Malicious WiFi Networks: A First Look

Andrew Zafft
Computer Science Department
Worcester Polytechnic Institute
Worcester, MA, USA

Emmanuel Agu
Computer Science Department
Worcester Polytechnic Institute
Worcester, MA, USA

Abstract— **WiFi networks have become ubiquitously deployed. Anecdotal evidence suggests that wireless networks are less secure than wired networks. However, very little quantitative data exists to characterize the insecurity levels of WiFi networks. In this paper, we take a first look at city-level WiFi security statistics for eighteen cities within the United States. For the purposes of this study, we define unsecure WiFi networks as those that do not use WEP, WPA2 or any WiFi security standard. We found that on average, 45 percent of WiFis studied were insecure, with Miami, FL having 81 percent of WiFi networks unsecure. We also found that Texas had several cities with a high number of blacklisted IP addresses. In the end, we found no strong correlation between WiFi insecurity rates and blacklisting rates, which we attribute to inadequate education of the inhabitants of cities on how to report malicious activity so that perpetuators can be blacklisted. Finally, we found that the percentage of secure WiFi networks in cities with municipal WiFi networks was comparable to that of cities without municipal WiFi networks.**

Index Terms— **Wireless measurement, performance, security, wireless routers**

I. INTRODUCTION

Many household devices that previously ran independently are now internetworked, allowing people to set thermostats, turn lights off and on, and even unlock car doors from anywhere in the world. The future is here, and the home environment is constantly connected to the outside world. Wireless networks have been a major driver in today's highly networked world. However, as wireless networks have become widely deployed, security concerns are also rising. It is well known that wireless networks are more vulnerable than wired networks. Wireless signals leak beyond buildings in which access points are installed and intruders can easily pick up these signals from parking lots or nearby buildings. Consequently, intruders can easily gain access if wireless networks are not secured using Wired Equivalent Privacy (WEP), WiFi Protected Access II (WPA2) or other security standard.

Wireless security is further complicated by the advent of open wireless networks. Many businesses provision wireless access to its customers and frequently these networks are not secured, and in some cases do not even require customers to authenticate or identify themselves.

Much of the insecurity of wireless networks has largely been anecdotal. Very little work has been done to quantify and understand the levels of security of existing WiFi deployments. As a result, many questions remain unanswered. What percentage of deployed WiFi networks is unsecured? Is malicious activity more prevalent on wireless networks than on wired networks? When detected, is malicious activity promptly reported and malicious networks blacklisted? Is malicious activity on WiFi networks more prevalent in some cities than others? These questions and many others remain unanswered.

One of the core themes of this paper is characterizing the levels of security in WiFi networks. We investigated WiFi security levels in eighteen US cities, selected primarily because they had a high number of blacklisted IP addresses. The intent of this paper is to characterize these cities and quantify what percentage of their WiFi networks is secure. Blacklisting rates are also examined and compared to WiFi insecurity rates. This paper examines Internet availability and attempts to determine if WiFi availability and WiFi network security are useful predictors of malicious WiFi activity.

We found that on average, 45 percent of WiFis studied were insecure, with Miami, FL having 81 percent of WiFi networks insecure. For the purposes of this study, we define insecure WiFi networks as those that do not use WEP, WPA2 or any WiFi security standard. We also found that Texas had several cities with a high number of blacklisted IP addresses. We found a limited correlation between WiFi insecurity rates and blacklisting rates. In part, we attribute this to inadequate education of the inhabitants of cities on how to report malicious activity so that perpetuators can be blacklisted. Finally, we found that the percentage of secure WiFi networks in cities with municipal WiFi networks was comparable to that of cities without municipal WiFi networks.

The rest of the paper is as follows. Section 2 discusses related work. Section 3 describes our methodology for gathering and analyzing data. Section 4 presents our results. Section 5 describes future work. Section 6 presents our conclusion and Section 7 acknowledges the help of key individuals that helped us in this work.

II. RELATED WORK

This paper was loosely inspired by the work of Andre Kalafut, Craig Shue and Minaxi Gupta [1]. While their paper focused on detecting and characterizing the distribution of abnormal behavior in Autonomous Systems (ASes) on the Internet, we applied similar concepts to public WiFi environments looking at the distributions of malicious

978-1-4673-2129-7/12/$31.00 ©2012 IEEE

or unsecure WiFi networks and investigating their correlations with black listed IP addresses. Also related to our project is the work of Hu *et al* [2], which presented the city-wide spread of malware on WiFi networks as an epidemiological study. This study used a simulated environment to investigate the potential prolific explosion of malware in WiFi congested cities while our study focused on evaluating statistics gathered on real world cities.

Geolocation, or the attempt to identify the physical longitude and latitude of a (WiFi) location was a core theme of this paper. Many different applications utilize geolocation, such as mobile applications and GPS devices. BambaGueye, et al [3] performed research using geolocation to predict the physical location of a server or client, and then seeing if this physical location was where it was expected to be. In contrast our paper did not propose to use geolocation to prevent illegal activity, merely to make the gathering of statistical information an easier affair.

III. METHODOLOGY

There were three distinct phases to this project. First was the data acquisition phase where we identified and obtained the desired datasets. Second was the data manipulation phase in which we modified the data to facilitate analysis. Finally was the data analysis phase where we constructed the charts and drew our observations.

A. Data Acquisition Methodology

No new data was generated during the course of this project. All data sets came from existing vendors, originating from both commercial and non-commercial sources. This created various difficulties as commercial ventures sell blacklisted IP address data to generate a profit. Our methodology had multiple steps. First, we obtained a listing of blacklisted IP addresses. Next we used geolocation techniques to associate the blacklisted IP addresses with a physical community or city. Finally, we acquired wireless statistics for a given area. Each of these sub-tasks was tackled separately.

Obtaining blacklisted IP addresses: While attempts were made via email to secure free blacklisted IP addresses from several companies, we were mostly unsuccessful. In the end, we obtained a copy of global blacklisted IP addresses from a German website, UCEPROTECT-Network[1]. This information was not as up-to-date nor as complete as the for-profit lists but was sufficient for this paper. Since this was the primary source for quantifying illegal activity, it proved to be invaluable.

Associating blacklisted IP addresses with geographic locations: We defined a community by mapping a list of blacklisted IP addresses to a geographic community. While our analysis required a static mapping, in practice IP addresses were not always assigned to one defined geographic area. However, IP addresses often were assigned to organizations in blocks or assigned to residences through fixed commercial ISPs. As a rough approximation, we assumed IP addresses assigned through ISPs were fixed within an individual city. While simplistic, this assumption suffices for this paper but remains an area of improvement in the Future Work section. As with blacklisting sites, there were both commercial and publicly available sources for mapping an IP to a geographical location. Maxmind provided one such tool named GeoIP[2]. The GeoIP tool contained a database of IP addresses and their corresponding global location information, namely city, state, country, longitude and latitude. While the location data was not always supplied for all countries worldwide, most of the US locations contained complete information making this tool sufficient for use in this study.

Obtaining security statistics of WiFi deployments: We wanted to view statistics of WiFi deployments such as percentage of secure access points, and the number of blacklisted IP addresses occurring within specific WiFi deployments for cities. We wanted to be able to investigate the correlations between network availability and wireless security usage. To get WiFi statistics, we contacted WiGLE[3], a company that has been capturing wireless access point location information and their security settings through war-driving applications on a large scale for nearly a decade. While WIGLE did not provide raw data directly, or allow specific queries, summary data was provided.

Gathering information about cities studied: In order to get a better sense of what types of cities would be investigated in this paper, we needed information about city attributes such as population and education level. We gathered some general statistics through the US Census Bureau's[4] recent 2010 census. Additionally, we wanted to investigate the effect of free WiFi deployments funded by a city (or municipal WiFi deployments). We reviewed Wikipedia's[5] website for determining if a city had a municipal WiFi. While we would have liked to use a more authoritative source for determining the availability of WiFi deployments, a central source of municipal WiFi data was not available. The Future Work section details our plans to use an authoritative database of WiFi deployment statistics.

B. Data Manipulation Methodology

Gathering the information was just the first step. Next we manipulated the data in multiple steps. The data that was acquired was largely global in nature. As was noted above, our WiFi security analysis focused on cities in the United States. As can be seen in **Error! Reference source not found.** below, the US ranked fairly low in terms of total blacklisted IP addresses (31st), coming between Italy and the Philippines. Despite this, the US was chosen as its culture and Internet usage was the most familiar to the authors. From the US dataset we chose a subset of twenty cities to examine in detail. Fifteen cities were chosen because they

[1] http://www.uceprotect.net/en/index.php

[2] http://www.maxmind.com/

[3] http://wigle.net/

[4] http://www.census.gov/

[5] http://www.wikipedia.org/

had the highest number of blacklisted IP addresses (initially not normalized by population or number of IP addresses). We then hand picked five additional cities: Riverside, CA, Tempe, AZ, Herndon, VA, Spokane, WA and Schenectady, NY. We felt these cities would be interesting to examine in detail.

TABLE I. COUNTRIES RANKED BY TOTAL BLACKLISTED IPs TAKEN FROM UCEPROTECT – NETWORK DATASET

Ranking	Country	Ranking	Country
1	India	30	Italy
2	Vietnam	31	United States
3	Brazil	32	Philippines
4	Russia		
5	Pakistan		

A difficult aspect of defining a community was identifying it in such a way that was consistent and expressible between the different datasets. In the end, we chose to use longitude and latitude coordinates to define the extents of a city or community. This was convenient given that the MaxMind GeoIP dataset included this value. For simplicity we limited all cities to be bound by a rectangular "box" of coordinates.

We associated the blacklisted IP addresses to their source city using the GeoIP dataset. For each city, we grouped the data by the city name field. From this, we determined the minimum and maximum longitude and latitude coordinates to form the bounding box for the city. As some GeoIP records did not have a city name linked to them, we took the extra step to recalculate the number of blacklisted IP addresses that fell within the bounding box by using the calculated longitude and latitude. We acknowledge that in reality, the geographic shape of a city may not be rectangular. We did not attempt to verify the longitude and latitude coordinates for a city, we simply assumed that each city fit within the calculated rectangular box. Other ways to more accurately specify city extents are described as Future Work.

Early trials showed that database comparisons using the *bigint* formats, which held the IP addresses in the GeoIP dataset, were possible however exorbitant amounts of time were required for each comparison. Elaborate indexing did not bring the running time to a reasonable level. Eventually, we were able to speed up the comparisons by dividing the IP address ranges into subsets and performing comparisons on these limited slices. Sub-dividing in this manner reduced the execution time to a manageable amount.

For consistency, the same bounding box used to specify city extents in the blacklist dataset was used to calculate total number of available IP addresses in the GeoIP dataset. To retrieve WiFi deployment statistics, the city-specific bounding boxes were passed along to WiGLE.net, and they provided the summarized WiFi deployment statistics to us.

C. Analysis Methodology

The data manipulation phase involved processing our data to generate data to answer our research questions. As this was primarily an investigational study, we generated derived statistics of fields such as IP address availability, WiFi network security, and number of blacklisted IP addresses. Some of these fields were normalized by dividing raw numbers by the population of a city or the number of available IP addresses. These statistics are used to plot the charts in our observation section. A table of the full set of data can also be seen in Table 2 below.

IV. RESULTS

We now describe our observations from the data.

Observation 1: A few cities had IP densities that were far above the median.

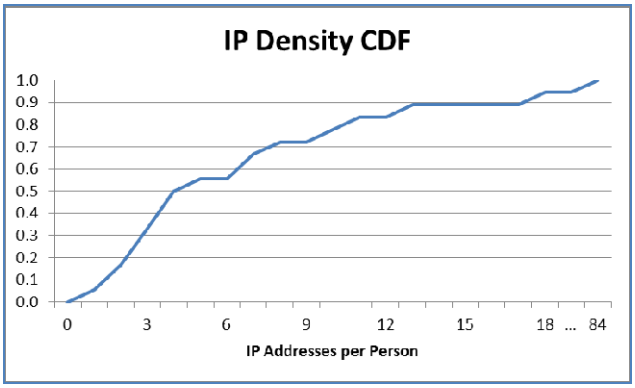

Figure 1 IP Addresses per Person CDF

Figure 1 is a cumulative distribution function (CDF) of the number of IP addresses per person in the investigated cities. The median value was 4.78 IP addresses per person. The values ranged from just under 1 (Arlington, TX) to just over 83 (Herndon, VA). The two cities with the highest IP address per person were Herndon, VA (83) and Atlanta, GA (17.22). We speculate that Herndon's high IP density may be due to the concentration of technology companies and government offices in that city.

TABLE 2: CONSOLIDATED DATASETS USED IN THIS EXPERIMENT

City	Total WiFi Networks	Secure WiFi Networks	Security Rate	Population	IP Addresses	Blacklisted IPs	IP Density	WiFi Density
Arlington, TX	20,739	13,492	0.65	365,438	317,029	59	0.87	0.0654
Atlanta, GA	149,164	82,493	0.55	420,003	7,230,973	165	17.22	0.0206
Chicago, IL	589,629	357,441	0.61	2,695,598	9,198,939	213	3.41	0.0641
Dallas, TX	179,413	90,728	0.51	1,197,816	5,730,943	449	4.78	0.0313
Denver, CO	173,289	105,588	0.61	600,158	6,043,433	60	10.07	0.0287
Herndon, VA	4,448	2,097	0.47	23,929	1,995,040	107	83.37	0.0022
Houston, TX	645,566	287,917	0.45	2,099,451	25,909,387	385	12.34	0.0249
Las Vegas, NV	234,175	133,123	0.57	583,756	1,819,279	97	3.12	0.1287
Los Angeles, CA	421,312	254,907	0.61	3,792,621	10,439,952	240	2.75	0.0404
Miami, FL	127,516	36,788	0.29	399,457	2,885,771	223	7.22	0.0442
New York, NY	397,041	254,891	0.64	8,175,133	23,019,797	252	2.82	0.0172
Orlando, FL	60,661	33,551	0.55	238,300	1,646,997	96	6.91	0.0368
Philadelphia, PA	151,851	98,655	0.65	1,526,006	4,710,118	108	3.09	0.0322
Riverside, CA	19,940	14,218	0.71	303,871	541,425	82	1.78	0.0368
San Antonio, TX	173,421	90,547	0.52	1,327,407	2,486,790	177	1.87	0.0697
Schenectady, NY	3,922	2,726	0.70	66,135	620,539	9	9.38	0.0063
Spokane, WA	10,548	7,398	0.70	208,916	428,249	9	2.05	0.0246
Tempe, AZ	12,024	4,889	0.41	161,719	1,077,266	89	6.66	0.0112

Observation 2: Most cities studied had about the same percentage of secure WiFi access points.

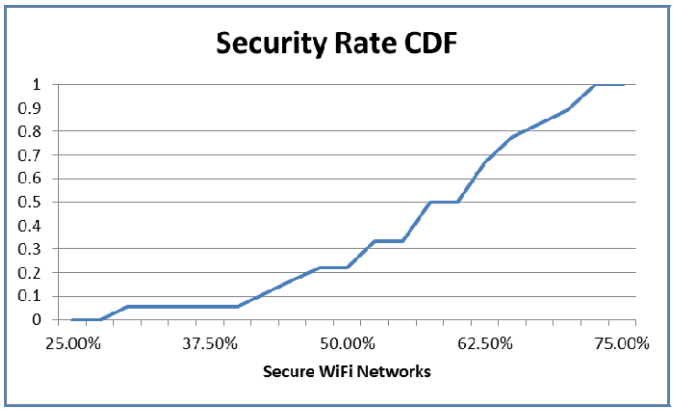

Figure 2 WiFi Network Security CDF

Figure 2 is a CDF of the percentage of all WiFi access points that were secure in each city. The security rate is determined by dividing the number of WiFi networks in a city that is secured with any type of wireless security protocol (WPA, WPA2, AES, TKA, etc.) by the total number of WiFi networks in that city. Most of the sites were closely clustered around the average (57%) with two-thirds of all sites falling within one standard deviation of the average. However, Miami, FL had a security rate that was well below the average. It is a little unclear why Miami has such a low security rate. This may be due to the presence of a large municipal level WiFi network within the city, increasing the number of unsecured WiFi networks.

Observation 3: The ratio of secure to unsecure WiFi networks varied.

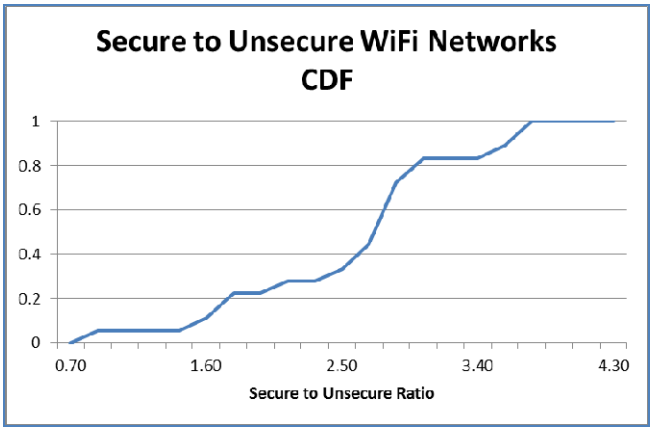

Figure 3 Secure to Unsecure ratio of WiFi Networks CDF

In Figure 3 above, we measure network security using a different metric. In this figure, we are comparing WiFi networks with security measures (defined in Figure 2) divided by WiFi networks with no security measures, which we call 'unsecure' networks. In this figure, a ratio larger than one means that secure networks are more prevalent than unsecure networks. The majority of sites are clustered near the average (2.53), with most being slightly above or slightly below the average. It is interesting to note that only one city (Miami, FL) has more unsecure networks than secure networks.

Observation 4: Security rates varied within the largest networks.

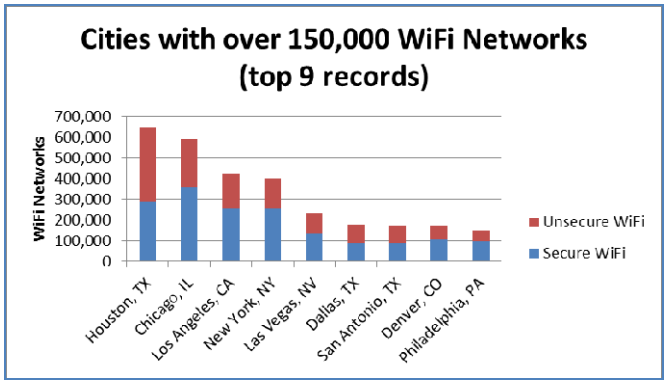

Figure 4 Secure and Unsecure WiFi Network Breakdown in Top 9 Cities

Figure 4 shows a further look into the insecurity of the nine largest WiFi cities (cities having more than 150,000 WiFi networks). The top site, Houston, had the third lowest security rating with 645,566 WiFi networks, and only 287,917 (45%) contained any detectable security measures. Other sites within the top nine were closer to the average security rate (57%). Interestingly enough, the cities selected in this study were chosen because of the number of blacklisted IPs found in the UCEPROTECT dataset, and 22% of the cities used in this study were located in Texas. At the same time, cities in Texas held 3 of the top 7 spots when sorted by total WiFi networks. In total, Texas accounted for 30% of all WiFi networks examined in this study. There could be several reasons for the high prevalence of WiFi networks here. State programs could be highly supportive of wireless networks. Or it could be that the war-diving tools which are used to populate WiGLE.net's database has a higher adoption rate in Texas. Or this could indicate a high "tech-savvy-ness" of the local population.

Observation 5: WiFi density and IP density have standard S-curves.

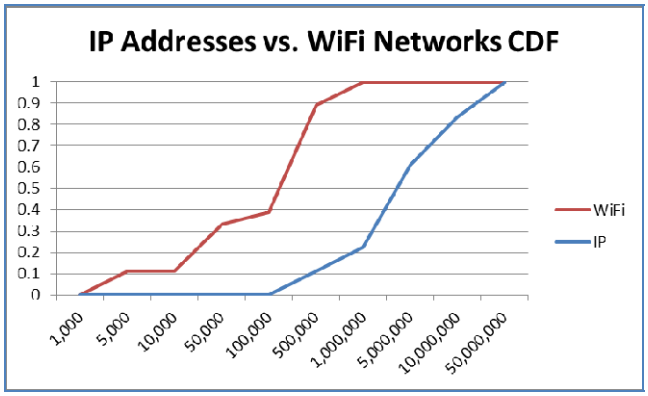

Figure 5 Comparison of IP Density and WiFi Density CDFs

Figure 5 is a combined CDF of the density of IP addresses and WiFi networks. In this study, the minimum number of IP addresses per city was 317K, and the maximum topped out at 25.6M. For WiFi networks, the minimum was 3.9K per city and the maximum was 645K. Two interesting points are Las Vegas, NV and Herndon, VA. Both have similar numbers of IP addresses (between 1.8M and 1.9M), but there was a dramatic difference in the number of WiFi networks. Herndon had 4K networks and Las Vegas had 234K networks. The average ratio of IP addresses to WiFi networks in this study is 63:1. Las Vegas's ratio was 7.8:1 which means it has nearly ten times as many WiFi networks as other cities in this study. Herndon's ratio was 449:1, meaning it has one-seventh as many WiFi networks as the average city in this study. There could be several reasons for this dramatic difference. As with the previous figure, this could again be due to the method used to detect WiFi networks. Las Vegas probably has a high level due to all the hotels, convention halls and other assorted resort destinations seeking to provide WiFi access to travelers. Herndon, on the other hand, is a suburb that has more technology businesses which have lots of IP addresses and where there is less of a need to lure of travelers with open networks.

Observation 6: There was a slight correlation between unsecured networks and blacklisted IP addresses.

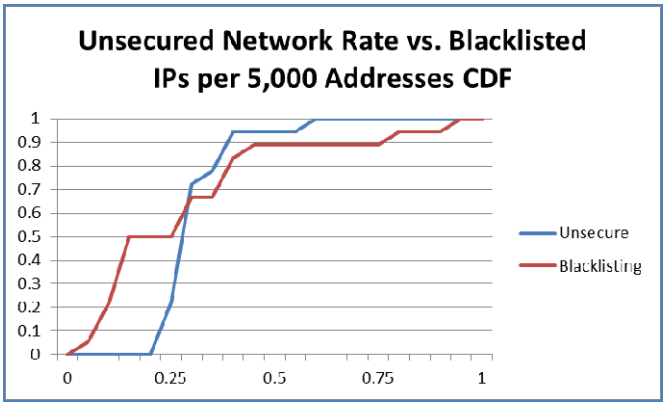

Figure 6 Unsecured WiFi Networks overlaid with Blacklistings per 5,000 IP Addresses CDF

Figure 6 is a CDF comparing the rate of unsecured WiFi networks versus the blacklisting rate per 5,000 IP addresses. You will notice that there is a cross-over point between the two charts at the 0.25 position on the x-axis. In this chart, we can see a limited correlation with unsecure network rates and the number of blacklisted IP addresses moving similarly. This appears to verify that communities with less secure WiFi networks would have more malicious activities and hence more blacklisted IP addresses.

Observation 7: The security rates of cities with municipal WiFi were comparable to cities without municipal WiFi.

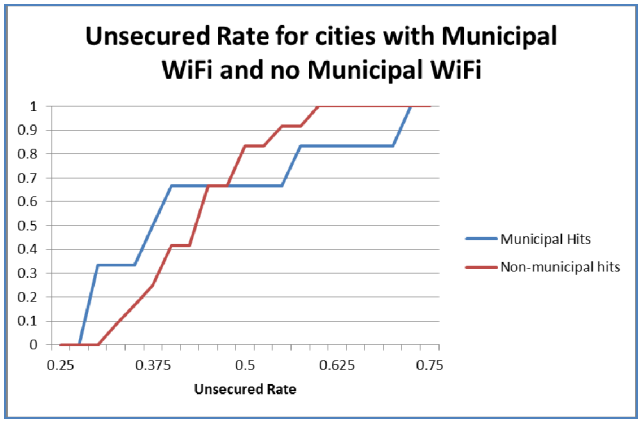

Figure 7 Security Rate of cities with municipal WiFi overlaid with security rate of cities without municipal WiFi

Figure 7 above shows that the security rates of cities with municipal was comparable to that of cities with no municipal WiFi. This could be because the number of access points in the city's municipal WiFi was small compared to the total number of WiFi in each city.

V. FUTURE WORK

Several areas of improvement remain for this project including many which could improve the overall results or reduce any inherent errors. First, there were only 18 cities examined in this project. Ideally, this set size should be closer to 50 or 60 to get a good sampling of wireless and blacklisting behavior. Additionally, more information could be obtained if the WiFi dataset provided by WiGLE.NET included data on security protocols. Also linking IPs to their parent Autonomous Systems and including the network type (commercial, government, private) would be interesting.

There also needs to be more examination in the people involved in this experiment. Specifically, there should be an investigation of the correlation between security or blacklisting rates and the "technical savvy-ness" of regions. The rationale for this exploration would be because the decision by an access point owner to secure it or the decision to report any observed malicious activity depends heavily on their technical knowledge. Essentially, blacklisting IP addresses is a reporting based system, which relies on technically aware users. No datasets discovered contained any information about the entity reporting the illegal activity. Without this information, only half of the blacklisting picture can be analyzed. Furthermore, there was little to no indication regarding the lifespan of a blacklisted IP address. Are these IP addresses blacklisted forever? Do they (or should they) be removed from this list after a fixed period of time? These are questions that need to be asked as they speak to the core functionality of the blacklisting process.

Better tools for localizing IP addresses are also needed. Is assigning an IP to an AS sufficient to link the IP to the city? Is there a better method than that which was used in this study? Additionally, what is the proper way to define the extents, the longitude and latitude, of a city? In this project, we used a rectangular bounding box yet few cities are exactly rectangular. Some additional thought here could improve the results of any follow along experiments.

Similarly, this study could be expanded from a US-centric study to a global study. The datasets collected contained global data, and therefore it would be a fairly simple affair to expand this study. Eventually, this study could spawn into an online database into which various entities could submit their own data. Such a database would be searchable and advance the overall understanding of the security of WiFi networks.

VI. CONCLUSIONS

There are several conclusions that we can draw from this study. First and foremost is that education should be on the frontline of any effort to combat illegal activities. From tips for identifying illegal behavior to available avenues to report these activities when they do occur, the general public should be informed to both acknowledge that there is a problem, and provide insight on how this problem might be alleviated. In lock step with this should be procedures to educate the population on WiFi security. Our study found that on average 45% of WiFi networks were insecure. In this day and age, no wireless router should be accidentally unsecured.

In terms of gauging where public policy should focus, cities that have less than 75% of all routers secured should focus on security education. Furthermore, cities with a high IP density should focus on educating the public about procedures to identify and blacklist illegal activity.

Above and beyond this study, however, cities should regularly monitor the online activity of their residents. While this certainly should not be on a granular level (i.e. household), there should still be statistical analysis to gauge the quality of this communication medium. Because internet security is maintained entirely on the local level, local attention must be paid to prevent abuse.

ACKNOWLEDGMENT

We would so like to thank Bobzilla from WiGLE.NET. His effort provided all the wireless datasets. Finally, we would like to thank CAIDA for their assistance in responding to questions regarding internet measurement practices.

REFERENCES

[1] A. Kalafat, C Shue, and M. Gupta, "Malicious Hubs: Detecting Abnormally Malicious Autonomous Systems." In Proceedings of IEEE INFOCOM 2010 Conference.

[2] H. Hu, S. Myers, V Colizza, and A Vespignani, "WiFi Networks and Malware Epidemiology." In Proceedings of the National Academy of Sciences (PNAS), 2009.

[3] B Gueye, A Ziviani, M Crovella, and S Fdida, "Constraint-Based Geolocation of Internet Hosts." IEEE/ACM Transactions on Networking. Vol. 14. No. 6, 2006