi-structured data from specific devices on the Internet of Things. Using the data of the device (such as vendor or version), rather than data created by the device is a new concept in analyzing data mining on the Internet of Things.

3.3 Research Questions

In order to address the gaps above research will need to provide an automated and scalable framework to classify SCADA devices, identify the types and amount of SCADA devices on the Internet of Things, and access the vulnerabilities of these devices on the Internet of Things. Based on the needs of these gaps to be filled, the following research questions were asked:

- Can SCADA devices in the Internet of Things be located in an automated manner?
- How many SCADA devices are in the Internet of Things?
- How many of these SCADA devices are vulnerable?

4 Methodology, Experiments, and Results

4.1 Introduction and Approach

In order to solve these different research problems, becoming accustomed to the different tools and technologies to be used was essential. Experimentation and research can be broken into the three different categories of exploration, SCADA collection, and vulnerability assessment. Initial research was done on the Shodan Search engine to see the capabilities of this tool and how it could be used to test different devices, particularly SCADA devices. Having this knowledge led to the creation of different testing tools. These tools were created using the python programming language. Ensuring that these custom tools could be used to effectively test for multiple vulnerabilities in a scalable manner was essential to the success of filling the aforementioned research gaps. After the completion of these custom-made programs, they were tested on a subset of devices to ensure that they would work correctly. Collection of known types of SCADA devices was then conducted in order to better understand these devices, as well as form a strong set of known SCADA devices. A method was created to locate all SCADA devices on the Internet of Things using a classification algorithm as well as the previously mentioned collection of SCADA devices. This classifier is able to locate SCADA devices to be tested with our vulnerability assessment programs. The complete research design for this process can be seen in Figure 4. This process leads to a new paradigm for collecting and analyzing the vulnerability of different devices on the internet. In this experiment SCADA and critical infrastructure devices are used as an example, but this methodology can be applied to other devices as well.

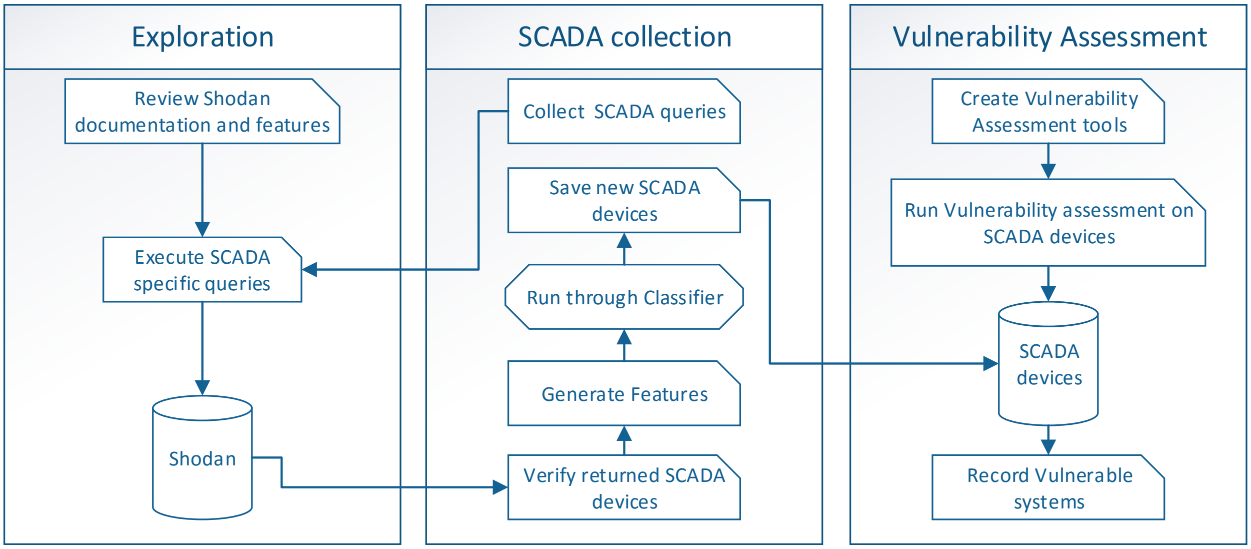
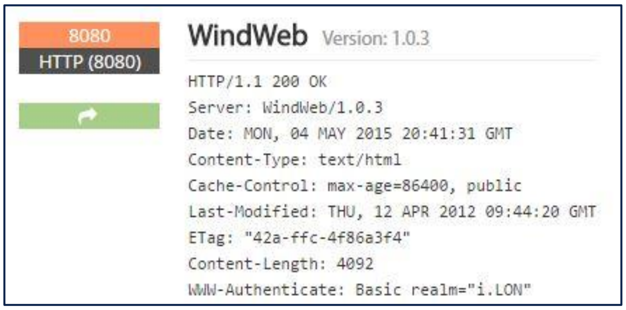

Exploration
Review Shodan documentation and features
Execute SCADA specific queries
Shodan
SCADA collection
Collect SCADA queries
Save new SCADA devices
Run through Classifier
Generate Features
Verify returned SCADA devices
Vulnerability Assessment
Create Vulnerability Assessment tools
Run Vulnerability assessment on SCADA devices
SCADA devices
Record Vulnerable systems

Figure 4: Methodology

4.2 Exploration

In order to accurately locate and test the vulnerability of devices on the Internet of Things, understanding of the different tools was needed. This section includes the exploration Shodan, which is the main tool to locate these devices. Shodan allows for the search and exploration of all devices on the Internet of Things. This exploration can be done either through an API or through a web interface. Shodan indexes different devices by randomly scanning IP ranges and ports and saving a number of returned fields. These fields can be used to analyze different attributes of these devices. Full exploration allowed for the understanding of the different devices which could be located using this tool.

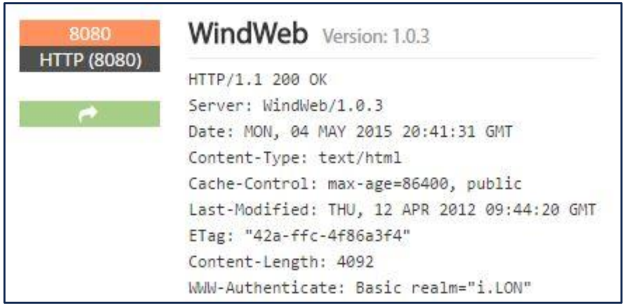


```
HTTP/1.1 200 OK
Server: WindWeb/1.0.3
Date: MON, 04 MAY 2015 20:41:31 GMT
Content-Type: text/html
Cache-Control: max-age=86400, public
Last-Modified: THU, 12 APR 2012 09:44:20 GMT
ETag: "42a-ffc-4f86a3f4"
Content-Length: 4092
WWW-Authenticate: Basic realm="i.LON"
```

Figure 5: Device data example

4.2.1 Methodology and work completed

Exploring the devices on Shodan was the first step leading to the collection of SCADA devices. The first step of exploration was accessing the web interface to do basic exploration of the types of devices which Shodan indexed. This was done by looking at many of the premade queries from other people to search Shodan. Most of these queries were for printers and web cameras, but there were a small amount for SCADA and embedded devices. Learning how to use filters such as the port and IP ranges to search for specific results was also important for narrowing down specific types of devices we wanted to see.



WindWeb Version: 1.0.3

```
HTTP/1.1 200 OK
Server: WindWeb/1.0.3
Date: MON, 04 MAY 2015 20:41:31 GMT
Content-Type: text/html
Cache-Control: max-age=86400, public
Last-Modified: THU, 12 APR 2012 09:44:20 GMT
ETag: "42a-ffc-4f86a3f4"
Content-Length: 4092
WWW-Authenticate: Basic realm="i.LON"
```

As more types of devices were investigated, we started looking for similarities in the headers of the devices. These headers usually included general data about the type of device which was being accessed. Some of the header information can be seen in Figure 5. This led to more specific searches for SCADA devices. In particular the first types of SCADA devices we searched for were I.LON and Niagara SCADA systems. Our original research looked for the vulnerability of these devices using default password testing (Patton, et al., 2014). This led to the interest of being able to locate all devices of a specific type using Shodan. These devices were collected using the Shodan API which we worked with extensively.

Understanding the data returned from the API as well as being able to parse this data into a database led to the ability to complete exploration in an automated and efficient manner. All data collection with the API was done using an unlimited access account for academic use. After discussions with the creator of Shodan, we were able to receive access to an unlimited professional account which gave us access to all records in JSON format which allowed for much faster collection of data. The collection of this data can be done manually using scanning tools, but it will be more intensive than using the data which is available through Shodan. Shodan offers an anonymous and easy way to collect data about specific devices, which is why we explored this tool.

4.2.2 Results

Our original collection of data was gathered using the API which would then parse the data into a MySQL database. The collection of about 700,000 different devices was completed in order to analyze the security in an automated manner. These devices included SCADA devices, printers, servers, and web cams. We also scanned specific IP ranges of colleges to test the vulnerability of different organizations. The ability to use the unlimited business account led to a much larger results set. We created a database with about 630 million records which was inclusive of about one month of data from Shodan from January 2015. All of the data collected included only the most populated fields which were: IP address, port number, timestamp, and device data. Much of the other data fields provided by Shodan is very sparse (about 20-30% data available) and was not good for analysis of a large set of devices. All of the 630 million records in our database were inserted via JSON files. Due to the unstructured nature of the data, some of the insertions into the database included incorrect information. For example, some of the device data may be included in the IP field, rather than being populated with a correct IP. In order to mitigate this issue, specific query restrictions were made. In order to ensure an IP was selected, a regular expression was made to ensure the first character of the IP field was a number. For the ports available on Shodan, any port which had less than 50 distinct devices relating to that port were removed.

After the collection of this data, a select distinct count SQL query was done with regular expressions on the IP addresses in this dataset. The result of this query was that a total of 180,695,201 unique device IPs are contained in Shodan for the month of January. Due to the fact that the entire internet is scanned every two weeks by Shodan, it can be assumed that this number is the total number of directly accessible devices connected to the internet on the scanned ports available in Shodan. The full list of ports can be found in Appendix A. A distinct port query was also conducted which allowed us to find all of the aforementioned ports which are available on Shodan. A total of 151 ports were located in our dataset. The top 10 most populated ports can be seen in Figure 6.

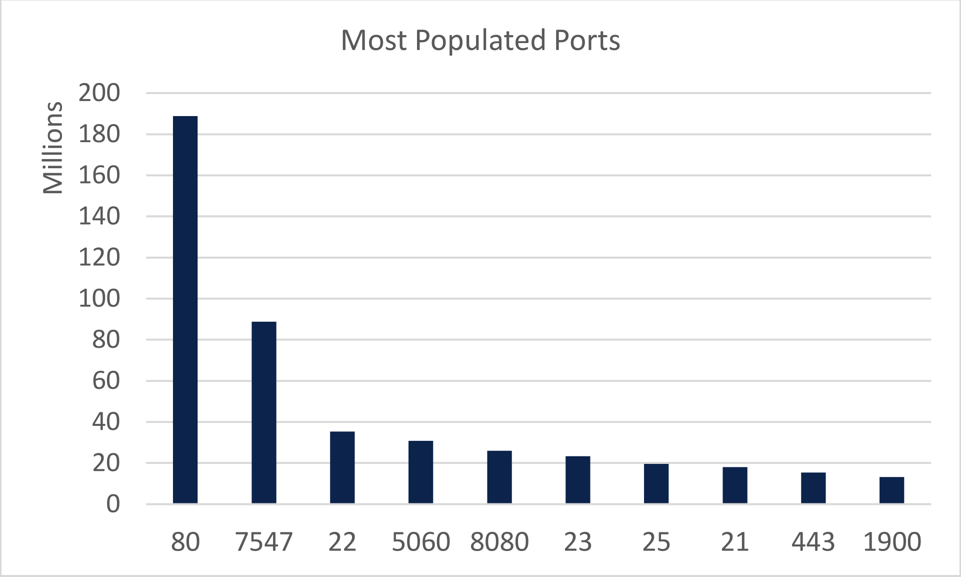


Figure 6: Most populated ports in Shodan

4.2.3 Discussion

Collecting devices from the Internet of Things can be challenging, but using Shodan is one of the easiest ways to accomplish this task. The quick and accurate access to all of the devices scanned by Shodan can greatly reduce the time needed for device discovery using other tools. The drawback to using Shodan for collection of results is the lack of information you may need for device analysis. During our collection, the ability to gather more SCADA specific ports would have greatly increased the ability for us to analyze these and other devices.

4.3 SCADA collection

One of the main goals of our research was the collection of SCADA devices using Shodan. With basic analysis of the device data, it can usually be determined if a device is a SCADA device or if it is not a SCADA device. We needed to do this in a much more automated manner in order to quickly categorize all the SCADA devices on the Internet of Things. In order to make this process more automated we decided to use classification algorithms. Random forest and Naïve Bayes were used with the devices data feature. The classification of SCADA devices was based on multiple features. In order to train the classifier, 45 different SCADA specific queries which had a very high probability of returning SCADA devices were used. These keywords were collected by looking at premade queries in Shodan and by creating our own keywords for Shodan. The keywords which were created were located by finding relevant banner information in SCADA device documentation and then testing these keywords in Shodan. Some examples of these queries include: Rockwell Automation, 8650 ION, Honeywell BNA, and plc port:102. A full list can be seen in the

Appendix C

SCADA Queries Used for Classifier in Appendix C. These keywords returned a total of 92,012 unique records from Shodan. We also had 26 different non-SCADA specific queries which returned devices that should not be classified as SCADA devices. A total of 64,211 distinct non-SCADA records were returned from Shodan. Once this data was collected from Shodan, the classifier was trained and used to classify SCADA devices using data from Shodan.

4.3.1 Methodology

The first step in locating SCADA devices using Shodan was the preparation of data. This is where the SCADA and non-SCADA devices were collected from Shodan using known SCADA queries such as Honeywell Excel and 8600 ION. A subset of these results was manually verified to ensure that the majority were actually SCADA devices. The same method was done to collect the non-SCADA devices. For each of these devices Shodan collects we retrieve 30 different data fields, which can be seen in the

Port Number	**Number of devices**	**Port Number**	**Number of devices**	**Port Number**	**Number of devices**	**Port Number**	**Number of devices**
7	397035	623	577628	3749	25330	8090	475059
11	6610	626	37334	3784	4181	8098	66188
13	167087	631	1354506	3790	351	8139	426
15	4905	666	2979	4022	30615	8140	398
17	57615	771	20681	4040	81528	8181	981
21	18020071	789	10903	4369	295204	8333	7970
22	35306131	992	35408	4443	11143	8443	837421
23	23255072	993	2307122	4444	211040	8649	100900
25	19592417	995	2239856	4500	8465839	8834	311
26	1483579	1023	31706	4949	36533	8888	13795
37	270430	1177	231	5000	1363482	9051	22998
53	11006474	1200	55914	5001	12735	9100	409060
67	19223	1234	956517	5008	12367	9151	22424
79	108768	1434	336854	5060	30807264	9160	6493
80	188845478	1471	26496	5094	19435	9200	140749
81	6170994	1604	15405	5222	482436	9943	3134
82	720356	1723	8960570	5353	19185	9944	25263

83	33679	1900	13242527	5357	2840891	9981	49903
84	18883	1911	48394	5432	1185609	9999	121990
88	17709	2067	14365	5560	41180	10000	40336
102	6211	2082	3097853	5632	212505	10001	8733
110	12778071	2083	9922	5900	1509139	10243	1162308
111	4822490	2086	2689854	5901	42012	11211	311720
119	55597	2087	704	5985	688847	16010	19353
123	3912207	2123	714867	5986	11249	20000	582
137	6613066	2181	48172	6000	166802	25565	211627
143	11663809	2222	1205281	6379	180113	27017	102211
161	12353878	2323	438714	6666	15730	28017	97155
195	7621	2375	18443	7071	639	32764	14070
389	981846	2376	360	7547	88749014	44818	7993
443	15333882	2404	409	7657	19618	47808	34350
444	42229	2455	19887	7777	1277459	49152	5804356
445	1203714	2628	32076	8000	243948	50100	4098
465	1200827	3000	252528	8069	110031	55553	749
500	11545875	3128	1041221	8080	25953862	55554	370
502	29026	3306	12502920	8081	4763	62078	861336
515	727045	3388	14805	8087	941	64738	97138
523	24049	3389	9132392	8089	27736		

Table 3: Full port list

* Ports with less than 50 devices associated with them were not included

Appendix B

Full List of Services Available in Shodan in the Appendix. Many of the fields are not used due to the aforementioned problems with sparse data. We continued with analysis on port and banner data because of the completeness of this data.

The next step was generating N-grams from the data, which helped locate the most popular terms in the device data field of the different collected SCADA devices. This is done for both SCADA and non-SCADA devices so that N-grams in both sets could be removed. Features were then created so that the classifier could be trained. There were 15 features in total which will be used for classification. They included a combination of the N-grams, Port, Length, and the number of grams in the data. The classifier was then trained using the SCADA and non-SCADA data sets retrieved from Shodan.

4.3.2 Process

Following the above methodology, we needed to apply the classifier to sets of unclassified data from Shodan. Our original test included the testing of 498,929 records which were randomly collected from Shodan. We wanted to evaluate the classifier based on data which had not been used to train the classifier. Once the devices in this set were classified as SCADA devices, they were verified with domain experts familiar with SCADA devices. The final chosen algorithm is Random Forest which classified 13.2% of the devices from the 498,929 set as SCADA devices. This classifier can be used on a larger set of data in order to analyze the amount of SCADA devices on the Internet of Things.

4.3.3 Results

The above set of results was classified with a confidence level of .8, which led to a total of 66,229 distinct SCADA devices. This leads to the conclusion that about 13.3% of devices located in Shodan are SCADA devices. The classifier can be run of much larger sets of results making the analysis of large sets of data scalable. These results are then parsed into a database in order to conduct vulnerability analysis on them. A breakdown of the SCADA vs Non-SCADA devices can be located in Figure 7.

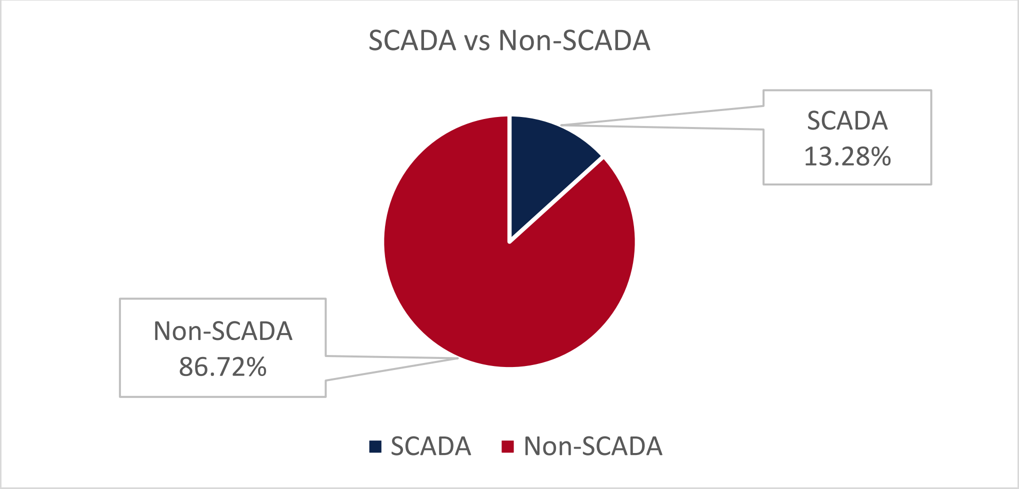


Figure 7: Classification of SCADA vs Non-SCADA

4.3.4 Discussion

Using data mining to locate discover device types is a novel idea. Most past research has looked at structured data from devices on the Internet of Things, but not the devices themselves. Having the ability to classify all the different types of devices on the Internet of Things will allow for advanced security testing of all devices on the internet. This same data mining methodology can also be applied to other devices, such as medical devices. The above methodology can be applied to the complete set 630 million records from Shodan, but limitations with processing power and time restricted the analysis of the whole data set. With greater processing power, most SCADA devices on Shodan can be located using the same methods. Preemptive collection of critical sectors can be used for vulnerability testing. Once devices like this are located, vulnerability assessment can be completed for the located devices.

4.4 Vulnerability Assessment

In order to conduct a vulnerability assessment on the located SCADA devices, novel vulnerability assessment programs needed to be created. We tested two different security threats to which SCADA devices are currently susceptible. The first line of experimentation examined the use of default passwords on SCADA devices. The second line examined if SCADA devices located on Shodan were susceptible to current vulnerabilities in the National Vulnerability Database. Both of these experiments were read-only and no device data was modified or read. Special precautions were put into place in these programs to ensure the confidentiality, integrity, and availability of the devices being tested.

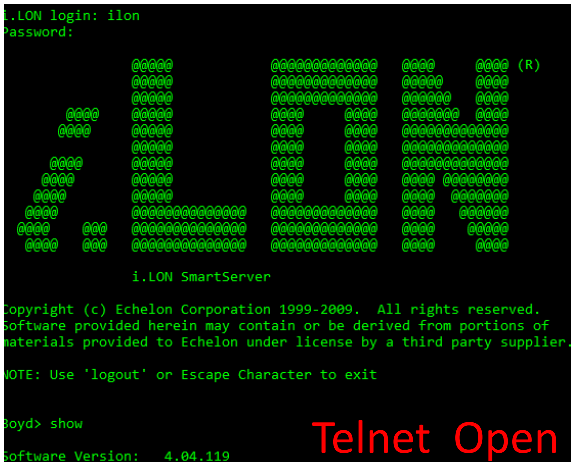
4.4.1 Study design using default password analysis on Telnet

One of the largest security threats is the ability for people to locate default credentials in device documentation and exploit these critical devices in different ways by simply by logging in. Many SCADA devices allow for remote access to devices through multiple protocols. The main protocol we looked at was Telnet on port 23. This protocol is used to remotely modify different devices using a full shell on the device. Some SCADA devices come with this setup as a default option using default credentials (Power Measurement, 2003) which allow for a very low barrier to exploitation of these devices. The main reason Telnet is tested is because Telnet is an unsecure protocol and there is a much higher accuracy when testing passwords on Telnet. Past work has been done on testing for passwords on Telnet, which supported our decision to use Telnet

(Tillapart, Thumthawatworn, & Santiprabhob, 2002) (Internet Census, 2012). In order to effectively test default passwords on Telnet, we gathered some common default passwords from multiple SCADA devices, and chose the most common to test against all devices. We tested all the SCADA devices we collected only once to ensure that they would not be locked out due to multiple password attempts. Specific credentials were used for different types of known SCADA devices. For example, the nickname for a particular subset of SCADA devices was PowerLogic ION8600. A similar name of ION8600 was used as the device in the password database. Because these two devices were similar, the username and password for that specific device was used. The credentials of username: "admin" and password: "password" were used for devices which no default credentials in our password database. The program we created would test the SCADA devices and report back which devices were vulnerable to this basic password test.

Process

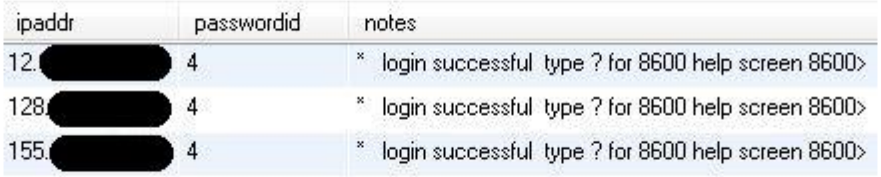
The creation of this program began with the analysis of different SCADA devices on the Telnet protocol. During this analysis, it was noted that most SCADA devices are protected with password authentication. In order to authenticate remote users the devices will have one of the following words in the Telnet response: login, user, or password. If these do not show up in the Telnet information then the device is most likely not using password protection. The program we created will read the entire Telnet response, and will write a default username at the end of the response in the username or login box. The program will then search for another response using



```
i.LON login: ilon
Password:

                @@@@@            @@@@@@@@@@@@@   @@@@     @@@@ (R)
                @@@@@            @@@@@@@@@@@@@   @@@@@    @@@@
                @@@@@            @@@@@@@@@@@@@   @@@@@@   @@@@
        @@@@    @@@@@            @@@@     @@@@   @@@@@@@  @@@@
       @@@@     @@@@@            @@@@     @@@@   @@@@@@@@@@@@@
                @@@@@            @@@@     @@@@   @@@@@@@@@@@@@
      @@@@      @@@@@            @@@@     @@@@   @@@@@@@@@@@@@
     @@@@       @@@@@            @@@@     @@@@   @@@@ @@@@@@@@
    @@@@        @@@@@            @@@@     @@@@   @@@@  @@@@@@@
   @@@@         @@@@@@@@@@@@@@   @@@@@@@@@@@@@   @@@@   @@@@@@
  @@@@    @@@   @@@@@@@@@@@@@@   @@@@@@@@@@@@@   @@@@    @@@@@
   @@@@   @@@   @@@@@@@@@@@@@@   @@@@@@@@@@@@@   @@@@     @@@@

                i.LON SmartServer

Copyright (c) Echelon Corporation 1999-2009.  All rights reserved.
Software provided herein may contain or be derived from portions of
materials provided to Echelon under license by a third party supplier.

NOTE: Use 'logout' or Escape Character to exit


Boyd> show

Software Version:   4.04.119
```

Figure 8: Example of vulnerable SCADA device on Telnet

the field password. If this field is located then the program will write the default password at the end of the field. If the response does not contain the words: user:, username:, login:, or password: then the device is seen as vulnerable and it is added to a database of vulnerable devices. If one of these is found in the Telnet response then it is likely that the password is incorrect and the tested device is asking to enter a new password or login again. Doing this basic vulnerability assessment will work with most devices on Telnet as they use similar keywords for user access.



ipaddr	passwordid	notes
12.[redacted]	4	* login successful type ? for 8600 help screen 8600>
128.[redacted]	4	* login successful type ? for 8600 help screen 8600>
155.[redacted]	4	* login successful type ? for 8600 help screen 8600>

Figure 9: Examples of Vulnerable PowerLogic Devices

Results

Many different devices are vulnerable on Telnet and allow full access to devices by simple using default passwords. An example of what a vulnerable system on Telnet can be seen in Figure 8. The unique set of 92,012 known SCADA devices were tested using the methodology above. Out of this set of devices, 1,665 devices were found on Telnet. These devices were tested against the default password database. Out of these devices, 442 were vulnerable to a basic default password attack. This is about 26% of all Telnet devices (Figure 10), and .48% of the whole SCADA subset. Once the devices were tested, they were stored in a vulnerable systems database for further verification. An example output from the vulnerable systems database can be seen in Figure 9.

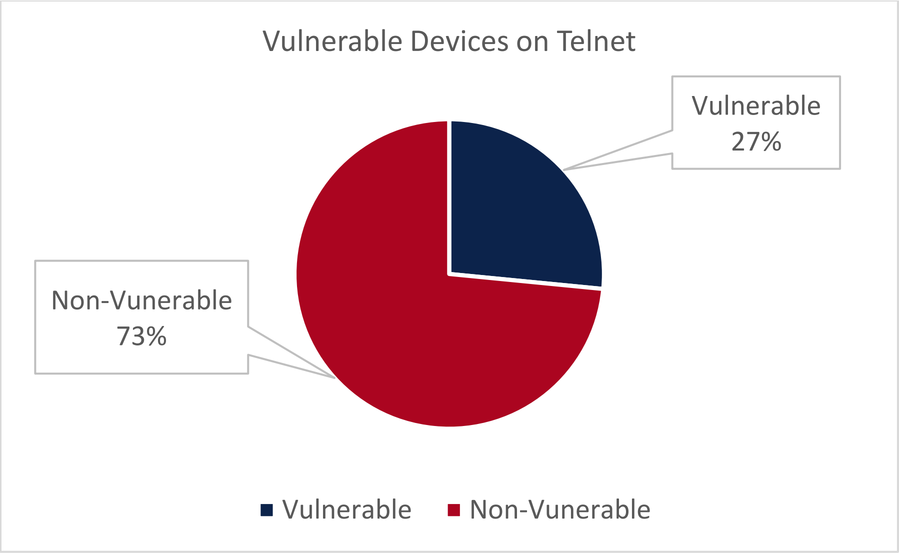


Figure 10: Percentage of Vulnerable Telnet Devices

4.4.2 Study design using the National Vulnerability Database

Another major threat facing SCADA devices is the exploitation of known vulnerabilities for specific devices which are left unpatched (Idaho National Laboratory, 2011). The ability to effectively test if devices are susceptible to past vulnerabilities can make securing devices much simpler to implement. In order to find the vulnerabilities we imported all of the vulnerabilities from the National Vulnerability Database. We would then compare this to the specific devices in the data field and then search for a matching version number in the same devices data field. This was a similar approach to the work done in (Genge & Enăchescu, 2014) where they tested a subset of devices on Shodan against the National Vulnerability Database. This would allow us to accurately locate devices which were using the same product and version from our SCADA device set which had current exploitable vulnerabilities. The National Vulnerability Database also provides a Common Vulnerability Scoring System (CVSS) which categorizes vulnerabilities by severity. This method can be applied to any set of device to quickly ensure that a network is not running any vulnerable devices.

Process

The first step of this process was collecting all of the National Vulnerability Database (NVD) entries for vulnerability analysis. All of the vulnerabilities are available on the NVD website in XML format. A script was created to parse all the unique CVE numbers, Vendors, Products, and versions into a database. This would allow for quick availability of all vulnerabilities for different devices. Another python script was then created which would do a cross referencing test of the devices in the NVD against the device data provided from Shodan. Once a device was found then all of the versions were tested against the devices that had the product from the NVD. In order to ensure our python script was not picking random numbers that may show up in device data, several precautions were put into place. The version number must follow the product name in the device data and it must not have any other numbers or periods before the version number. An example would be searching for the version number 5.2. In order to make sure correct devices were located our script ensured that this search would not match, for example, 1.5.2. This allowed for the accurate collection of the devices. Once the devices were seen as vulnerable, they were saved to a vulnerable systems database. Products which had less than three characters were not counted due to the possibility of having a match in device data which was not associated with the actual vulnerability. An example would be the product “ie”. This was dropped because it matched too many devices where this product was found in the device data, but was not the actual product. Because many products with specific versions may have multiple vulnerabilities, the discovered devices are only categorized by either being vulnerable or not vulnerable.

Results

SCADA devices which were located from the classifier were parsed into a database for further vulnerability analysis. A total of 66,229 SCADA devices from the classified results were tested using the above methodology. Only the last four years (2012-2015) of vulnerabilities from the NVD were used to assess device security. A total of 2067 vulnerabilities were located in the SCADA test set which was inclusive of 1559 unique devices. A device can have multiple vulnerabilities which is why the total number of vulnerabilities is larger than the total number of devices. About 2.3% of SCADA devices from the collected subset are vulnerable. Most of the vulnerabilities found were related to popular software which SCADA devices usually run with. Some of the popular vulnerabilities are located in Figure 11. These results relate to the idea that many of the vulnerabilities are found in commercial off-the-shelf software which has not been patched.

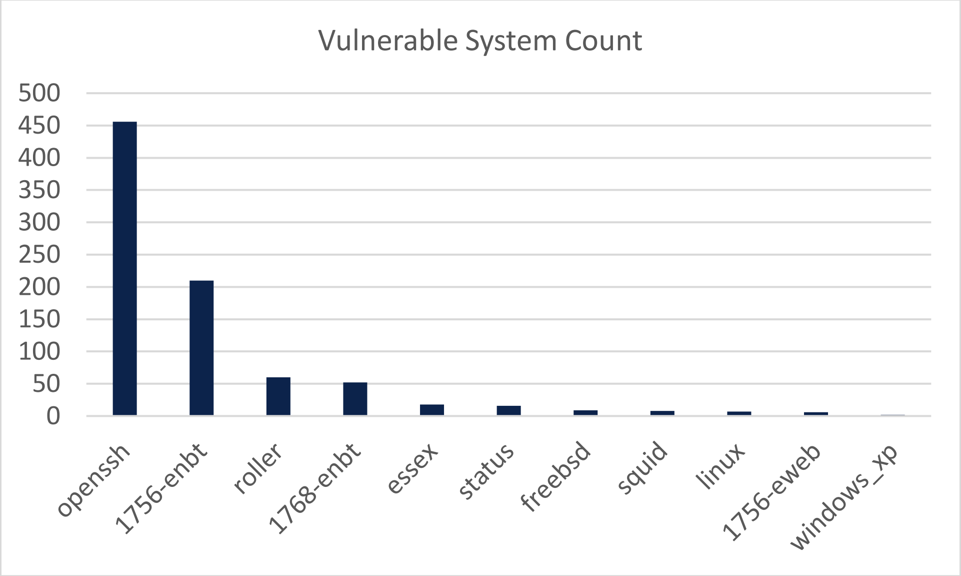


Figure 11: NVD Vulnerable System Count

4.4.3 Discussion

Being able to test the vulnerability of devices in an effective and automated way will lead to the improvement of the security of the tested devices. Multiple methods for testing the security of devices exist, but many use real-time analysis rather than past collected data. This change in the testing of devices leads to a much faster method for testing devices with less overhead. Making sure that devices which control critical infrastructure are secure is very important due to their role in many organizations. Making sure that SCADA systems have basic security measures to ensure their confidentiality, integrity, and availability can save many companies and organizations from disastrous consequences. Using the above mentioned collection and analysis techniques, other vulnerability assessment methods can be used as well to further secure devices.

5 Conclusion

Using the Internet of Things to access critical systems is a feature which many companies need in order to function properly. Unfortunately many of these companies rely on security through obscurity to protect these devices by placing them unprotected on the internet. Research has shown that it is very easy to locate specific critical devices on the internet and test the vulnerability of these devices in an automated manner. This can be done by classifying devices into different types, such as SCADA, collecting these devices, and testing the vulnerability of these devices in different ways. Vendors and companies must realize the risk of leaving unsecure devices on the internet as they can be easily exploited. Companies and vendors should have policies in place which do not allow for the use of default credentials. Policies also need to be made to ensure devices have an update schedule in order to ensure that they do not have any existing easily identifiable vulnerabilities. The best option is to not allow these devices to connect directly to the external internet, but rather only allow access through tools such as network address translation or access through a VPN. Making sure these devices are secure will help strengthen critical infrastructure, which millions of people rely on every day.

National Science Foundation Acknowledgement

The view expressed in this thesis are those of the author and do not reflect the official policy or position of the University of Arizona, or the National Science Foundation (NSF). This material is based upon work supported by the NSF under Grants

No. DUE-1303362 and No. SES-131463. I also acknowledge the funding and program support provided by the NSF Scholarship-for-Service (SFS) Program.

6 References

Bin, S., Yuan, L., & Xiaoyi, W. (2010). Research on data mining models for the internet of things. *2010 International Conference on Image Analysis and Signal Processing* (pp. 127-132). IEEE.

Bodenheim, R., Butts, J., Dunlap, S., & Mullins, B. (2014). Evaluation of the ability of the Shodan search engine to identify Internet-facing industrial control devices. *International Journal of Critical Infrastructure Protection, 7(2)*, 114–123.

Chi, S. (2011). Automated Object Identification Using Optical Video Cameras on Construction Sites. *Computer-Aided Civil and Infrastructure Engineering*, 368-380.

Evans, D. (2011). *The Internet of Things: How the Next Evolution of the Internet Is Changing Everything.* Cisco Internet Business Solutions Group (IBSG).

Fleury, A., Noury, N., & Vacher, M. (2009). Supervised classification of Activities of Daily Living in Health Smart Homes using SVM. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society*. (pp. 6099-6102). IEEE.

Genge, B., & Enăchescu, C. (2014). ShoVAT: Shodan-based vulnerability assessment tool for Internet-facing services. *Security and communication networks*, 1-18.

Hentea, M. (2008). Improving Security for SCADA Control Systems. *Interdisciplinary Journal of Information, Knowledge, and Management*, 73-86.

Idaho National Laboratory. (2011). *Vulnerability Analysis of Energy Delivery Control Systems.* Idaho Falls: Idaho National Laboratory.

Igure, V. M., Laughter, S. A., & Williams, R. D. (2006). Security issues in SCADA networks. *Computer and Security 25*, 498-506.

Internet Census. (2012). *Port scanning /0 using insecure embedded devices*. Retrieved from Internet Census 2012: http://internetcensus2012.bitbucket.org/paper.html

Matherly, J. (2015, Feburary). Shodan Data Interview. (E. Gross, Interviewer)

Nicholson, A., Webber, S., Dyer, S., Patel, T., & Janicke, H. (2012). SCADA security in the light of Cyber-Warfare. *Computer and Security 31*, 418-436.

OSVDB. (2015). *OSVDB Search*. Retrieved December 2014, from Open Sourced Vulnerability Database: http://osvdb.org/

Patton, M., Gross, E., Chinn, R., Forbis, S., Walker, L., & Chen, H. (2014). Uninvited Connections: A Study of Vulnerable Devices on the Internet of Things (IoT). *2014 IEEE Joint Intelligence and Security Informatics Conference*, (pp. 232-235).

Power Measurement. (2003). ION 7500 7600 User's Guide.

Robles, R. J., & Choi, M.-k. (2009). Assessment of the Vulnerabilities of SCADA, Control Systems and Critical Infrastructure Systems. *International Journal of of Grid and Distributed Computing*, 27-34.

Shodan. (2015). *Shodan*. Retrieved from Shodan: https://www.shodan.io/

Tillapart, P., Thumthawatworn, T., & Santiprabhob, P. (2002). Fuzzy Intrusion Detection System. *AU JT 6.2*, 109-114.

Tsai, C.-W., Lai, C.-F., Chiang, M.-C., & Laurence, Y. T. (2014). Data Mining for Internet of Things: A Survey. *IEEE Communications Surveys and Tutorials*, 77-97.

Williams, M. P. (2014). *Distinguishing Internet-facing ICS devices using PLC programming information.* Wright-Patterson Air Force Base: AIR FORCE INSTITUTE OF TECHNOLOGY.

Zmap. (2015). *Zmap*. Retrieved from Zmap The Internet Scanner: https://zmap.io/

Appendix A

Full Port List from Collected Results

Port Number	Number of devices	Port Number	Number of devices	Port Number	Number of devices	Port Number	Number of devices
7	397035	623	577628	3749	25330	8090	475059
11	6610	626	37334	3784	4181	8098	66188
13	167087	631	1354506	3790	351	8139	426
15	4905	666	2979	4022	30615	8140	398
17	57615	771	20681	4040	81528	8181	981
21	18020071	789	10903	4369	295204	8333	7970
22	35306131	992	35408	4443	11143	8443	837421
23	23255072	993	2307122	4444	211040	8649	100900
25	19592417	995	2239856	4500	8465839	8834	311
26	1483579	1023	31706	4949	36533	8888	13795
37	270430	1177	231	5000	1363482	9051	22998
53	11006474	1200	55914	5001	12735	9100	409060
67	19223	1234	956517	5008	12367	9151	22424
79	108768	1434	336854	5060	30807264	9160	6493
80	188845478	1471	26496	5094	19435	9200	140749
81	6170994	1604	15405	5222	482436	9943	3134
82	720356	1723	8960570	5353	19185	9944	25263
83	33679	1900	13242527	5357	2840891	9981	49903
84	18883	1911	48394	5432	1185609	9999	121990
88	17709	2067	14365	5560	41180	10000	40336
102	6211	2082	3097853	5632	212505	10001	8733
110	12778071	2083	9922	5900	1509139	10243	1162308
111	4822490	2086	2689854	5901	42012	11211	311720
119	55597	2087	704	5985	688847	16010	19353
123	3912207	2123	714867	5986	11249	20000	582
137	6613066	2181	48172	6000	166802	25565	211627
143	11663809	2222	1205281	6379	180113	27017	102211
161	12353878	2323	438714	6666	15730	28017	97155
195	7621	2375	18443	7071	639	32764	14070
389	981846	2376	360	7547	88749014	44818	7993
443	15333882	2404	409	7657	19618	47808	34350
444	42229	2455	19887	7777	1277459	49152	5804356
445	1203714	2628	32076	8000	243948	50100	4098
465	1200827	3000	252528	8069	110031	55553	749
500	11545875	3128	1041221	8080	25953862	55554	370
502	29026	3306	12502920	8081	4763	62078	861336
515	727045	3388	14805	8087	941	64738	97138
523	24049	3389	9132392	8089	27736		

Table 3: Full port list

* Ports with less than 50 devices associated with them were not included

Appendix B

Full List of Services Available in Shodan

Service	**Description**
asn	The autonomous system number (ex. "AS4837").
data	Contains the banner information for the service.
ip	The IP address of the host as an integer.
ip_str	The IP address of the host as a string.
Port	The port number that the service is operating on.
timestamp	The timestamp for when the banner was fetched from the device in the UTC timezone.
hostnames	An array of strings containing all of the hostnames that have been assigned to the IP address for this device.
domains	An array of strings containing the top-level domains for the hostnames of the device. This is a utility property in case you want to filter by TLD instead of subdomain.
location	An object containing all of the location information for the device
Opts	Contains experimental and supplemental data for the service. This can include the SSL certificate, robots.txt and other raw information that hasn't yet been formalized into the Banner Specification.
Org	The name of the organization that is assigned the IP space for this device
Isp	The ISP that is providing the organization with the IP space for this device.
Os	The operating system that powers the device.
uptime	The number of minutes that the device has been online.
link	The network link type. Possible values are: "Ethernet or modem", "generic tunnel or VPN", etc.
title	The title of the website as extracted from the HTML source.
html	The raw HTML source for the website.
product	The name of the product that generated the banner.
version	The version of the product that generated the banner.
devicetype	The type of device (webcam, router, etc.).
info	Miscellaneous information that was extracted about the product.
cpe	The relevant Common Platform Enumeration for the product or known vulnerabilities if available.

Table 4: Full list of Shodan services

Appendix C

SCADA Queries Used for Classifier

SCADA Queries Used for Classifier	
8600 ION	Rockwell Automation/Allen-Bradley Communications Adapter
AKCP	Rockwell Automation/Allen-Bradley Human-Machine Interface
Cimplicity	Rockwell Automation/Allen-Bradley Programmable Logic Controller
Citect	RuggedCom
ClearSCADA	S7-200
EIG Embedded Web Server	S7-300
HMI, XP277	scada
honeywell BNA	Scalance
honeywell Excel	Schneider Electric
ION 7550	Server: MoxaHttp/1.0
ION 7650	server: Niagara Web Server
IPC@CHIP title:Start	serverview
Modicon	Set-Cookie: SoftPLC=
niagara_audit-login	SIMATIC
opto 22	Simatic HMI
plc port:102	Simatic S7
PLC5	Simatic -S7 -HMI
Portal0000.htm	SLC-5
powered by SpiderControl TM country:DE	SSH-2.0-moxa_1.0
PowerLogic	title:adcon
PowerLogic PM800	WINCCFLEXIBLE
Rockwell Automation	wince Content-Length: 12581
	XA/21

Table 5: SCADA Queries Used for Classifier