Location, Location, Location: Mapping Potential Canadian Targets in Online Hacker Discussion Forums

Richard Frank*, Mitch Macdonald, and Bryan Monk
International CyberCrime Research Centre
School of Criminology, Simon Fraser University
Burnaby, BC, Canada
{rfrank, mcmacdon, bmm8}@sfu.ca

***Abstract*— The goal of this paper was to analyze hacker forums to better understand the threats they pose to Canadian critical systems specifically and cyber-security more generally. To facilitate the data collection, a customized web-crawler was developed to specifically capture the structured content posted to forums. Three hacker forums were selected for analysis that represented different facets of the hacker community: carding (data theft), coding (malware development and deployment), and security (distribution of vulnerabilities). We identified and geolocated user disclosed IP addresses to try to identify critical systems and determine the extent as well as context in which critical systems were openly discussed by forum users. In total, 311,501 analyzable IP addresses were extracted from the data with 3,168 (1%) geolocated to Canada. The prevalence of Canadian IP addresses does not indicate their potential for exploitation, although it does highlight a perceived heightened interest in Canadian critical systems by hacker forum users. Potential at-risk systems included government agencies, universities across Canada, and private industries within the transportation network, namely aviation and shipping firms.**

Keywords— hacker forums; web-crawling; geolocation; critical infrastructure

I. Introduction

Targeted attacks against critical infrastructures are increasing on a global scale. Critical systems are rapidly being connected to the Internet, affording attackers opportunities to target virtual systems that operate and monitor physical structures through various modes of cyber-attack. Cyber-attacks are hostile operations that undermine the function of computer networks [15] with political, militaristic, or economic goals [31]. In Canada, a cyber-attack has been officially defined as, "the unintentional or unauthorized access, use, manipulation, interruption or destruction (via electronic means) of electronic information and/or the electronic and physical infrastructure used to process, communicate and/or store that information" [28]. That is, cyber-attacks range from rendering computer networks inaccessible for end-users, to the manipulation of virtual or physical equipment, and to the deletion or theft of sensitive data. Industrialized countries are most vulnerable to cyber-attacks due to increasing reliance on digital technologies.

The authors of this paper would like to thank Public Safety Canada, who funded this research through their Cyber Security Cooperation Program.

Prior to the widespread application of computer technology, attacks against critical infrastructures were limited to physical strikes during wartime conflicts between state-actors; however, due to the expansion of the Internet, power dynamics between state-actors and non-state actors have drastically shifted, such that non-state actors are now capable of attacking critical systems [17, 23]. The pervasiveness of computer technology, relative and widespread computer proficiency, and inexpensive anti-security software have created the necessary conditions for small groups of hackers to pose a serious threat to national critical infrastructure. Subsequently, there is a growing body of research that is focusing on data collected from hacker forums. The current study contributes to this literature by identifying and geo-locating Canadian IP addresses posted to hacker forums, through which potential targets of cyber-attack can be detected.

II. Literature Review

A. Critical infrastructure

Critical infrastructures are the physical and virtual systems, facilities, technologies, networks, assets, and services that support the health, safety, and economic well-being of communities, as well as the effective functioning of government [26, 37]. It has been suggested that infrastructures are 'critical' if: they provide essential, routine functions; no handy, rapid substitutes exist; sudden disruption in these functions causes nontrivial harm; and they are embedded in wide, functionally reciprocal, integrated systems [7]. In Canada, critical systems are often interconnected and interdependent within and across provincial and territorial boundaries, as well as international borders, forming massive, physically dispersed systems that are controlled by many different players, mostly within the private sector [14]. It is argued that the collapse of critical systems would cause crippling failures in essential services, leading to societal, political, and economic disaster [6]. The Canadian national critical infrastructure system spans 10 sectors [29]:

1) Agriculture (Agriculture and Agri-Food Canada);
2) Energy and Utilities (Natural Resources Canada);
3) Finance (Finance Canada);
4) Government (Government of Canada);
5) Healthcare (Public Health Agency of Canada);

978-1-5090-2857-3/16 $31.00 © 2016 IEEE
DOI 10.1109/EISIC.2016.31

6) Information and communication technology (Industry Canada);
7) Manufacturing (Industry Canada, Department of National Defense).
8) Safety (Public Safety Canada);
9) Transportation systems (Transport Canada); and
10) Water systems (Environment Canada)

Critical infrastructure originally referred to physical equipment, but the definition has been broadened to include networks of people, machines, and software [10]. Critical systems are increasingly digitized and connected to the Internet, which provides many benefits to both operators and users. The Internet supports a global network of information and communications systems that directly or indirectly connects these systems [31], such that digital systems now form the backbone of the national critical infrastructure system [34]; however, the complexity and interdependency of this larger system has increased its vulnerability to unpredictable threats emanating from cyberspace [11, 31]. Fortunately, the majority of past cyber-attacks have not caused significant damages to critical systems [3, 8]. Nonetheless, cyber-attacks have the potential to cause physical damage, disrupt vital services, or threaten the economy [2, 25].

B. Risks to critical infrastructures

The merger of old and new technologies create unintentional interaction effects such as incompatible software and vulnerabilities, which compromise the security of critical systems and create inherent difficulties for security assessment [16]. Vulnerabilities are flaws in computer security, information, or software systems that can be accessed and subsequently exploited by attackers seeking root access to the system. Security patches are intended to protect vulnerabilities against exploit tools, although patches are not 'fix all' solutions as overlooked or yet to be discovered vulnerabilities are sure to remain. For instance, in large, interconnected systems, updating one component risks interfering with the operation of other components, creating a disincentive for organizations to modify any working system, but in the process leaving them vulnerable. For example, a vulnerability called 'Microsoft Internet Explorer RDS ActiveX' was first observed, and subsequently patched in 2006, yet was still the second most observed vulnerability in January 2014—likely the result of end-users not patching their systems with the latest updates [36]. But perhaps most importantly, the average 'time to attack' is shortening following the disclosure of vulnerabilities, occurring within hours of discovery, whereas the average release time for patches is increasing, taking anywhere from months to one year to release.

While the number of disclosed vulnerabilities have remained consistent over time—6,253 in 2010 compared to 6,549 in 2014 [32, 33]—the number of new zero day vulnerabilities have nearly doubled in the past 5 years—14 in 2010 to 24 in 2014 [32, 33]. Zero day vulnerabilities are 'holes' within software that are unknown to its developers or vendors. If discovered by attackers, they can be exploited without awareness by the software developer. Many vulnerabilities, both disclosed and undisclosed, have manifested into real world threats. Analysis suggests that the exploitation of vulnerabilities accounts for a significant portion of attacks against critical infrastructures [21, 35]. A 2015 study surveying 26 nations throughout North and South America found vulnerabilities to be the top threat to critical systems within the government sector specifically, although the financial, information and communications technology sectors were also frequently exploited [35].

C. Role of Hackers

The increasing reliance on large, complex systems has increased their vulnerability to attackers [10]. Current events indicate that cyber-attacks are a growing threat and it is not a question of *if* hackers will attack critical infrastructures, but *when* and *to what degree* [11]. But attacks do not occur in a vacuum: the underlying social, political/ideological, and economic forces that operate within the online hacker community are at the root of cyber-security issues [16, 17].

Discussion forums have been widely embraced by hacker communities [19] due to the many advantages afforded to users. The most obvious advantage is the opportunity for like-minded individuals to connect with one another on a massive scale, irrespective of physical time and space. Discussion forums also provide a degree of anonymity, which serves to counter detection from law enforcement agencies [20]. Though interactions are transient in nature [4], hacker forums nonetheless foster interactive virtual environments that coordinate the exchange of social capital: the collective advantages gained through cooperation between individuals and groups [5]. In hacker forums social capital consists of opportunities for criminal collaboration, access to exploit tools, the diffusion of vulnerabilities, and stolen data markets [30]. Collaboration of this sort is necessary for the continued success of hacker forums and contributes to the persistent and growing issue of cyber security [38, 39, 40].

III. Methods

To analyze discussions within a hacking forum, they must be searchable and conducive to analysis. Relying on the forums' search capabilities both for searches and analysis would have required a very significant effort on the analyst's part, assuming the analysis could actually be performed with the tools available, all the while not being banned from the forum for unusual behaviour. To solve this problem, after selecting some hacking forums to download (Chapter III.A), custom software was designed to collect structured content from these forums (Chapters III.B). IP addresses were then extracted (Chapter III.C) and geolocated to allow for the geographical analysis of this information (Chapter III.D).

A. Target Selection

As part of a larger report, a manual survey of online hacker discussion forums identified 129 hacker forums on the public Internet that cater to English-speakers, although this is likely a gross underestimation of the number of hacker forums that exist. Of the 129 hacker forums, 34 required registration or, in the most underground forums, for existing users to invite or vouch for new users. The remaining 95 had no barriers to entry, that is, they did not have prerequisites for registration and their content was also accessible to non-registered observers. Only these open forums were considered as candidates for data collection.

Forum type	Coding	Security	Carding
Users	26,831	8,535	116,950
Total posts	147,604	464,572	952,858
Threads	16,809	49,849	99,032
Post:Thread	8.78	9.32	9.62

Table 1 – Size of target hacker forums at the data collection.

Dataset	Carding	Coding	Security	Total
United States	127,139	2,050	2,866	132,055
China	78,122	469	7,590	86,181
Brazil	5,616	31	927	6,574
Russia	5,913	255	169	6,337
Indonesia	5,009	2	812	5,823
India	5,005	29	133	5,167
United Kingdom	4,334	342	248	4,924
France	3,161	914	226	4,301
Venezuela	3,540	1	174	3,715
Canada	2,953	19	196	3,168
Ukraine	2,909	65	121	3,095
Germany	2,663	86	327	3,076
Saudi Arabia	2,572	12	10	2,594
Republic of Korea	2,104	14	83	2,201
Netherlands	1,759	65	153	1,977
Japan	1,747	50	81	1,878
Spain	1,456	280	60	1,796
Denmark	1,615	6	23	1,644
Romania	1,379	15	40	1,434
Thailand	1,270	5	97	1,372

Table 2 – Counts of IP addresses for the top 20 countries.

Province	Coding	Security	Carding	Total
Unknown		6	256	262
Alberta	3	11	128	142
British Columbia	4	4	250	258
Manitoba		4	40	44
New Brunswick		1	24	25
Newfoundland and Labrador			4	4
Northwest Territories			1	1
Nova Scotia		2	19	21
Ontario	5	33	621	659
Prince Edward Island			2	2
Quebec	7	45	1,543	1,595
Saskatchewan		2	63	65
Yukon			2	2

Table 3 – Number of IPs within each province.

Most information posted to open forums is freely or publicly accessible on the Internet, so forums members do not have any reasonable expectation of privacy. Adherence to this criterion satisfied ethical concerns expressed by the project's granting agency, and our university. To address any issues concerning privacy and the potential uncovering of personally identifiable information, the analysis does not identify specific forum users, but rather the content posted to the forums subject to analysis. These precautionary measures were guided by previous research designs that also analyzed data from hacker forums [18].

The candidate forums were prioritized for data collection based on post-to-thread ratio (Post:Thread), with the 10 largest being selected. A higher post-to-thread ratio indicates that, on average, there were more iterations of discussion within each thread, which was deemed important for the analysis of the data. A lower post to thread ratio is indicative of less discussion and interactivity between users and more lurking behaviors. While the visibility of posts increases criminal opportunities, discussions suggest a greater likelihood that the content is being utilized. Three forums were selected for final analysis: a *Coding* forum, a *Security* forum and a *Carding* forum (see Table 1), each representing a different facet of the online hacker community.

B. *Data Collection*

Data was collected using software called the Open Discussion Forum Crawler (ODFC). ODFC is a customized web-crawler designed to capture the content posted to discussion forums. This software captures relevant information from a user-selected forum by parsing and downloading its webpages, adhering to user-specific 'rules'. The captured data is then stored in a database which is designed to resemble the universal structure of discussion forums and is navigated in the same way: each forum has many sub-forums, which in turn have many threads of discussion, each with at least one post.

Due to the thematic grouping of sub-forums, and the chronological listing of threads and posts therein, each webpage within a discussion forum follows a strict structure. Each sub-forum has a title, followed by a description of the topic(s) of discussion, the number of replies, and views. Threads and posts exhibit the same implicit structure, as each post repetitively lists the Author, body, and date (for example). This structure is consistent from post-to-post and across threads. ODFC downloads each webpage and then performs extensive clean-up. Rules (defined per forum, by the researchers) then help ODFC extract relevant data from each sub-forum, thread, and post. The information exchanged in restricted sub-forums or within private messages are not open source data and, thus, cannot be collected by ODFC. For full details of this process, see [24].

C. *Identifying potential Canadian targets*

While it is widely believed that hacker forums contribute to cyber-attacks, this was not empirically tested until 2012 [38]. Modest support was found for the correlation between the number of exploit tools exchanged through hacker forums and the number of corresponding vulnerabilities exploited in computer systems. Though this research was an important first step, it squarely focused on the tools used to attack critical systems, researchers have not sought to uncover the potential targets of attack. Of the millions of computers connected to the Internet, most are identifiable only by their Internet Protocol (IP) address [1]. Thus, IP addresses are crucial pathways that may be leveraged by attackers to exploit critical systems. The current study adds to this body of research by identifying and geo-locating Canadian IP addresses posted to hacker forums, through which potential targets of cyber-attack can be detected.

D. *Geolocation*

The current version of IP addresses (IPv4) follow a consistent and recognized pattern. An IP address consists of four

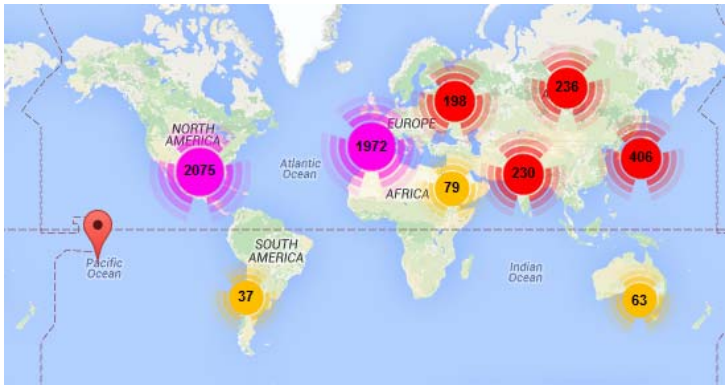
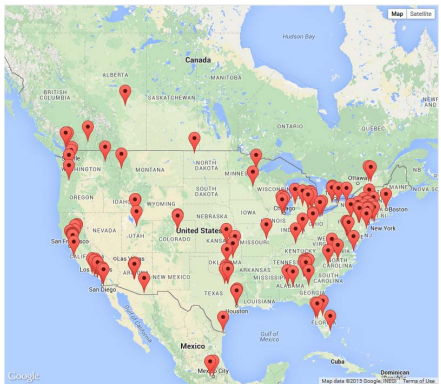


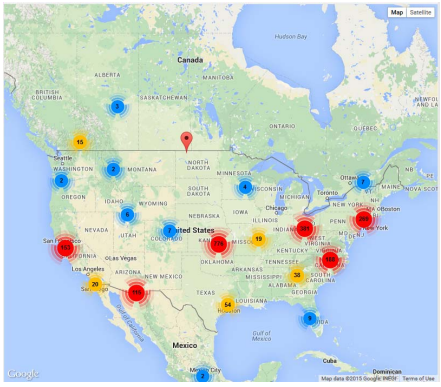
Figure 1 – Global distribution of IP addresses in the Coding Forum

IP Address	Internet Service Provider	City, Province
67.212.77.13	Netelligent Hosting Services	Montreal, QC
72.12.166.17	Mountain Cablevision Ltd.	Hamilton, ON
69.50.171.122	Skyway West	Port Coquitlam, BC
66.154.102.142	Assertivenet via Synergy	Vancouver, BC
67.215.134.151	Source Cable Ltd.	Hamilton, ON
68.150.40.92	Shaw Communications	Edmonton, AB
24.150.162.84	Cogeco Cable Solutions	Burlington, ON
70.52.6.84	Bell Canada	Saint-Eustache, QC
69.10.145.211	Rackforce Hosting	Kelowna, BC

Table 4 – Sample of Canadian IP addresses extracted from the coding forum



a) Exact locations



b) Clustered locations

Figure 2 – Locations of IPs in North America, from the coding forum

sets of numbers, or blocks, ranging from 1 to 255 separated by a decimal. An example IP address, 66.183.3.81, shows the variation that can exist with single, double and triple digit values within each block. The inherent pattern of IP addresses made automated data extraction and analysis consistent. First, using Regular Expressions (RegEx) ODFC was used to extract all IP addresses from the captured data while also determining the frequency of each IP address. Most IP address were identified only once, while some were posted multiple times within or across the hacker forums. After this data was extracted by ODFC, MaxMind's GeoLite [13] database was used to assign the geographical information to each IP address. GeoLite is an offline geolocation database that is updated frequently. Its accuracy varies depending on the granularity level: 99.80% at the country level; 90% at the province/state level; and 83% at the city/municipal level. Since this study focuses on addresses within Canada, the 99.80% accuracy rate at the country level provided sufficient confidence in the accuracy.

An initial analysis was done at the country level to identify a relevant sample of Canadian IP addresses. A lookup was made in the GeoLite database to retrieve the longitude/latitude coordinate for each IP address extracted from the data. The volume of Canadian IP address extracted from the coding and security forums were manageable, so each sample was analyzed in its entirety; however the volume of Canadian IP addresses extracted from the carding forum was too large for exhaustive review and was thus randomly sampled. After a list of Canadian IP addresses were identified, a manual analysis of the registrant data for that IP address assessed their relevance to the research objective. A 'WhoIS' lookup of this IP address reveals data concerning the Internet Service Provider (ISP), the registered holder, and the geo-coordinates of the computer system. Through these identifying pieces of information, it was possible to detect potential Canadian targets within the data.

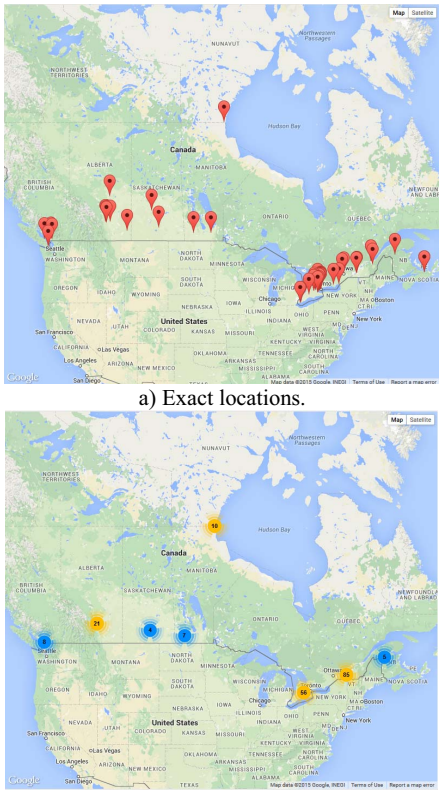
IV. Results

A total of 324,901 IP addresses were extracted from the data with 311,501 (95.88%) analyzed to determine their country of origin. The remaining 13,400 IP addresses were linked to internal home networks that are neither public, nor allocated by the ISP; therefore, any data associated with the IP address is invalid and not suitable for analysis.

The IP addresses discussed in the dataset were found to be from 183 countries. As expected, the United States was by far the biggest contributor of IP addresses to the hacker forums, most likely due to its very large computer-based infrastructure, and "rich" resources, making it a very attractive target. Although China is a significantly larger country in terms of population, it placed in second for overall frequency in IP addresses, possibly due to it having a smaller footprint on the Internet, or as a result of its internet monitoring and firewalling efforts. Canada ranked

IP Address	Internet Service Provider	City, Province
130.15.158.38	Queen's University	Kingston, ON
142.204.13.100	Seneca College	North York, ON
142.58.154.54	Simon Fraser University	Burnaby, BC
67.215.6.66	GloboTech Inc.	Saint-Quentin, NB
173.239.137.90	Rogers Communications	Ottawa, ON
192.139.15.1	Universite Du Quebec	Quebec City, QC
136.159.99.156	University of Calgary	Calgary, AB
131.104.106.246	University of Guelph	Guelph, ON
131.104.106.245	University of Guelph	Guelph, ON
130.179.56.109	University of Manitoba	Winnipeg, MB
130.179.56.108	University of Manitoba	Winnipeg, MB
140.193.56.113	University of Manitoba	Winnipeg, MB
137.122.185.41	University of Ottawa	Ottawa, ON
137.122.185.43	University of Ottawa	Ottawa, ON
192.139.15.33	Universite Du Quebec	Quebec City, QC
142.150.165.157	University of Toronto	Toronto, ON
129.97.74.12	University of Waterloo	Waterloo, ON
130.63.117.121	York University	Toronto, ON
130.63.127.59	York University	Toronto, ON

Table 5 – Canadian IP addresses extracted from the security forum.



a) Exact locations.

b) Clustered locations

Figure 3 – Canadian IP addresses from the Security Forum

10th on this list, ordered by frequency. Interestingly, although Canada contains approximately 10% of the population of the US, the number of IP addresses located within Canada, with respect to the US, was only 2.4%. The frequency of the top 20 countries is shown in Table 2.

Focusing on Canada, analysis showed that 3,080 (1.01%) of unique IP addresses were traced to Canadian geo-coordinates. Of these, the vast majority (2,953, or 95.88%) of the IP addresses were from the carding forum and only 19 were from the coding forum. Broken down by province, the majority (1,543, or 50.01%) of IP addresses came from the Quebec, followed by Ontario (621, or 20.16%). It would be expected that Ontario would take the major share, especially from the carding forum, given that the major financial center of Canada is in Toronto, Ontario. A provincial breakdown is shown in Table 3.

While 3,080 (or 1.01% of the entire set of IPs extracted) may seem insignificant, Canada's share of the public Internet was 28 million [9] of the estimated 13.3 billion (0.2%) devices connected to the Internet in 2013 [27]. Thus, findings suggest that Canadian IP addresses were actually overrepresented in the data. Results for each hacker forum are discussed in sequence according to the total volume of IP addresses extracted from the data. Relevant IP addresses were also manually analyzed to examine the context in which they were used during discussion in the coding (Chapter IV.A), security (Chapter IV.B) and carding (Chapter IV.C) forums.

A. *Coding Forum*

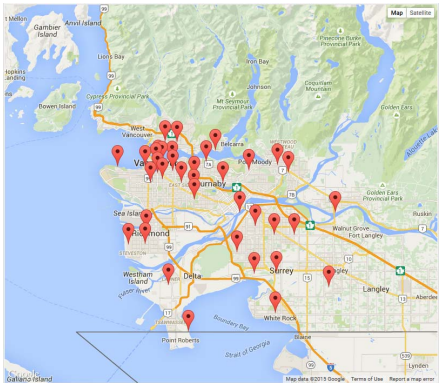
The coding forum contained 2.28% of the IP addresses retrieved from the aggregate data, 74.54% of which belonged to publicly designated holders and 0.36% were traced to Canadian geo-coordinates. The global distribution of IP addresses within the coding forum is shown in Figure 1, with the distribution of IP addresses only in North America shown in Figure 2. For Figure 2a, repeated IP addresses, or different IP addresses which geocode to the same location are shown as a single point, while for Figure 2b, repeated IP addresses, or different IP addresses which geocode to the same location are counted multiple times.

Interestingly, each of the identified ISPs were unique (see Table 4 for a sample), although most were registered to individual holders with only one being registered to a private entity, Teekay Shipping (Canada) Ltd, in Kelowna, British Columbia. Teekay Shipping Ltd. is one of the world's largest marine energy transportation companies. The shipping industry is vital to the global distribution supply chain and economy and, thus, is a key part of the transportation sector. Discussion indicated that the IP address registered to Teekay Shipping may have been used as a proxy to relay data to another IP address. The user that disclosed the IP address stated they "only collect information" when they can "sell the data" and have "clients that request information" at which point they "find ways to get it". The motivation for compromising this system remained unclear, but it is likely that the system had already been comprised.

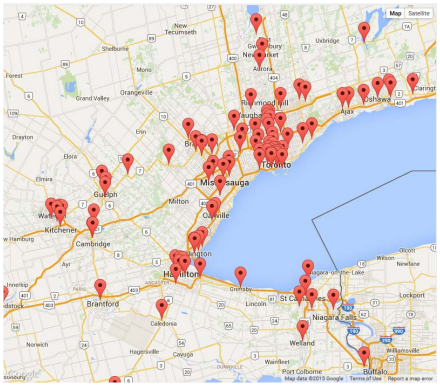
IP Address	Internet Service Provider	City, Province
70.38.37.75	Iweb Technologies	Montreal, QC
184.107.89.134	Iweb Technologies	Montreal, QC
66.36.149.176	AEI Internet Inc.	Montreal, QC
128.43.0.0	Department of National Defence	Ottawa, ON
128.43.255.255	Department of National Defence	Ottawa, ON
216.123.207.124	Telus Communications Inc.	Fort McMurray, AB
204.101.161.0	Bell Canada	Burnaby, BC
167.44.255.255	Government Telecommunications Agency	Hull, QC
167.44.0.0	Government Telecommunications Agency	Hull, QC
207.102.0.0	Telus Communications	Bella Bella, BC
137.94.0.0	Royal Military College of Canada	Kingston, ON
137.94.255.255	Royal Military College of Canada	Kingston, ON
66.183.33.81	Telus Communications Inc.	New Westminster, BC

Table 6 – Canadian IP addresses extracted from the carding forum.

Figure 4 – Clustered map of IP addresses discussed in the Carding Forum.



a) Exact locations of IP addresses in the Greater Vancouver Regional District.



b) Exact locations of IP addresses in the Greater Toronto Area.

Figure 5 – Locations of IPs from the carding forum

B. Security Forum

The security forum accounted for 5.90% of all IP addresses extracted from the aggregate data, of which 92.54% were registered to publicly designated holders and 1.15% traced to Canadian systems specifically. The distribution of Canadian IP addresses within the security forum is shown in Figure 3, with Figure 3a showing the exact locations of the IP addresses (multiple points that geolocate to the same XY coordinate are shown once), and Figure 3b showing the clusters of IP addresses (where multiple points that geolocate to the same XY coordinate are represented multiple times).

Of the total volume of Canadian IP addresses posted to the security forum, 55.10% were unique (see Table 5 for a subset). Results of the geolocation analysis span potentially comprised systems in 8 provinces across 36 cities. At the provincial level, 57.89% of Canadian IP addresses were traced to ISPs and holders in Ontario. At the municipal/city level, 34% of unique IP addresses were traced to Montreal. 33 different ISPs were identified within this sample with OVH Hosting Inc. being linked to 29.63% (32 of the 108) of the unique IP addresses, although all of the IP addresses linked to this ISP were registered to individual holders, not critical infrastructure operators.

75.93% of Canadian IP addresses were assigned to individual holders with the remaining 24.07% registered to public and private entities. Most common were those registered to universities and, to a lesser extent, health services and aviation companies. The Canadian IP addresses linked to universities were included in posts for rootkits that were advertised for sale within the security forum. Relevant discussion indicated that these IP addresses were sourced from the Shodan database, a repository for vulnerable IP addresses that are registered to critical infrastructures around the world. In contrast, the IP addresses associated with the avionics company were retrieved by scanners, identifying consumer-grade routers.

C. Carding Forum

The largest number of IP addresses were extracted from the carding forum, accounting for 96.12% of all identified IP addresses. Furthermore, 96.59% of the IP addresses extracted from this sample were analyzable (not registered to an internal network) with 1.02% traced to Canada. The distribution of Canadian IP addresses within the carding forum is shown in Figure 4. Looking at individual geo-coordinates did not yield a usable map (due to the large number of pins, no details were visible), and hence is not presented. Focusing in on the various

larger regions/cities, Figure 5a shows the location of IP addresses identified in the Greater Vancouver Regional District, while Figure 5b shows the Greater Toronto Area, which is the financial centre of Canada. These maps clearly show that the majority of activity is following the population densities and financial centers and thus is mainly in or near Vancouver, Toronto, and Montreal.

Due the large volume of Canadian IP addresses posted to the carding forum, the data selected for manual analysis was randomly sampled; 288 of 2953 Canadian IP addresses were included in the analysis. Table 6 contains Canadian IP addresses linked to operators of critical infrastructures spanning government agencies, aviation companies, financial institutions, and technology firms. IP addresses registered to government agencies were most frequently uncovered during analysis, accounting for 9.75% of the random sample, including the Department of National Defence, Government Telecommunications Agency, and the Royal Military College of Canada. The availability of this information to hacker forum users poses a serious threat to the integrity of these critical systems and Canadian national security. Interestingly, the IP addresses for the Department of National Defence was posted to the carding forum a year prior to the January 2011 cyber-attack that targeted its computer networks.

V. Discussion

By analyzing Canadian IP addresses posted to hacker forums and identifying their geo-coordinates, it was established that pertinent information can be used to target Canadian critical infrastructures and this information is accessible through hacker forums. The use of the IP addresses in the context of discussion does not provide indication to their potential for exploitation; however, the prevalence of Canadian IP addresses across hacker forums highlights the potential threats to Canadian systems. While IP addresses themselves are not indicative of significant threat to any one system, they provide a crucial pathway that allows attackers to infiltrate critical systems. In many countries, critical infrastructures are connected to the same computer networks as civilian systems [1], which means attackers are likely forced to attack other systems sharing no association with the target(s). This includes a variety of entities—including government agencies, public institutions, and industries— such as those identified through the geolocation techniques used in this study. Furthermore, having these IP addresses available in a public forum allows others to access and collect this information, providing shortcuts for the target selection. These shortcuts serve as criminal opportunities which decrease the costs associated with offending. Thus, just the fact that these IP addresses are available increases their potential vulnerability.

Utilizing the features of ODFC, the geolocation of IP addresses specific to Canada became a feasible process. Being able to identify pockets of data to determine the context in which it is used allows for a rapid assessment of emerging threats. For example, Figure 5 shows the geo-coordinates of the Canadian IP addresses extracted from the carding forum. Detecting significant changes in "hotspots", through the rapid increase of activity in a specific region of Canada for example, could potentially identify not only new vulnerabilities, but also their impacts on cyber-security threats in that region or industry sector. In this way, geolocation techniques compliment traditional data mining techniques like keyword analysis. Hotspot mapping has shown success in determining changes in crime patterns within different environmental contexts. The application to cybercrime has seen limited use although the potential exists for its use to identify emerging risks.

Despite the unique opportunities that geolocation analysis lends to potential target identification, these methods are not without limitations. As previously mentioned, IP addresses posted in hacker forums merely suggest that information used to attack these systems are exchanged through hacker forums and do not provide indication of the potential threat level. To this point, the exchange of IP addresses lacked context, as there were generally no discussions to suggest the way in which the information was to be used. In addition, the results provide only a glimpse of the content available online and would likely vary by sample, or from one hacker forum to the next. This is due to the fact that forum users would likely be inclined to share information that they themselves find intriguing, and is thus related to the focus of the forum.

Future research may attempt to establish connections between a random sample of IP addresses made available in hacker forums, and actual attacks on their corresponding systems in the real world to better understand the validity of these techniques. These objectives could be achieved by gaining access to data collected by intrusion detection systems and correlating IP addresses across datasets, while controlling for attack vectors. Research should also consider a similarly proposed design using IP addresses geo-coordinated to the United States due to its extent massive infrastructure related to the Internet and computer technologies, and it's interconnectedness with corresponding Canadian infrastructure. Furthermore, this analysis used data captured over a fixed period of time, with analysis following. Such a strategy has the benefit of providing consistent data for analysis. Progress has been made since the conclusion of this study, and ODFC has been adapted to capture data continuously, thus allowing for real-time monitoring of malicious forums. The ability to monitor online criminal networks in real time provides a host of advantages, including the rapid assessment of emerging threats. Complementary analysis tools could be designed to coincide with ODFC to allow it to highlight content meeting specified criteria as it is encountered. These tools would enable longitudinal analysis to be performed on this data to attempt to determine trends and themes as they emerge. Sentiment analysis could also be incorporated to gain a better understanding of the post content and how sentiment fluctuates over time, and whether a major attack influences sentiment and/or the emergence of other potential targets. These ideas are left for future work.

VI. Conclusions

Through data mining of data and manual revision, this study sought to identify information pertinent to critical infrastructures, with specific emphasis on Canadian systems, posted to publically accessible hacker forums. Results indicate that Canadian IP addresses were overrepresented in the data, meaning that Canadian critical systems are of particular interest to hacker forum users. IP addresses linked to information and

communication technologies, finance, banking and, to a lesser degree, government systems, were relatively common in comparison to the total volume extracted from the data. Much of the discussion involved the tools used to attack websites or servers, especially those linked to financial systems and government agencies. Specific details were relatively scarce, although generally discussion was largely transient, so it is not surprising that potential targets of attack were not disclosed in hacker forums. Nonetheless, detecting regional hotspots and mobility patterns through real-time monitoring of activity could potentially inform cyber-security threat assessments and target hardening strategies by region and sector.

VII. References

1) Applegate, S. D. (2011). Cybermilitias and political hackers: Use of irregular forces in cyberwarfare. IEEE Security & Privacy, (5), 16-22.
2) Benjamin, V., & Chen, H. (2012). Securing Cyberspace: Identifying Key Actors in Hacker Communities. In Intelligence and Security Informatics (ISI), 2012 IEEE International Conference (pp. 24-29). IEEE.
3) Brenner, S. W. (2008). Cyberthreats: The emerging fault lines of the nation state. New York: Oxford University Press.
4) Broadhurst, R., Grabosky, P., Alazab, M., & Chon, S. (2014). Organizations and Cyber Crime: An Analysis of the Nature of Groups engaged in Cyber Crime. International Journal of Cyber Criminology, 8(1), 1-20.
5) Burt, R. S. (2000). The Network Structure of Social Capital. Research in Organizational Behaviour, 22, 345-423.
6) De Bruijne, M. & Van Eeten, M. (2007). Systems that Should Have Failed: Critical Infrastructure Protection in an Institutionally Fragmented Environment. Journal of Contingencies and Crisis Management, 15(1), 18-29.
7) Demchak, C. C. (2006). Embracing surprise in resilient complex critical infrastructures: Rapid crisis response lessons from military organizations and the Atrium model. In SEMA/ECMA Conference on "Future Challenges for Crisis Management in Europe", Stockholm, Sweden.
8) Denning, D. E. (2011). Cyber-conflict as an emergent social problem. In T. J. Holt & B. Schell (Eds.), Corporate hacking and technology-driven crime: Social dynamics and implications (pp. 170-186). Hershey, PA: IGI-Global.
9) Dingman, S. (2015). Beyond the smartwatch: Canada finds its place in the internet of things. The Globe and Mail.
10) Egan, M. J. (2007). Anticipating Future Vulnerability: Defining Characteristics of Increasingly Critical Infrastructure-like Systems. Journal of Contingencies and Crisis Management, 15(1), 4-16.
11) Geers, K. (2009). The cyber threat to national critical infrastructures: Beyond theory. Information Security Journal: A Global Perspective, 18(1), 1-7.
12) Geers, K. (2010). The challenge of cyber attack deterrence. Computer Law & Security Review, 26(3), 298-303.
13) GeoLite from MaxMind. http://dev.maxmind.com/geoip/legacy/geolite/
14) Graham, A. (2014). Canada's Critical Infrastructure: When is Safe Enough Sage Enough? National Security Strategy for Canada Series. The Macdonald-Laurier Institute.
15) Hathaway, O. A., Crootof, R., Levitz, P., Nix, H., Nowlan, A., Perdue, W., & Spiegal, J. (2012). The Law of Cyber-Attack. California Law Review, 100(4), 817-886.
16) Hellstrom, T. (2007). Critical infrastructures and systematic vulnerability: Towards a planning framework. Safety science, 45(3), 415-430.
17) Holt, T. J., & Kilger, M. (2012). Know Your Enemy: The Social Dynamics of Hacking. The Spartan Devils Chapter of the Honeynet Project.
18) Holt, T. J., & Lampke, E. (2010). Exploring stolen data markets online: products and market forces. Criminal Justice Studies, 23(1), 33-50. DOI: 10.1080/14786011003634415
19) Holt, T. J., Strumsky, D, Smirnova, O., & Kilger, M. (2012). Examining the Social Networks of Malware Writers and Hackers. International Journal of Cyber Criminology, 6(1), 891 903.
20) Holz, T., Engelberth, M., & Freiling, F. (2009). Learning more about the underground economy: a case study of keyloggers and dropzones. In European Symposium on Research in Computer Security (ESORICS).
21) IBM. (2014). IBM X-Force Threat Intelligence Quarterly 1Q 2014. Retrieved from http://business review-vodcasts.com/ibm-securitycentre/whitepapers/ibm_x-force.PDF
22) Luo, X., Chang, R., & Chan, E. (2005). Performance Analysis of TCP/AQM under denial-of service attacks. In Proceedings of the 13th International Symposium for Modeling, Analyzing, and Simulating Computer Telecommunication Systems, September 27-29, 2005, pp. 97-104.
23) Lynn, W. J. (2011). Remarks at the 28th Annual International Workshop on Global Security. June 16, 2011.
24) Macdonald, M., Frank, R., Mei, J., Monk, B. "Identifying Digital Threats in a Hacker Web Forum", International Symposium on Foundations of Open Source Intelligence and Security Informatics, Paris, France, 2015.
25) Miller, B., & Rowe, D. (2012) A survey of SCADA and critical infrastructure incidents. In Proceedings of the 1st Annual conference on Research in information technology (pp. 51 56). ACM.
26) Office of the Auditor General of Canada. (2012). 2012 Fall Report of the Auditor General of Canada: Protecting Canadian Critical Infrastructure Against Cyber Threats. (Office of the Auditor General of Canada No. FA1-2012). Ottawa, ON.
27) Press, G. (2014, Aug). Internet of things by the numbers: Market estimates and forecasts. Forbes Magazine. Retrieved from: http://www.forbes.com/sites/gilpress/2014/08/22/internet-of things-by-the-numbers-market-estimates-and-forecasts/
28) Public Safety Canada. (2010). Canada's Cyber Security Strategy. (Public Safety Canada No. PS4 102/2010E). Ottawa, ON.
29) Public Safety Canada. (2014). Action Plan for Critical Infrastructure (2014-2017). (Public Safety Canada No. PS4-66/2014E). Ottawa, ON.
30) Radianti, J., & Gonzalez, J. J. (2007). Understanding hidden information security threats: The vulnerability black market. In System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference (pp. 156c-156c). IEEE.
31) Sigholm, J. (2013). Non-State Actors in Cyberspace Operations. Journal of Military Studies, 4(1), 1-37.
32) Symantec. (2011, Apr). Symantec Internet Security Threat Report: Trends for 2010, 16.
33) Symantec. (2015, Apr). Symantec Internet Security Threat Report: Trends for 2015, 20.
34) Ten, C. W., Manimaran, G., & Lui, C. C. (2010). Cybersecurity for critical infrastructures: attack and defense modeling. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on Smart Grid, 40(4), 853-865.
35) Trend Micro. (2015). Report on Cybersecurity and Critical Infrastructure in the Americas. Organization of American States. Retrieved from http://www.trendmicro.com/cloudcontent/us/pdfs/security-intelligence/reports/critical-infrastructures-west-hemisphere.pdf
36) Trustwave. (2015). Malware Statistics. SpiderLabs. Retrieved from http://www.trustwave.com/support/malware-statistics.asp
37) United States Department of Homeland Security. (2013, Oct 24). "What is Critical Infrastructure?" Retrieved from: http://www.dhs.gov/what-critical-infrastructure
38) Wang, Q. H., Yue, W. T., & Hui, K. L. (2012). Do hacker forums contribute to security attacks? In E-Life: Web-Enabled Convergence of Commerce, Work, and Social Life (pp. 143-152). Springer Berlin Heidelberg.
39) Yip, M., Shadbolt, N., & Webber, C. (2013). Why forums? An empirical analysis into the facilitating factors of carding forums. In Proceedings of the 5th Annual ACM Web Science Conference (pp. 453-462). ACM.
40) Yip, M., Shadbolt, N., Tiropanis, T., & Webber, C. (2012b). The Digital Underground Economy: A Social Network Approach to Understanding Cybercrime. In The Proceeding from the Conference on Digital Futures '12, October 23-25, 2012, Aberdeen, UK.