

MIS 510: Cyber Analytics Project

Team: Never Off Guard

SUMEET BHATIA | AADIL HUSSAINI | SNEHAL NAVALAKHA | MO ZHOU

Table of Contents

Introduction	2
Hacker Web	3
Data Collection.....	3
Research Question 1	3
Data Analysis.....	3
Findings/Discussion.....	4
Research Question 2	5
Data Analysis.....	5
Findings/Discussion.....	5
Shodan	7
Research Question 1	7
Background Information	7
Data Collection/Analysis	7
Findings/Discussion.....	8
Research Question 2	9
Background Information	9
Data Collection/Analysis	9
Findings/Discussion.....	10
References	11

Introduction

As computers become more ubiquitous throughout society, the security of networks and information systems is a growing concern. There is an increasing amount of critical infrastructure that relies on computers and information technologies, and advanced technologies have enabled hackers to commit cybercrime much more easily now than in the past. Thus, it has become important for researchers to look into cyber security issues, such as botnet activity, digital forensics and malware analysis.

Our research seeks to contribute to the existing literature on cyber security by conducting analytics on data collected from two sources – Hacker Web and Shodan:

- Hacker Web: A collection of 18 major online forums where a community of international hacking enthusiasts participate in various kinds of discussions regarding cyber security topics.
- Shodan: A search engine for finding open and vulnerable ports and devices on the internet (“Internet of Things”)

We have looked at two main research questions for each of the aforementioned data sources. We present our data collection and analysis methods, before discussing our results and findings.

Hacker Web

Hacker Web is a collection of 18 major online forums where a community of international hacking enthusiasts participate in various kinds of discussions regarding cyber security topics. The individual posts on each of these forums, along with post date/time and authorship information, have been stored in a MySQL database. Hacker Web has been used in various research contexts and has proven to be a reliable source of cyber security-related data.

For our research purposes, we have chosen the four English forums to extract data from: “Elitehack,” “Hackhound,” “iCode” and “VCTool.”

Data Collection

We followed the following steps to collect our data from the four English forums via the Hacker Web MySQL database:

1. Downloaded and configured HeidiSQL
2. Connected to Hacker Web database using provided credentials
3. Run SQL queries (i.e., `SELECT * FROM [table] WHERE upper([column]) LIKE “%[KEYWORD]”`)
4. Converting the results of the queries into CSV files
5. Used MS Excel and IBM Many Eyes for various analytics

For the scope of our research, we chose to only look at the forum posts (i.e., “elitehackposts,” “hackhoundposts,” “icodeposts” and “vctoolposts”), since they were more likely to contain the keywords that we were looking for.

Research Question 1

Through browsing many cyber security-related literature, we found that there were two words that appeared very often: “victim(s)” and “target(s).” Thus, we wanted to analyze data collected from Hacker Web using these two keywords.

So our first research question is the following:

How frequent do posts with either of the two keywords (i.e., “victim(s)” and “target(s)”) appear on each of the four English forums? How does the frequency vary between the forums across time?

Data Analysis

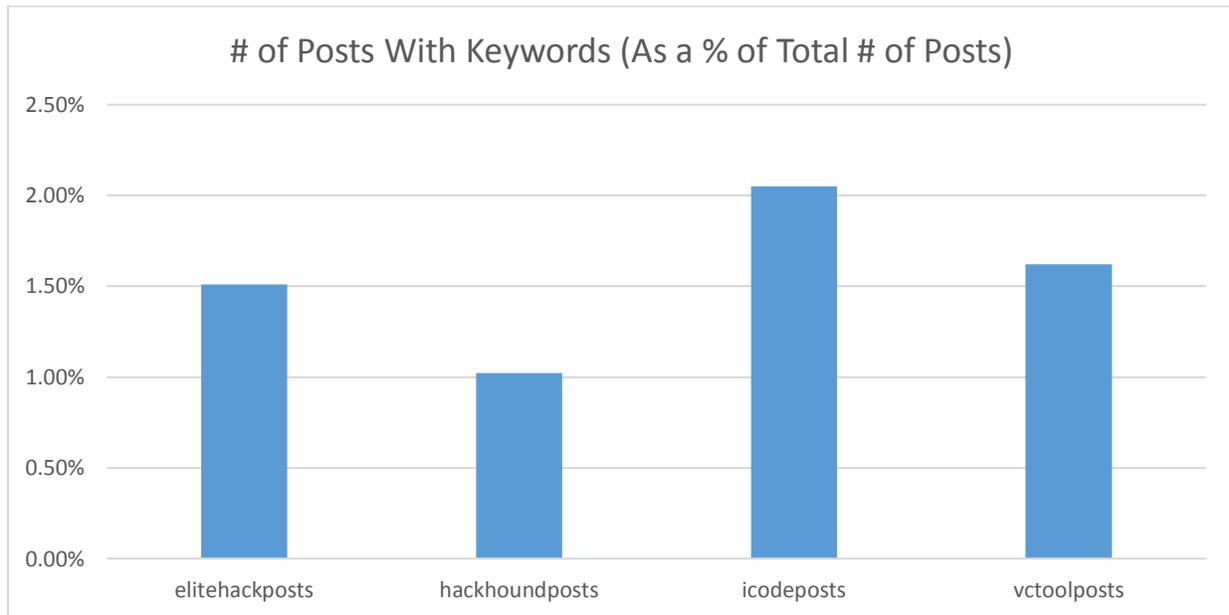
We first queried the four English forums individually to find the total number of posts (without any keywords) for each forum. Then we used the keywords “victim(s)” (SQL: “%VICTIM%”) and “target(s)” (SQL: “%TARGET%”) and queried all four forums individually to find the total number of posts with either of the two keywords. Finally, we calculated the percentage of total posts that contained either of the keywords.

Next, we wanted to look at how this percentage has changed over time, specifically over the most recent five-year period of 2009-2013. Because “Elitehack” and “Hackhound” are relatively new forums, there was

little to no data for years prior to 2012. Thus, we only compared the temporal trends for “iCode” and “VCTool.”

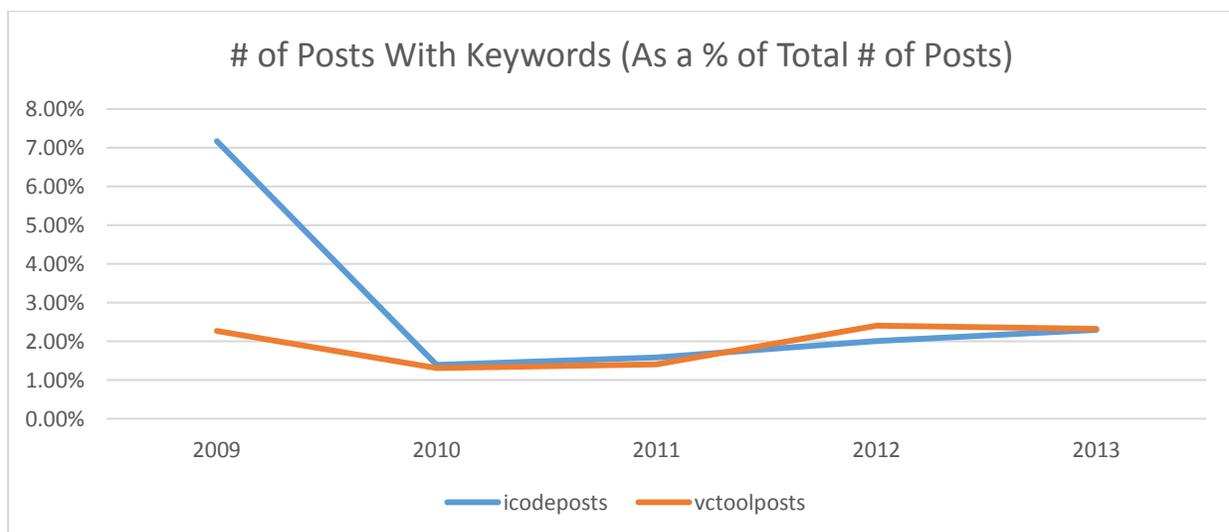
Findings/Discussion

The following chart depicts a comparison of this percentage for each of the four forums:



As we can see, the percentage of posts with either the keyword “victim(s)” or “target(s)” is above 1% for every forum, with “iCode” having the highest at 2.05% and “Hackhound” at 1.02%. It is likely that the reason “iCode” and “VCTool” have higher percentages than “Elitehack” and “Hackhound” because the former two forums were started earlier, so our queries resulted in a larger amount of data for calculating these statistics.

The following graph shows a comparison of these trends:



From the graph above, we see that there was an unusually high percentage with the two keywords in the “iCode” forum in 2009. Then, it dropped to a similar level as “VCTool.” In both forums, the percentage has generally grown over time, suggesting an increase in postings that mention “victim(s) and/or “target(s).”

Research Question 2

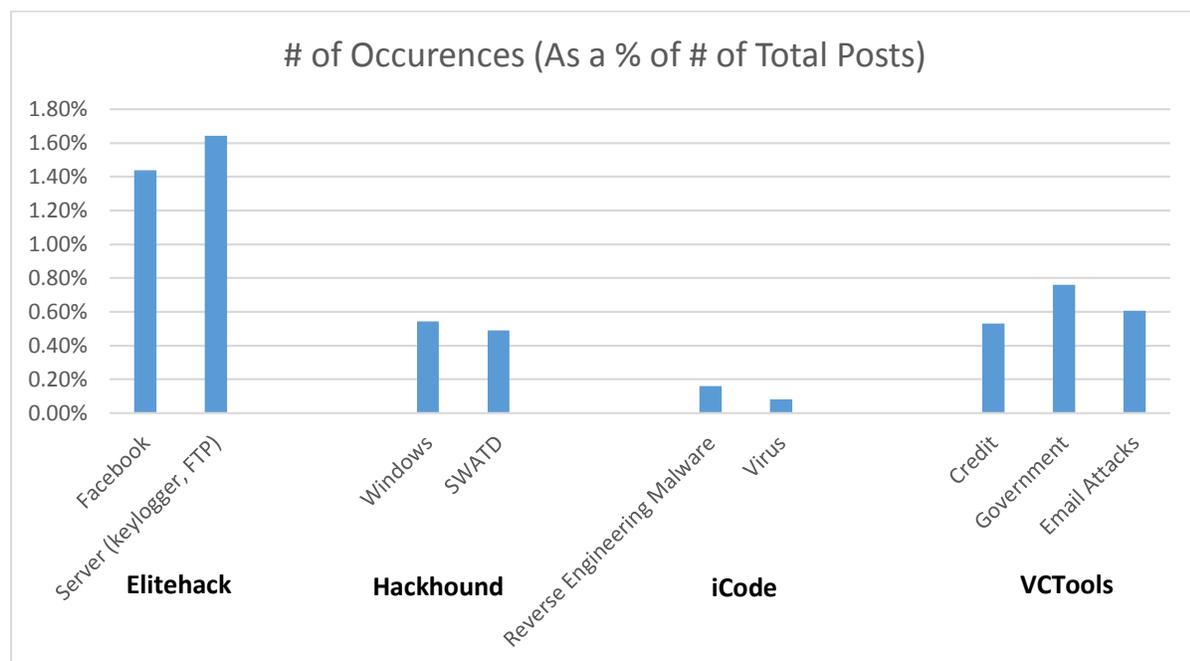
What are the most frequently mentioned topics within each forum and across all four forums?

Data Analysis

Again, for this question, we used the same query in Question 1 to find the total number of posts in each English forum. What is different is that we then used IBM Many Eyes to conduct a Word Tag analysis to find the most frequently mentioned topics on each forum. Finally, we calculated the percentage of total posts that contained each of these topics.

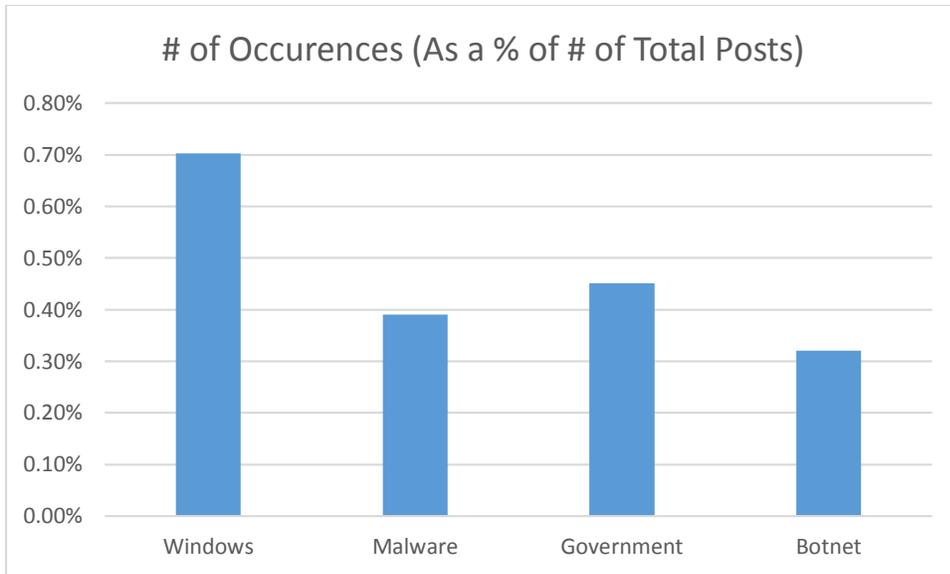
Findings/Discussion

The chart below shows the most frequent topics that are mentioned on each of the four English forums:



It is apparent from our analysis that each of these forums has its own “flavor.” Elitehack seems the most suitable for hackers who are more interested in social media hacking (e.g., Facebook), while iCode tends to serve the more traditional crowd who are more interested in malware and viruses. Those who enjoy talking about specific technologies and hacking techniques are most likely to visit Hackhound, but those who like to target the government or phish for personal information from individuals may like VCTools better.

Then, we aggregated the data from all four forums and found the most popular topics:



The results show that unsurprisingly, “Windows” is the most talked about, at 0.7%. The other hot topics are “government” (0.45%), “malware” (0.39%) and “botnet” (0.32%).

Shodan

Shodan is a powerful source for finding open and vulnerable ports and devices on the internet (“Internet of Things”). It interrogates ports, grabs the resulting banners and indexes the banners for searching. The different filters available for searching include: IP address, hostname, port, latitude and longitude, operating system, city, country, and device data. When used maliciously, Shodan can be used to exploit unprotected servers or devices. For research purposes, however, it can be a very useful tool to find potentially vulnerabilities in certain systems that are connected to the internet.

Research Question 1

Samsung has tried to go “SoLoMo” using its SmartTV. It has tried to integrate internet and Web 2.0 features with television sets. Our first research question on SmartTVs is divided into the following parts:

1. How many SmartTVs are publicly-facing and respond to Shodan’s search query? What is the geographical distribution of these SmartTVs and are all of them exploitable?
2. What percentage of SmartTVs is publicly visible where the Webkit vulnerability in the device could be exploited?

Background Information

Samsung SmartTV is essentially a Linux device configured with a Webkit-based browser used to load web pages and applications. Webkit is an open-source HTML rendering engine that has a significant presence in Google Chrome and Apple Safari browsers.

We chose Samsung SmartTV as our research device because it is a relatively new device in the market and it is not a big enough platform for a lot of people to contribute to its development. However, it uses a browser based on the Webkit technology, which exposes the device to a range of security exploits such as cross-site scripting attacks, denial-of-service attacks and unexpected application termination or arbitrary code execution. Thus, we feel it is important to research SmartTV’s vulnerability on the web.

Data Collection/Analysis

To get the Samsung SmartTV results, we first researched on Shodan’s website (<http://www.shodanhq.com/>) regarding various types of banner information the Samsung SmartTV gives us.

We found that the tag “Content-Length:345 Server:Swift1.0” was highly prevalent in the SmartTV banner. Then, we ran a query on Shodan using a Python

```
results = api.search('Content-Length:345 Server:Swift1.0' , page=1, limit=3000, offset=None)
# Show the results
print 'Results found: %s' % results['total']

for result in results['matches']:
    data = result['region_name']
    data2 = result['city']
    data3 = result['org']
    data4 = result['country_name']
    data5 = result['os']
    data6 = result['port']

    worksheet.write(row, col, data )
    worksheet.write(row, col+1, data2 )
    worksheet.write(row, col+2, data3 )
    worksheet.write(row, col+3, data4 )
    worksheet.write(row, col+4, data5 )
    worksheet.write(row, col+5, data6 )

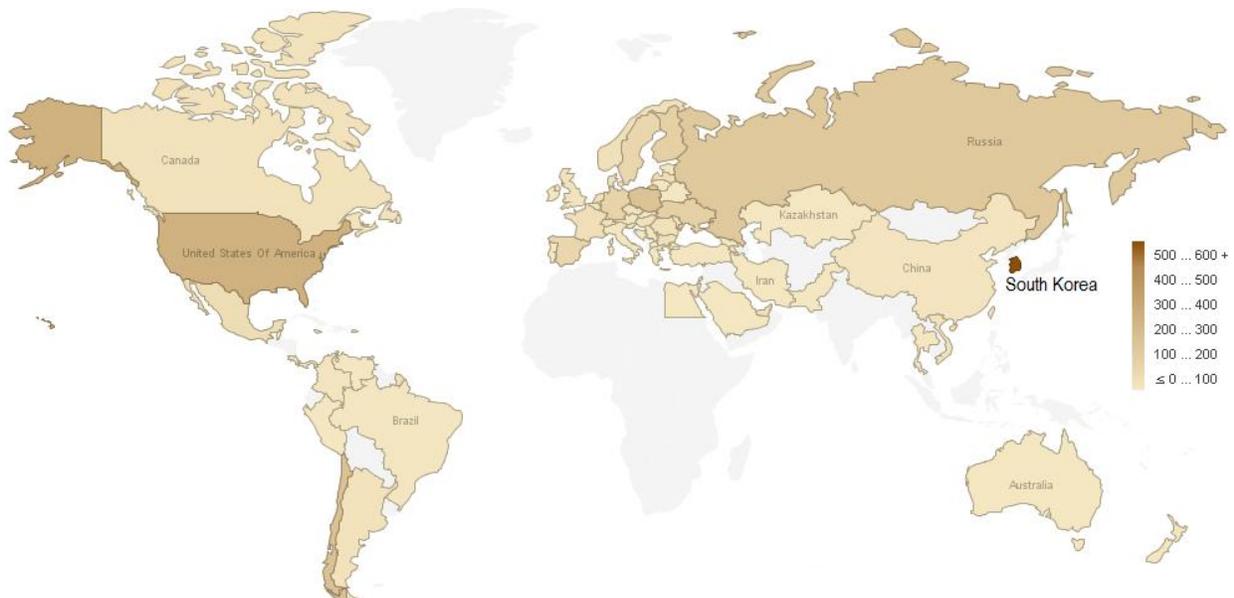
    row += 1
```

script (as seen in the figure), we retrieved 350,968 records. These were the number of SmartTV that responded to the Shodan query. For our research purpose, we limited our search to 3,000 devices. We assume that the 3,000 records is a random sample, giving us an approximate look and feel of the entire dataset. Finally, we used these records to analyze the geographical distribution of the devices and their exploitability (based on the operating port).

Findings/Discussion

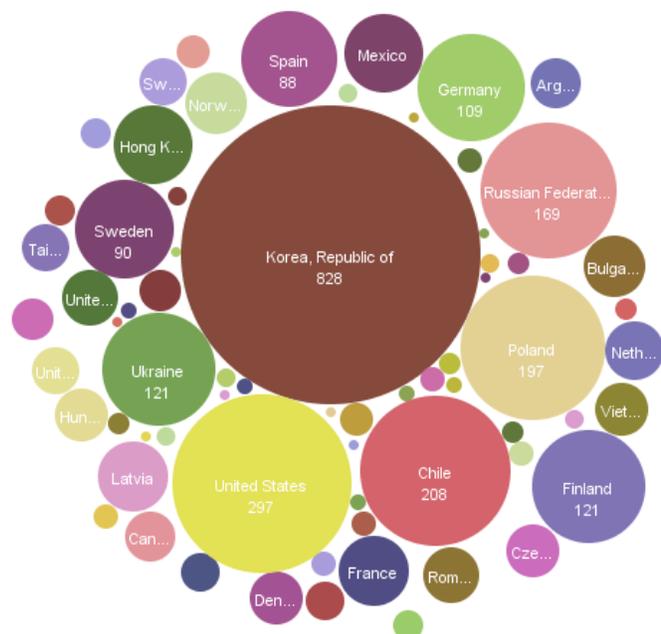
Research Question 1 Part 1

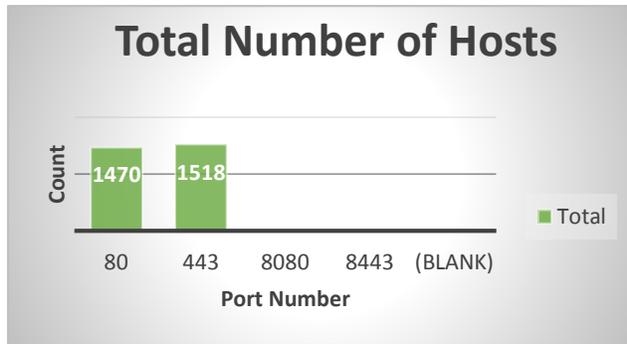
Our analysis shows that not surprisingly, the Republic of Korea (where Samsung is based) dominates the list of countries. It is followed by United States and Chile is ranked #3 as seen in the figures below.



Other countries where the number of devices exceeds 100 are Ukraine, Germany, Russia, Poland and Finland.

Digging a level deeper, we found that although the majority of the SmartTVs work on Port 443, which greatly enhances security, a large number of them gave access on Port 80, which makes them quite vulnerable. The figure below shows the breakdown of the ports given access to out of the 3,000 search results.





Research Question 1 Part 2

We analyzed Samsung's financial data for its SmartTV sales. Statistics show that Samsung has sold around 12 million SmartTVs as of Q1 2013. As mentioned earlier, we tracked down 350,968 devices from our Shodan query. This implies that there at least 2.92% of the devices are publicly visible.

The Samsung SmartTV has features such as built-in webcams, web-based applications and browsers. The Webkit vulnerability exposes hundreds of thousands of SmartTVs, potentially giving hackers access to multiple ways to get sensitive information from the users with a single hack.

Research Question 2

How vulnerable are the traffic signal systems in the United States? Which are the cities that are most vulnerable to getting their traffic signal systems hacked?

Background Information

While the internet has enabled many of the public communication systems to be accessed digitally, there are growing concerns about their lack of security. Engineers recently hacked into the Traffic Signal System in Los Angeles. The engineers programmed the signals so that red lights for several days would be extremely long on the most congested approaches to the intersections, causing heavy gridlock. Such incidents raise the following security concerns:

1. Traffic Signal Systems can be hacked, which will affect the normal flow of traffic
2. The traffic cameras used for monitoring traffic can be hacked, which is a privacy and security risk

Our objective is to find out how vulnerable the Traffic Signal Systems are in the United States, specifically in which cities.

Data Collection/Analysis

We searched the Shodan database to find information about servers for Traffic Signal Systems. We used two tags that were most prevalent in the header information such servers. Below is our detailed process:

1. Searched for header keywords in the Shodan Database
2. Wrote a Java application to extract the data row by row return it to Python
3. Wrote a loop using Python to input and store the data row by row in MS Excel
4. Used the results in output for analysis

```

try:
    #shodansssearch = 'atz_executive'
    results = api.search('atz_executive', page=1, limit=3000, offset=None)
    # Show the results
    print 'Results found: %s' % results['total']

    for result in results['matches']:
        data = result['region_name']
        data2 = result['city']
        data3 = result['org']
        data4 = result['country_name']
        data5 = result['os']
        data6 = result['port']

        worksheet.write(row, col, data )
        worksheet.write(row, col+1, data2 )
        worksheet.write(row, col+2, data3 )
        worksheet.write(row, col+3, data4 )
        worksheet.write(row, col+4, data5 )
        worksheet.write(row, col+5, data6 )

        row += 1

except Exception, e:
    print 'Error: %s' % e

```

```

try:
    #shodansssearch = 'atz_executive'
    results = api.search('Content-Length:345 Server:Swift1.0', page=1, limit=3000, offset=None)
    # Show the results
    print 'Results found: %s' % results['total']

    for result in results['matches']:
        data = result['region_name']
        data2 = result['city']
        data3 = result['org']
        data4 = result['country_name']
        data5 = result['os']
        data6 = result['port']

        worksheet.write(row, col, data )
        worksheet.write(row, col+1, data2 )
        worksheet.write(row, col+2, data3 )
        worksheet.write(row, col+3, data4 )
        worksheet.write(row, col+4, data5 )
        worksheet.write(row, col+5, data6 )

        row += 1

```

Findings/Discussion

First, we used the tag “atz executive” to get information about all the systems on Port 23 (i.e., FTP). This gave us information of all “ATZ Traffic Signal Systems” using the Operating systems C Executive. We were able to get access into the system and see live images. We found 216 records and then we sorted them by city and state. The analysis



showed that Louisiana had the most number of vulnerable systems, and the cities of Metairie and New Orleans in that state ranked in the top 2 across the country.

Then, we used the tag “Content-Length: 2861 Cache-Control: max-age=86400” to get information on Port 8080. We found that “Content-Length: 2861” corresponds to PIPS technology, which provides automatic license plate recognition. These devices offer a telnet connection for management that does not require authentication. We were able to observe the number plate information and live images. We were also able to modify the configuration settings. Again, we found that the city of Metairie in Louisiana had the most vulnerable Traffic Signal Systems.



References

- (n.d.). *Shodan Introduction* [PowerPoint slides]. Retrieved from <http://ai.arizona.edu/mis510/>
- Benjamin, V. (2014). *Cybersecurity Research Overview* [PowerPoint slides]. Retrieved from <http://ai.arizona.edu/mis510/>
- Freamon, D. *The Darius Freamon Blog*. Retrieved from <http://dariusfreamon.wordpress.com/tag/traffic-management/>
- Grad, S. (2009, December 1). *Engineers who hacked into L.A. traffic signal computer, jamming streets, sentenced*. Retrieved from <http://latimesblogs.latimes.com/lanow/2009/12/engineers-who-hacked-in-la-traffic-signal-computers-jamming-traffic-sentenced.html>
- Roberts, P. (2013, August 1). *Samsung Smart TV: Like A Web App Riddled With Vulnerabilities*. Retrieved from <https://securityledger.com/2013/08/samsung-smart-tv-like-a-web-app-riddled-with-vulnerabilities/>
- Segall, L., Fink E., *Samsung Smart TV security flaw let hackers turn on built-in cameras*. (2013, August 1). Retrieved from <http://www.wptv.com/news/science-tech/samsung-smart-tv-security-flaw-let-hackers-turn-on-built-in-cameras>
- Strategy Analytics. (2013, July 24). *Samsung Leads with 26 Percent of Global Smart TV Market Share in Q1 2013*. Retrieved from <http://www.strategyanalytics.com/default.aspx?mod=pressreleaseviewer&a0=5400>