Geographic distribution and content analysis of targets on five hacking-focused online discussion forums

Emily Choma
International CyberCrime Research
School of Criminology, Simon Fraser University
Burnaby, Canada
echoma@sfu.ca

Richard Frank *
International CyberCrime Research
School of Criminology, Simon Fraser University
Burnaby, Canada
rfrank@sfu.ca

***Abstract*— The goal of this paper was to present an analysis of five hacker forums to better understand the threats they pose to online systems. To facilitate the data collection, a customized web-crawler was used to capture five hacking-focused online discussion forums. We identified and geolocated user disclosed IP addresses to determine what types of systems hackers were discussing within their communities. In total, 11,062,793 posts were retrieved, 225,476 IP addresses extracted, resulting in a detailed analysis of 31,088 posts which contained those IP addresses. Results indicate that, while most of these IP addresses are posted as a result of automated scans for SOCKS proxies, there are posts which discuss more targeted attacks. These hackers are mainly targeting financial institutions and search engine companies.**

Keywords— hacking; discussion forum, critical infrastructure; Geolocation

I. INTRODUCTION

Cyber-attacks targeting the computer and control systems of critical infrastructures are increasing in both scope and prevalence. There have been reports of dramatic global increases in attacks targeting power grids, financial institutions, and transportation networks [1]. For example, approximately one quarter of organizations operating critical infrastructure report malware attacks [2]. Due to the connected nature of the Internet and the lack of capable guardianship in the form of computer security, malware is easily spread over the Internet. Virtual infrastructures are frequently targeted in these attacks to gain access to massive databases of information. In 2014 alone, three of the ten largest data breaches in history occurred and more than 822 million records were exposed worldwide, exceeding 2013's statistics [3]. The estimated proceeds of cyber-attacks surpass $US 1 trillion per year, affecting millions of individuals, corporations, and governments worldwide [4].

One strategy to combat this problem is to monitor and analyze the communal networks of the cyber criminals who commit these attacks. More specifically, their communication activity can be analyzed to determine potential threats. That is what this paper does to five online discussion forums specializing in hacker- and hacking- related topics.

A. Critical Infrastructure

Critical infrastructure is a term that encompasses all systems, processes, networks, technologies and assets that provide services essential to the security, health, safety and economic prosperity of individuals, in addition to providing services that support the day-to-day functions of governments [6, 7, 17, 18, 19]. The classification of critical infrastructure sectors varies across nations; however, sectors may include energy, finance, government, healthcare, information and communications technology, transportation, and water systems, among others [17, 18, 19]. Although critical infrastructures may stand alone (through the use of air-gapping), they are increasingly interconnected and interdependent both within and across sectors [6, 9, 18]. Consequently, attacks on one critical infrastructure may result in cascading effects across other systems [9]. Critical infrastructures are often privately owned, and intersect both local and international borders [6, 16]. The responsibility for securing critical infrastructure lies primarily with its owners and operators [25]. As such, they represent a challenging and dynamic security issue that requires collaboration among private sector stakeholders, local government departments, and international authorities [6, 23]. Cyber-physical threats to critical infrastructure are continually evolving due to the integration of network-connected systems, which has resulted in previously unforeseen vulnerabilities to critical infrastructures that traditionally operated on closed, proprietary software [6, 8, 12, 15]. These emerging cybersecurity vulnerabilities, coupled with the potential for attacks to have devastating effects on essential services, make critical infrastructures an attractive target for actors seeking to cause substantial disruption to the economy, health, and security of nations.

Beyond nature-based hazards, aging systems, and geo-political threats, the critical infrastructure risk landscape is evolving due to the incorporation of network-connected products and systems across all sectors [6, 15, 25]. Although many new technologies have positively impacted the efficiency and effectiveness of essential services, they also expand upon the scope of potential vulnerabilities [15, 25]. Automated industrial process systems, e-government services, smart power grids, Internet of Things (IoT) integration in manufacturing, and automated financial services are but a few of the recent technological developments that may expose sectors to greater risk [15, 25]. State and non-state sponsored attacks against

critical infrastructures have increased in recent years [26, 29]. As critical infrastructures become increasingly dependent on technology, this trend may continue. The 2007 distributed denial of service (DDoS) attacks against critical communications, financial, and government infrastructure in Estonia demonstrate how sectors and nations reliant on digital systems can be seriously disrupted by concerted cyber-attacks [28]. Although governments have traditionally viewed state-sponsored attacks on critical infrastructure as the most pressing threat, non-state actors also possess the capabilities to carry out a disruptive attack [26]. The threat posed by non-state actors has become more tangible in recent years due to the transition from closed, proprietary systems to commercial off-the-shelf software (COTS) with known vulnerabilities [15]. The spread of the Internet has also allowed for previously inaccessible hacking tools and techniques to be disseminated on online communities, disrupting the traditional power differential between state and non-state actors [8]. This threat is further substantiated by prior research indicating that critical infrastructure vulnerabilities and exploits are being discussed in online hacking communities [10, 16].

B. *Industrial Control Systems*

Industrial control systems such as supervisory control and data acquisition (SCADA) and distributed control systems (DCSs) are widely used across critical infrastructure sectors, including energy and utilities, industrial, manufacturing, water, and transportation sectors [6, 12, 16]. These systems are widely recognized as being particularly vulnerable to cyber-attacks [6, 8, 12, 15, 30]. Historically, industrial control systems were closed, proprietary systems that did not require strong cybersecurity protocols, often sending operating commands without authentication and in clear-text [6, 8]. However, the integration of operational technologies across critical infrastructure sectors has since lead to more open communication channels for industrial control systems, such as shared leased lines, Ethernet, and wireless networks [6, 12, 15]. Furthermore, COTS products are now widely incorporated into process control systems for critical infrastructures [6]. These design and operational changes in industrial control systems have important implications for the critical infrastructure risk landscape, as technical information about products in industrial control systems, including their vulnerabilities, are more widely known [6].

C. *Online Hacking Communities:*

Hacking forums serve as a medium for the exchange of knowledge, techniques and tools among members [21]. Online hacking communities are generally comprised of loosely connected networks and include members with a wide range of skill levels [10, 22]. These forums may be openly available, or accessible by invite-only [22]. Knowledgeable members who sell offensive hacking tools are often willing to provide technical support and competitive pricing to other members, thus widening the range of individuals who have the potential to carry out malicious cyber-attacks [20, 21]. Due to the easily accessible nature of many online hacking forums, the additional layer of perceived anonymity provided by the Internet, and the ability to freely disseminate knowledge and tools, online hacking forums are worthwhile to study in order to evaluate the current non-state hacker threat landscape [10, 22].

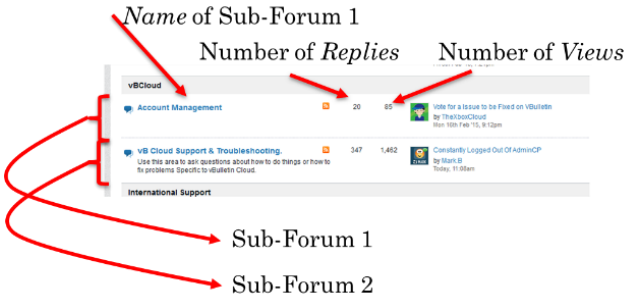

Figure 1 - Forum content

D. *This Study*

Much of the current literature has focused on describing or developing typologies of cyber-attacks, but has not directly examined the targets being discussed. By looking for IP addresses within these hacking communities, this study makes two primary contributions: 1) to the cyber security detection methods and analyses used to identify open source data (exchanged freely over the internet) and potential threats to critical infrastructure; and 2) to determine the extent to which and how various critical infrastructures are being targeted in hacker forums. The end goal of this specific research is to better understand the threats, and how they could be detected, within these communities. Ultimately, the long-term goal is to develop automated tools that combine these methods to streamline data analysis and facilitate real-time monitoring to identify potential threats as they emerge.

This paper first discusses the details of the data collection, clean-up and extraction process (Section II). Three types of analysis are conducted, geolocation, qualitative and targeted keyword searches, which are presented (Section III). The findings are then summarized and put back into the larger context as the paper is wrapped up (Section IV).

II. METHODS

A. *Data collection*

Data was collected using software called The Dark Crawler (see https://www.thedarkcrawler.com), a custom-built web-crawler designed to capture content posted to online discussion forums. The Dark Crawler (TDC) captures information from a

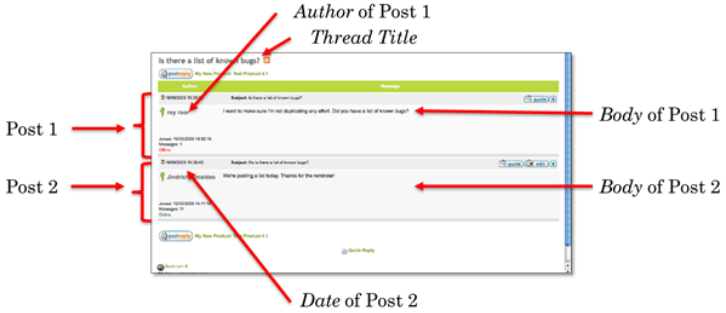

Figure 2 - Thread content

user-selected forum by downloading its web-pages, parsing the page apart with the use of forum-specific 'rules' to extract all useful information present on the forum, which is then stored in a database. For an overview, see [10]. The database is designed to resemble the structure inherent to 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 listings of threads, each webpage within a discussion forum follows a specific structure. A sample structure is shown in Figure 1, which illustrates multiple sub-forums displayed one after the other. Each sub-forum has a name, followed by a description of the topic(s) of discussion, the number of replies, and views. Figure 2 is similar to Figure 1, but depicts the structure of posts, rather than sub-forums. 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, but usually varies from forum to forum.

The data capture is as follows. TDC is configured to be able to parse apart each forum properly by specifying the location of a number of important pieces of data from each forum, sub-forum, thread, and post page. Locations of important pieces of data are specified using the format of the XPath standard [31], where each XPath query is applied to the HTML of the webpage and results in zero or more pieces of text. Each result that the XPath query returns might need further refinement which is done through a standard XQuery [32]. The XPath and XQuery combination, called a rule, results in the final data element being extracted (the author of a post, for example). Each data element that must be extracted from the forum, sub-forum, thread and post pages requires a rule. Following the configuration of all the rules for a forum, TDC retrieves the first page of each sub-forum, detects the number of pages and calculates the URLs for all of them. For each URL (i.e. each page of the sub-forum), TDC retrieves the content of the webpage, applies the corresponding rules to determine relevant thread information, then proceeds to download all pages of each thread. For each web-page retrieved, rules are applied to it to retrieve targeted data (author, date, body of the post, etc.) which are then stored in a database. Once the entire forum is downloaded into the database, the data can be extracted in various ways, such as retrieving posts containing keywords or those written by specific authors, or even constructing a social network of the entire forum.

B. Forum selection

The hacker forums analyzed for this study were downloaded by TDC in their entirety. These forums were selected for three reasons. First, they had large numbers of existing members, number of posts, and discussion threads. These characteristics increased the likelihood this forum contained relevant and relevant data. Second, the topics of its sub-forums were relevant to malicious hacking, which corresponds with the focus of this study. Finally, the discussion between forum members is freely or publicly accessible on the Internet, which means these individuals do not have any reasonable expectation of privacy. Adherence to this criterion satisfies ethical concerns associated with this study; however, to address issues concerning privacy and the potential uncovering of personally identifiable information, the domain of these hacker forum will not be identified, nor will the forum member's user-names.

ForumID	# of Posts	# of Authors	StartDate	EndDate	# of IP Addresses	# of Unique Ips	# of Posts with IP Addresses
10	935,449	65,935	4/7/2011 06:16	4/23/2018 00:00	113,616	31,391	5,474
11	1,145,378	51,533	4/8/2013 02:30	4/12/2018 22:25	26,513	9,475	92
157	7,136,836	212,348	4/23/2005 00:00	4/23/2018 00:00	35,291	16,571	5,989
159	1,581,238	33,535	2/9/2002 00:00	4/8/2018 00:00	34,694	10,763	16,115
160	263,892	9,505	Dates not available		15,362	4,909	3,418
Total	**11,062,793**	**372,856**			**225,476**	**73,109**	**31,088**

Table 1 - Details for the five hacking forums examined for this paper.

Continents	Number of IP Mentions		Number of Distinct IPs	
Africa	1914	(0.8%)	710	(1.0%)
Asia	43,689	(19.4%)	17,390	(24.7%)
Europe	34,434	(15.3%)	10,286	(14.6%)
North America	111,375	(49.4%)	30,190	(42.8%)
Oceania	2,670	(1.2%)	653	(0.9%)
South America	12,702	(5.6%)	9,041	(12.8%)
Unable to GeoLocate	18,692	(8.3%)	2,264	(3.2%)
Total	225,476	(100.0%)	70,534	(100.0%)

Table 2 - The distribution of IP addresses per continent.

Country	Mentions	Distinct
United States	106,712	28,119
China	23,964	7,712
(blank)	19,151	2,456
France	7,236	1,795
Venezuela	6,226	5,409
Netherlands	5,661	749
Russia	4,516	1,430
Brazil	4,452	2,631
Germany	3,344	877
Thailand	3,311	1,767
Indonesia	2,715	1,497
Canada	2,670	960

Table 3 - The top 12 countries discussed in terms of their IPs. Sorted by the number of mentions.

C. IP Extraction and GeoLocation Analysis

The current version of IP addresses (IPv4) follow a consistent and recognized pattern. An IP address consists of four 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. All IP addresses were extracted from all the posts collected using Regular Expressions (RegEx). Most IP address were identified only once, while some were posted multiple times within or across the hacker forums.

Geolocation refers to the process of identifying the location of internet devices (such as an IP address, a cellphone, or computer terminal) and involves the mapping of an internet protocol address to a real world geographic location of the host [25]. The result is either an address in the form of city/state/country, or a longitude/latitude pair. Accuracy is inconsistent since IP addresses are reserved for various service providers and not end users, and thus any attempt to geolocate an IP address yields information about the service provider and possibly only the rough location of the end-user. After IP address information was extracted from all the users' posts, MaxMind's GeoLite [33] 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 mainly on the country level for geographical analysis, the 99.80% accuracy rate provided sufficient confidence in the accuracy.

D. Content Analysis

Forum posts disclosing IP addresses were evaluated through a qualitative content analysis supplemented targeted keyword searches. The content analysis was approached inductively to allow for themes to emerge naturally from the data, however codes were also informed by the literature. This process was used to classify the data into general categories such as search engine optimization (SEO) tools, and to take note of cases and keywords that were particularly relevant to this study. Following this, several keyword searches were conducted across all posts on the five hacking forums to clarify the scope of discussion about specific terms, and to serve as a checking mechanism for potential coding inconsistencies. The keywords selected for query were informed by terms discussed in the literature in addition to key phrases that were noted during the coding process.

III. Results

For this paper, 5 hacking forums were downloaded in their entirety, totaling 11,062,793 posts from 372,856 registered users. These posts were then filtered to only those posts (31,088) which contained an IP address. In total there were 225,476 IP addresses across all five forums, although accounting for duplicates the extraction yielded only 73,109 IP addresses. These IP addresses were geolocated, and thus a longitude/latitude as well as city, state, country information assigned to them. This information, along with a descriptive summary of the five forums is shown in Table 1.

A. Geolocation Analysis

As seen in Table 2, there were 225,476 IP addresses mentioned across the five forums, yielding a total of 70,534 unique IP addresses (note, the 73,109 in Table 1 takes into account forums, thus, if an IP address is found in multiple forums, it's counted multiple times in Table 1, while only once in Table 2). Almost every second (49.4%) mentioned address belonged to computers within North America (defined as the United States and Canada). IPs belonging to the US accounted for 47.3% of the mentioned (39.9% of the unique) IP addresses, which is an unproportionally higher rate than the number of IP addresses assigned to the US (37.4%) [34]. Thus, although it could be argued that the number of unique IP addresses found in the hacking forums could be as a result of random chance (39.9% vs 37.4%), the US IP addresses are being discussed with much larger frequency than non-US addresses. That is, if hackers were finding, posting and discussing IP addresses with equal probability, it would be expected that approximately 37.4% of addresses found would belong to the US, but actually 47.3% do. This finding supports observations made elsewhere, which have found that the US makes up a disproportionate amount of traffic on the internet when it comes to sources of spam [36] and hosting malware (63% of malware hosts are in the US [35]). A list of the top 12 countries, in terms of their IPs being mentioned, is shown in Table 3.

B. Qualitative Analysis

To better understand the threat landscape posed by openly-accessible hacking forums, the data was broadly classified as relevant or irrelevant through identifying whether a post discussed offensive hacking tools, techniques, or targets. Relevant posts were then coded inductively. Posts that did not meet these criteria were deemed to be irrelevant. A substantial portion of the hacking forum posts and associated IP addresses analyzed were irrelevant to the current study. Irrelevant data comprised 26,665 of 31,088 posts, and 212,337 of 350,244 IP address records. Irrelevant data included automated proxy lists, pirated material, cracked software, and defensive security software (e.g., anti-malware software). However, all posts discussing exploitable vulnerabilities in security software were included in this analysis. Automated proxy lists comprised almost half of all IP address records, with 5,654 posts about proxies representing 158,618 IP addresses. The following section will discuss three major themes that emerged from the data and will evaluate other relevant concepts through analyzing a set of relevant keywords.

1) Theme #1: Hacking forum members as targets

The data indicates that hacking forum users are sometimes targeted through malicious software advertised on hacking forum threads. There were 98 posts representing 530 IP addresses, of which there were 132 distinct IP addresses. IP addresses assigned to the United States represented the largest percentage of both IP address records and distinct IP address records, at 65.8% and 36.4% respectively. The associated IP addresses were largely unrelated to the targeted users and

offenders, except for cases where a targeted user posted details about hackers who allegedly scammed them. A distribution of these IP addresses is shown in Table 4.

There were six distinct posts representing nine IP addresses that detailed some form of retribution against users who violate community norms through unfairly targeting or scamming other users. All six cases involved the aggrieved user doxxing the alleged scammer. For example:

> *"An information for the romanian BHW member that scammed me using an invitation from above through stolen paypal accounts [identifying information removed] I've reported you to the romanian authorities, maybe one of them will get his ass up and put yours into a cell." (user from forum 157)*

Five of the nine IP addresses related to doxxing targets were assigned to the US, two IP addresses were assigned to the Philippines, one IP address was assigned to Romania, and one IP address was assigned to Turkey.

2) Theme #2: Search Engine Optimization Spam

Another theme that emerged from the data was the prevalence of black search engine optimization (SEO) tools and techniques being used to redirect unsuspecting users to spam websites. The content of SEO spam discussion threads indicate that these malicious tools and techniques are being used to fraudulently generate advertising revenue, in addition to attracting users to websites laden with malware. Discussions about black SEO tools and techniques were responsible for 2,513 forum posts and 8,246 associated IP addresses (see Table 5 for top 15 countries with highest frequencies of IP addresses).

Country	# of IP Address Mentions	# of Distinct IP Address Mentions
United States	349	48
Russia	75	21
Unable to GeoLocate	30	17
China	25	19
South Africa	7	2
Ukraine	6	3
Canada	4	3
Chile	4	1
Malaysia	2	1
Netherlands	3	1
Libya	3	1
Spain	3	1
Philippines	2	2
Romania	2	2
Singapore	3	1
Germany	2	2
Turkey	2	2
Pakistan	2	1
Senegal	2	1
United Kingdom	2	1
France	1	1
India	1	1

Table 4 - IP address records for hacking forum member targets

South Africa's IP data, as seen in Table 5, stands out due to the distinct contrast between mentioned IP addresses vs. distinct IP address records, at 239 records and 3 records respectively. An evaluation of South Africa's data revealed that 237 IP records were in fact not IP addresses, but software version numbers (i.e., version 2.0.0.0 was recorded as an IP address).

3) Theme #3: Financial Sector

Consistent with prior research, the data showed that banking-related exploits, tools and services are widely discussed in hacking forums [10]. There were 152 carding and banking-related posts across all forums, representing 199 IP addresses, of which 21 were distinct. As many of the posts were related to carding dumps, the associated IP addresses were largely unrelated to the targeted users. A full breakdown of the IP addresses, by country, is shown in Table 6. Interestingly, the United States was not most frequently discussed country in this sample, contrary to the lists regarding targets and SEO spam.

Banking-related discussions primarily focused on selling credit card dumps and online money transfer accounts. However, one thread was found to have specifically targeted a Swiss financial organization:

> *"This company was conducting a promotion campaign 8 months ago. [...] Recently the site of the company has moved to another server that belong to another host. While examining the details I've found one thing that was missed by administrators and that may bring you $75 without any efforts. Your part of the deal - to be registered under me. To specify me as your referral (sponsor). The only one thing you need to do is to enter me as your referral when you are registering and I'll get my commissions that equal 9% of the sum*

Country	# of IP Address Mentions	# of Distinct IP Address Mentions
United States	2,414	988
Thailand	1,037	352
China	1,012	356
France	826	273
Unable to GeoLocate	644	154
Australia	569	119
South Africa	239	3
Germany	168	83
Malaysia	136	21
India	110	56
Russia	99	43
Japan	85	34
Syria	79	16
Brazil	64	43
United Kingdom	62	35

Table 5 - IP address records for SEO spam

of your free bonus. [...] Why I shall not open many accounts? [...] This company well watches that one person had only a unique account. And if it finds out that from one computer openly set of accounts - they immediately block these actions" (unknown individual's message posted by user from forum 160).

Although the poster provided the institution's name, they did not provide a target IP address. The IP addresses in this case were associated to China.

C. Keyword Searches

1) Zero-Day

In total, there were 40 posts with the keyword 'zero-day', representing 82 IP address records of which 59 were distinct. Various iterations of the keyword term 'zero-day' were searched to capture the data, and irrelevant posts were removed (e.g., "message is 0 days old"). None of the posts actively discussed how to exploit a zero-day vulnerability, nor did any of the associated IP addresses refer to potential targets. Most of these posts reported on and speculated about disclosed zero-day vulnerabilities, in addition to discussing how these vulnerabilities are addressed by software updates and anti-malware software. This type of discussion can be seen in the following three posts:

"The shell code in this attack calls back to IP address […] One could speculate that the server [...] was used by energycdn.com as one of their servers to host pirated content. Perhaps the server was compromised by whoever controls energycdn to host that content and then was reinfected by the perpetrator of this new malware variant. But weâ€™re [sic] speculating." (user from forum 157)

"As to browser security... "Beware the zero-day dragons" because NO browser is fully prepared for THOSE nasties." (user from forum 159)

"The script attempts do download malicious code from a web site in Russia that allows spammers to your the compromised client as a relay. It is yet unknown how the servers are being compromised..... Can we say "zero day"??? Mitigate by blocking all access to [the IP]" (user from forum 160)

2) Botnet

There were 20 posts and 188 IP address records related to the term 'botnet'. Most of these posts were unrelated to discussions about operating a botnet. The term was primarily used in posts discussing research, security software, and concerns about their own systems being infected, for example:

"I remember PrevX and BBC teamed up to do something similar. They bought a botnet and then used to it to SPAM, DDoS." (user from forum 159)

However, there was one theoretical discussion about how botnets could be used offensively:

"Depending on the configuration of the bot and threading that could of been a botnet of about 5-10k, maybe more. How it helps business? let me think 1. you dont overuse your bandwidth 2. you doesnt crash the server 3. you dont get suspended from hosting for overusing bandwidth limits. 4. you dont pay fines for overusing the bandwidth." (user from forum 157)

3) Trojan/RAT

There were 2,689 IP address records related to the terms 'trojan' and 'RAT' (abbreviation for *remote access trojan*). The term 'trojan' was primarily mentioned when concerned users posted their own logs, and in discussions about security software. However, there were also several occasions where users who wanted to download a tool for malicious purposes found that it contained a trojan.

4) Critical Infrastructure

Despite the wealth of literature indicating that industrial control systems such as SCADA are especially vulnerable to cyber-attacks, there was only one mention of the term 'SCADA' across all five forums. This term was listed in an intrusion detection software log published in one post. The term 'industrial' yielded 10 posts and 24 IP addresses across the five forums. There was no discussion about hacking into industrial systems, however there was one post discussing a cyber-attack against industrial and government systems:

"It seems that this supply-chain malware is a continuation of Operation Aurora, which was a Industrial/Governmental info attack that ha [sic] been going on for years. Originally implicated was the group Unit 61398 (Comment Crew), supplanted recently by the well funded Axiom Group. [...] Although this particular server has been brought down, within the malware code is the ability to connect to different servers in the future (I don't know if this has been reported as yet). Anyway, a typical Home user has nothing to worry about. Axiom does not even consider Chumps Like Us as being of any value." (user from forum 159)

Country	# of IP Addresses	# of Distinct IP Addresses
Bosnia & Herzegovina	92	1
Japan	77	2
No Country	11	5
United States	6	4
France	4	2
United Kingdom	4	2
China	3	3
Australia	1	1
Canada	1	1

Table 6 - IP address records for carding and banking

IV. Discussion and Conclusions

The victimization of hacking forum users by other members was contrary to expectations, given that marketplaces in online hacking communities have been found to be influenced by perceptions of trust and customer service, in a similar vein to legal marketplaces [20, 21]. The fact that some forum users doled out public shaming and reciprocal punishment via doxxing, often without admonishment from

other users, supports the notion that online hacking communities have their own set of social norms regarding acceptable behaviours and punishments.

Discussions relating to SEO tools and spam related to SEO dominated the relevant data, comprising 2,513 of all 4,423 relevant posts. Results from this study suggest that spam is still one of the most popular means of generating income. Although the analysis focused on SEO tools and related spam, discussions about several other forms of spam were prominent, notably social media spam, e-mail spam and to a lesser extent, phone spam. The prevalence of SEO spam and social media spam on hacking forums suggests that hackers are continually evolving their tools and techniques to maximize income.

Although government definitions of critical infrastructure sectors vary, the finance sector is often classified as a form of critical infrastructure [17, 18]. The results showed that, relative to other critical infrastructures such as industrial, energy, and manufacturing sectors, the financial sector is much more widely discussed on hacking forums. However, this focus is likely due to the quick and easy earning opportunities that hacking forums provide through selling access to credit card dumps, and online money transfer accounts such as PayPal and Western Union. Although some anti-establishment and anti-bank sentiment was found across the five hacking forums, this did not manifest in active discussions about targeting financial systems. Instead, discussions were restricted to basic carding and financial fraud, as well as hypothetical discussions about banking vulnerabilities. Even the forum post discussing how to exploit a particular Swiss financial organization appeared to be borne out of opportunity, rather than an intense desire to hack into financial systems.

There were several limitations that arose in this study due to time and resource limitations. First, only five forums could be analyzed, limiting the scope of potentially relevant data that could be revealed. Second, the sheer frequency of forum posts made it so that each post could only be coded once. Additionally, only one researcher qualitatively coded the data, which may make the findings more prone to bias. However, attempts were made to address this shortcoming through referring to key concepts in the literature during the coding process, and through conducting keyword searches primarily based on the literature. Another limitation of this study was the difficulties faced in accurately capturing the true number of IP addresses, as some of the IP records were in fact serial numbers and software version numbers.

Although the majority of IP addresses associated with posts across the five forums were irrelevant to this study, a small portion of the IP addresses were directly related to discussions about the hacking target or perpetrator. The IP locations of alleged scammers who had wronged hacking forum members provided some insight into the alleged scammers' location that would not otherwise be known. If IP address capture methods can be further refined to differentiate between software versions and true IP addresses, IP addresses may be more efficiently related to hacking targets and offending parties. Doing so may provide a novel approach to studying hacking forums, extending the analysis of hacking forums beyond traditional content and sentiment analyses.

Future research will expand upon the breadth of hacking forums selected for analysis. This should allow for a more balanced understanding of the current cybersecurity threat landscape on surface web hacking forums. Furthermore, this may allow for critical infrastructure targets on hacking forums to emerge, which was the original intention of this study. Future research will also involve the integration of additional keyword queries, in addition to conducting a more iterative coding process.

V. REFERENCES

1) Brocklehurst, K. (2015, Feb 1). Cyberterrorists Attack on Critical Infrastructure Could be Imminent. [Online]. Available: http://www.tripwire.com/state-of-security/security-data-protection/security-controls/cyberterrorists-attack-on-critical-infrastructure-could-be-imminent/

2) Borilin, S. (2014, Sep 24). So malware attacks against critical infrastructure are inevitable. What's next?. [Online]. Available: https://business.kaspersky.com/so-malware-attacks-against-critical-infrastructure-are-inevitable-whats-next-2/2647

3) Dingman, S., Silcoff, S., & Greenspan, R. (2014, Nov 24). Hacked: The escalating arms race against cybercrime. [Online]. Available: http://www.theglobeandmail.com/report-on-business/hacked-the-escalating-arms-race-against cybercrime/article21305464/?page=all

4) United Nations. (2011). "Cybersecurity: A global issue demanding a global approach". Available: http://www.un.org/en/development/desa/news/ecosoc/cybersecurity-demands-global-approach.html

5) Clark, R. M., & Hakim, S. (Eds.). (2017). *Cyber-physical security: Protecting critical infrastructure at the state and local level* (Vol. 3). Basel, Switzerland: Springer Nature Switzerland AG.

6) Ralston, P. A., Graham, J. H., & Hieb, J. L. (2007). Cyber security risk assessment for SCADA and DCS networks. *ISA transactions, 46*(4), 583-594. https://doi.org/10.1016/j.isatra.2007.04.003

7) Yusta, J. M., Correa, G. J., & Lacal-Arantegui, R. (2011). Methodologies and applications for critical infrastructure protection: State-of-the-art. *Energy Policy, 39*(10), 6100-6119. doi: 10.1016/j.enpol.2011.07.010

8) Geers, K. (2010). The cyber threat to national critical infrastructures: Beyond theory. *Journal of Digital Forensic Practice, 3*(2-4), 124-130. doi: 10.1080/15567281.2010.536735

9) Laugé, A., Hernantes, J., & Sarriegi, J. M. (2015). Critical infrastructure dependencies: A holistic, dynamic and quantitative approach. *International Journal of Critical Infrastructure Protection, 8*, 16-23. http://dx.doi.org/10.1016/j.ijcip.2014.12.0041874-5482/

10) Macdonald, M., Frank, R., Mei, J., & Monk, B. (2015). Identifying digital threats in a hacker web forum. *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Paris, 2015, 926-933. doi: 10.1145/2808797.2808878

11) Ducaru, S. D. (2017). The security of critical infrastructure in the age of multiple attack vectors: NATO's multi-faceted approach. *Europolity, 11*(1). Retrieved from https://doaj.org/article/2058930741a340898a69251d9bab2f3c

12) Ryu, D. H., Kim, H., & Um, K. (2009). Reducing security vulnerabilities for critical infrastructure. *Journal of Loss Prevention in the Process Industries, 22*(6), 1020-1024. https://doi.org/10.1016/j.jlp.2009.07.015

13) Johnsen, S. O., & Veen, M. (2013). Risk assessment and resilience of critical communication infrastructure in railways. *Cognition, Technology & Work, 15*(1), 95-107. doi: 10.1007/s10111-011-0187-2

14) Ridley, G. (2011). National security as a corporate social responsibility: critical infrastructure resilience. *Journal of Business Ethics, 103*(1), 111-125. doi: 10.1007/s10551-011-0845-6

15) Tuptuk, N., & Hailes, S. (2018). Security of smart manufacturing systems. *Journal of manufacturing systems, 47*, 93-106. https://doi.org/10.1016/j.jmsy.2018.04.007

16) Van Erp, J. (2017). New governance of corporate cybersecurity: A case study of the petrochemical industry in the Port of Rotterdam. *Crime,*

Law and Social Change, 68(1-2), 75-93. doi: 10.1007/s10611-017-9691-5

17) Public Safety Canada. (2014). *Forging a common understanding for critical infrastructure: Shared narrative.* Retrieved from the Public Safety Canada website: https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/2016-frgng-cmmn-ndrstndng-crtcalnfrstrctr/2016-frgng-cmmn-ndrstndng-crtcalnfrstrctr-en.pdf

18) Public Safety Canada. (2018). *Critical infrastructure*. Retrieved from Public Safety Canada website: https://www.publicsafety.gc.ca/cnt/ntnl-scrt/crtcl-nfrstrctr/index-en.aspx

19) United States Department of Homeland Security. (2017). *What is critical infrastructure?* Retrieved from United States Department of Homeland Security website: https://www.dhs.gov/what-critical-infrastructure

20) Leukfeldt, R., Kleemans, E., & Stol, W. (2017). The use of online crime markets by cybercriminal networks: A view from within. *American Behavioral Scientist, 61*(11), 1387-1402. doi: 10.1177/0002764217734267

21) Holt, T. J. (2013). Examining the forces shaping cybercrime markets online. *Social Science Computer Review, 31*(2), 165-177. doi: 10.1177/0894439312452998

22) Dupont, B., Côté, A. M., Boutin, J. I., & Fernandez, J. (2017). Darkode: Recruitment patterns and transactional features of "the most dangerous cybercrime forum in the world". *American Behavioral Scientist, 61*(11), 1219-1243. doi: 10.1177/0002764217734263

23) Public Safety Canada. (2016). *Risk management guide for critical infrastructure sectors*. Retrieved from Public Safety Canada website: https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/rsk-mngmnt-gd/rsk-mngmnt-gd-eng.pdf

24) Public Safety Canada. (2010, October 3). *Canada's cyber security strategy: For a stronger and more prosperous Canada*. Retrieved from Public Safety Canada website: https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/cbr-scrt-strtgy/cbr-scrt-strtgy-eng.pdf

25) Public Safety Canada. (2018). *2018-2020 Action plan for critical infrastructure*. Retrieved from Public Safety Canada website: https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/pln-crtcl-nfrstrctr-2018-20/pln-crtcl-nfrstrctr-2018-20-en.pdf

26) Public Safety Canada. (2017, September 29). *Horizontal evaluation of Canada's cyber security strategy: Final report*. Retrieved from Public Safety Canada website: https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/vltn-cnd-scrt-strtg/vltn-cnd-scrt-strtg-en.pdf

27) Public Safety Canada. (2010, May 28). *National strategy for critical infrastructure*. Retrieved from Public Safety Canada website: https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/srtg-crtcl-nfrstrctr/srtg-crtcl-nfrstrctr-eng.pdf

28) Herzog, S. (2011). Revisiting the Estonian cyber attacks: Digital threats and multinational responses. *Journal of Strategic Security, 4*(2), 49-60. doi: http://dx.doi.org/10.5038/1944-0472.4.2.3

29) Noguchi, M., & Hueda, H. (2018). An analysis of the actual status of recent cyberattacks on critical infrastructures. *NEC Technical Journal, 12*(2), 19-24. Retrieved from https://www.nec.com/en/global/techrep/journal/g17/n02/170204.html

30) Industrial Control Systems Cyber Emergency Response Team. (n.d.). *Overview of cyber vulnerabilities.* Retrieved from United States Department of Homeland Security website: https://ics-cert.us-cert.gov/content/overview-cyber-vulnerabilities#under

31) XPath, http://www.w3.org/TR/xpath/

32) XQuery, https://www.w3schools.com/xml/xquery_intro.asp

33) GeoLite from MaxMind. http://dev.maxmind.com/geoip/legacy/geolite/

34) Country IP Blocks. https://www.countryipblocks.net/allocation-of-ip-addresses-by-country.php. Retrieved Oct 27, 2018

35) NTTSecurity. 2017 Global Threat Intelligence Report. Available online: https://us.nttdata.com/en/-/media/nttdataamerica/files/americasd2/infrastructure_managed_services/gtir-ntt-security-ntt-data-04252017.pdf. Retrieved Oct 27, 2018

36) TrustWave. https://www.trustwave.com/support/labs/spam_statistics.asp. Retrieved, Oct 27, 2018