Networking in Child Exploitation – Assessing disruption strategies using registrant information

Russell Allsup, Evan Thomas, Bryan Monk, Richard Frank*, Martin Bouchard*
International Cybercrime Research Center
School of Criminology
Simon Fraser University
Burnaby, Canada
{rallsup, enthomas, bmm8, rfrank, mbouchard}@sfu.ca

***Abstract* — This research utilizes social network analysis to determine the success of three different disruption strategies on a child exploitation network extracted from the public internet. Using a custom-written web-crawler called LECEN, data from a set of hyperlinked child-exploitation websites was collected from the Internet. From these data, two types of networks were coded: the nodes of the first network consisted of only website domains, while the nodes of the second were generated using the registrant data, where the nodes represented the legal owners of those same domains. Three attack scenarios were carried out on these two networks: two types of hub attacks (one focused on in-degree and one focused on out-degree) and a bridge attack. Using these disruption strategies, it was found that bridge attacks were more suitable for disrupting the domain networks, while both hub-attacks could be favored when disrupting the network of registrants. These findings have implications for law enforcement, as it provides real-world applications to disruption where registrants may be targeted directly.**

Keywords: Child exploitation, social network analysis, web-crawler, internet

I. INTRODUCTION

Child sexual offenders have historically been quick to adapt technological advances, such as photography [9] and film [10] for the purposes of exploiting children. The movement of CEM to the internet has enabled child sexual offenders to form virtual communities online [11, 25, 26, 27], allowing them to more easily, and secretively, access and trade CEM, recruit co-offenders and/or business partners, as well as validate their deviant behavior amongst other child sexual offenders [9, 10].

Given the harm posed by the online spread of CEM, many organizations play a role in targeting the material online. These include law enforcement agencies, but also non-police regulatory bodies. Organizations such as INTERPOL continue to aid global police agencies by developing and maintaining the International Child Sexual Exploitation Image database [4], providing international law enforcement agencies access to known child exploitation images [6]. Private organizations, such as Microsoft© [13, 14], Google© [15], and NetClean [16] have developed technological solutions which assist law enforcement in identifying, analyzing, and storing CEM found online. Despite these efforts, current strategies to limit such content do not appear to have significantly reduced the problem. Law enforcement strategies intended to combat CEM online have included chat-room stings [7, 1], honey trap sites [2], injunctions issued against websites hosting CEM [1, 2], as well as traditional investigatory techniques adapted for online use [2, 3]. These strategies tend to falter in effectiveness for a variety of reasons. Law enforcement efforts have primarily viewed CEM targets by individual websites, divorced from the larger network of websites they operate and survive within [1].

These issues facing law enforcement and other organizations in targeting and combating online CEM points to a need for new strategies, and new ways of understanding the structure and nature of child exploitation online. Social network analysis (SNA) has been proven useful at delineating the structure of criminal organizations, such as terrorist groups [34], drug trafficking organizations [35], as well as child exploitation domains online [1, 5, 7]. From a law enforcement perspective, understanding the topography and disruption strategies of child exploitation networks holds significant value as it may detail how best to use limited resources to maximize the impact of attacks. For example, it has been suggested that when targeting online child exploitation networks, the strongest disruption attacks depended on the structure of the network, as well as the specific intention of law enforcement [1]. For example, hubs have been identified as useful for the control of specific information within terrorist networks [29], and using hub attacks has been shown to be useful in reducing density or clustering within child exploitation networks [1]. Similarly, research has posited bridge attacks as being particularly useful for small world networks such as the internet [30, 31].

What makes this research project unique, and helps improve upon previous studies focusing on child exploitation content online is the use of SNA for comparing attacks from the perspectives of both the web domains, and the registrants. Typically networks are created by linking nodes (i.e. website domains in this paper) to each other based on the number of links connecting them, and attack strategies have looked at removing those nodes based on various SNA measures. However, this assumes that attacking the websites themselves is the best strategy, and thus does not take into account the possibility that multiple websites could actually belong to a single actor (individual or group). A novel attack scenario is presented in this paper where, rather than attacking the website

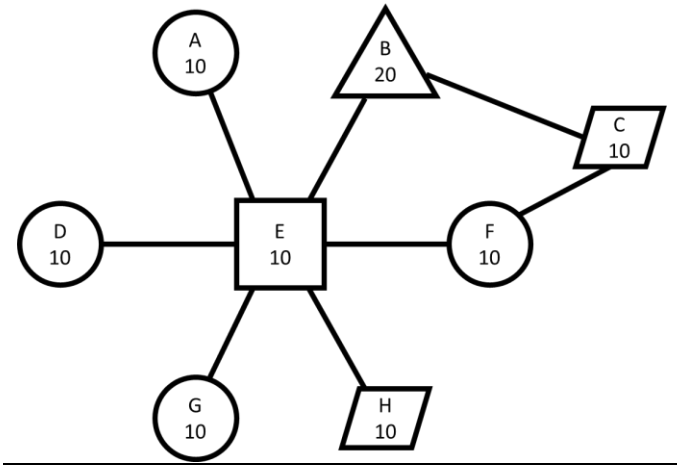

Figure 1 - Sample Network. Shapes denote website owners and numbers denote number of images on that website.

domains themselves, the actor(s) responsible for the nodes are also considered as potential targets (i.e. the registrants). This opens up the possibility that the removal of a single individual/group will remove multiple nodes simultaneously, which could significantly change the target prioritization for law enforcement. For example, using the network depicted in Figure 1, a general (non-SNA) attack focusing on only the number of images would attack node B as it has the most (20) images in the entire network. However, if the focus of the attack is to disrupt the network, then removing node B would have negligible effect as the network structure hardly changes. Instead, attacking node E would cause greater disruption as it would fragment the network into a small network (B, C, F) and 4 isolates (A, D, G and H). Unfortunately, this would only remove 10 images from the network, and would allow the owner of the content to easily replace it on another website. By taking into account the actors, options which may otherwise appear as non-optimal become arguably strong alternatives. By attacking the actor who owns nodes A, D, G and F (none of which are important, from a "key player" perspective) and thereby removing 40 images, the network is fragmented into a small network (E, B, and C) and a single isolate (H). This may lead to more disruption, even if the equivalent number of nodes is removed from the domain network. It is this latter scenario that this paper explores on a real-world set of websites that contain CEM. The attack is also compared to an attack of similar scale, without knowledge of the registrant information.

The three disruption strategies used to determine impact within the network are bridge attacks (nodes with the highest betweenness centrality scores) and two hub attacks, with one hub attack focusing on in-degree centrality and the other focusing on out-degree centrality. The change in specific network metrics (density, ties, average distance, distance based-cohesion) allows selecting both the most appropriate disruption strategy, as well as determining through which strategy to attack the network; removing prominent domains (nodes) or removing the domains belonging to prominent registrants (actors).

The structure of the paper is as follows. In section II, a description of the web-crawler used to extract the child exploitation network is given, while also describing our data collection method, construction of the network, as well as specific attack measures. In section III.A, the overall topography of the network is presented, as well as a break-down on the results of the three separate disruption attacks, these being a bridge attack and two hub attacks focusing on in-and out-degree centrality respectively. In section IV, an interpretation of the results is given, and an argument made for situational prioritization of domain owners within the online child exploitation network. Section V concludes, highlighting research limitations and future research propositions.

II. METHODS

Web-crawlers are the tools used by search-engines to navigate the Internet and collect information about each website and webpage. Search engines use them to collect data which allows users to perform queries to find information. They can also be used to seek out specific content, such as child exploitation material (CEM), as in the case of the study presented in this paper. Given a starting webpage, web-crawlers will recursively follow the links out of that webpage, until some user-specified termination conditions apply. For each web-page, the web-crawler will keep track of all the links between other websites and follow them to retrieve those as well.

To perform this research a software tool called the "*L*ocation *E*xtraction of *C*hild *E*xploitation *N*etworks" (LECEN) was utilized. LECEN is a customized web-crawler designed at the International Cybercrime Research Centre at Simon Fraser University [18]. It builds upon the traditional capabilities of an existing customized web-crawler called CENE in order to investigate and analyze CEM networks online [5, 6]. LECEN has the unique ability to identify registrants, their physical address, and the domains which belong to them, allowing us to identify potential major players based on an individual's location within the network. Therefore, this research seeks to measure the impact three different disruption attacks will have based on two different strategies for prioritizing removal of nodes within the network, while identifying commonalities and differences between both approaches.

A. Data Collection

LECEN starts by downloading a set of webpages which have been identified by the operator as containing CEM. In total, LECEN used 1, 269 starting webpages (or seeds) spread across 132 different domains in order to identify other CEM material on the public internet. These webpages were identified through a previous crawl performed by LECEN as containing either confirmed CEM, or material which was identified as indeterminate based on the judgment of investigators within the Royal Canadian Mounted Police's external hash database. For each webpage, the source HTML was retrieved, the content analyzed, and if certain requirements were not met the webpage was dropped from the queue and no further analysis was performed. If the requirements were met (namely, that a confirmed CEM hash value was located, the webpage was considered 'on topic', stored in the database, and the process repeated for all links within that webpage.

Without scripted rules in place LECEN (as well as all web-crawlers) would ceaselessly and blindly capture all web content.

To counter these limitations, LECEN has been designed with specific rules which ensure only relevant materials are captured during the crawling process. As LECEN collects and follows links outside of the starting seed sites, it simultaneously collects user requested statistics on each webpage, including relevant key-words, code-words, image hash values, and the presence of videos. To ensure accuracy when crawling, LECEN must identify a minimum of seven unique user-specified keywords or code words on each web-page in order to extract that webpage. The threshold of seven was previously identified as an indicator of the potential presence of CEM in previous research [7], but it does not exclude false positives. Because seeds had confirmed CEM, all domains examined in this study should be considered as directly or indirectly associated with a domain containing illegal material (as opposed to necessarily containing CEM themselves).

For each image it encounters, LECEN queries an external MD5 hash database provided by the Royal Canadian Mounted Police in order to determine whether CEM is located on any crawled webpage. MD5 is a widely used and secure hash algorithm, which is used primarily for calculating digital signatures [28]. The images within the database are classified into three distinct categories. Category 1 is confirmed child pornography imagery under section 163.1(1) of the Canadian Criminal Code. Category 2 contains content displaying pornography or an individual engaged in explicit sexually activity, where the age of the individual was uncertain. Category 3 images consist of other video files and digital images found alongside CEM, but are not CEM themselves. Last updated June 1, 2012, the database contained 702,997 Category 1, 2,109,813 Category 2 and 49,419,190 Category 3 image hash-values. For the purposes of this study, any webpage that contained Category 1 or 2 images was automatically deemed to be of interest.

Unlike normal web-crawlers, LECEN has been designed to incorporate Whois lookup functionality as it is crawling. Thus, LECEN's Whois functionality retrieves the administrative, technical and registrant owner contact information for each domain it encounters, allowing LECEN to determine the physical location of each type of contact. The registrant of a domain is considered the domains' legal owner. It should be noted that all registrant information collected and used was public and in accordance with ICANN WHOIS lookup terms of use [36].

B. Constructing the Network

LECEN began crawling with 1,269 seed webpages collected from previous research at the ICCRC. With those seeds LECEN crawled the public internet from December 26, 2014 to January 10th, 2015 while searching for the presence of category 1 or 2 imagery. In total, LECEN retrieved 54, 156 web-pages and all images located within. From these downloaded webpages, LECEN identified a total of 4,937 different category 1 and 2 images (1,569 category 1 images and 3,368 category 2 images). These images were spread across 1,166 different webpages on 82 servers/domains. The resulting web-crawler data was used to construct two networks (Table 1). It should be stated that at no point does LECEN contravene or enter password protected websites.

The first network, referred to as the "Domain Network", was focused on the domains of the websites, where the nodes consisted only of website domains, while edges in the network represented the number of hyperlinks between the two corresponding domains. The second network, referred to as the "Registrant Network", focused on the registrant data, where the nodes represented the legal owners of those same domains identified in the Domain Network, with the edges representing the number of hyperlinks between the sites that those registrants owned. Ties within the network were the summed links between domains. For example, using Figure 1, domains {A, D, F, G} are registered to Registrant1, and domain {C, H} are registered to Registrant2. Any links between {A, D, F, G} and {C, H} would be summed as the number of links between Registrant1 and Registrant2. The registrant network was also used to evaluate which registrant owned the most important domain group in terms of (a) betweenness centrality, (b) in-degree centrality, and (c) out-degree centrality. The properties of each node include the number of images on that domain, and the owner (name and address) of the owner.

The challenge in comparison arises from the discrepancy between using the LECEN data (i.e. registrants) which are aggregate data of the domain data and the domains themselves which are individual data. Further, simply creating a network of registrants loses valuable data on the network structure: the path between nodes has practical value in terms of traveling between domains as a user, an important point when discussing "disruption". In order to mitigate valuable data loss on network structure, the network was viewed from the perspective of registrants. Thus, during the exploratory attack scenario nodes were removed from the domain network, while the registrant view of the network was used to prioritize the nodes which would be removed in the Registrant Prioritized Attack.

C. Attack Measures

This research focuses on identifying the most impactful disruption attacks which would most affect the extracted child exploitation network. To meet this objective, both bridge attacks (utilizing betweenness centrality) and hub attacks (utilizing degree centrality) were used to determine which nodes, having been removed, would most impact the structure and characteristics of the network. Both bridge and hub attacks were used on both the domain network, as well as the registrant network in order to determine from which perspective law enforcement would be most disruptive.

Bridge attacks identify and target nodes with high betweenness centrality scores, where betweenness centrality indicates the degree to which node brokers access between otherwise unconnected others [19]. The betweenness centrality of the node *u* is defined as:

$$B_u = \sum_{\substack{u,w \in V, \\ u \neq v \neq w}} \frac{\sigma_{v,w}(u)}{\sigma_{v,w}}$$

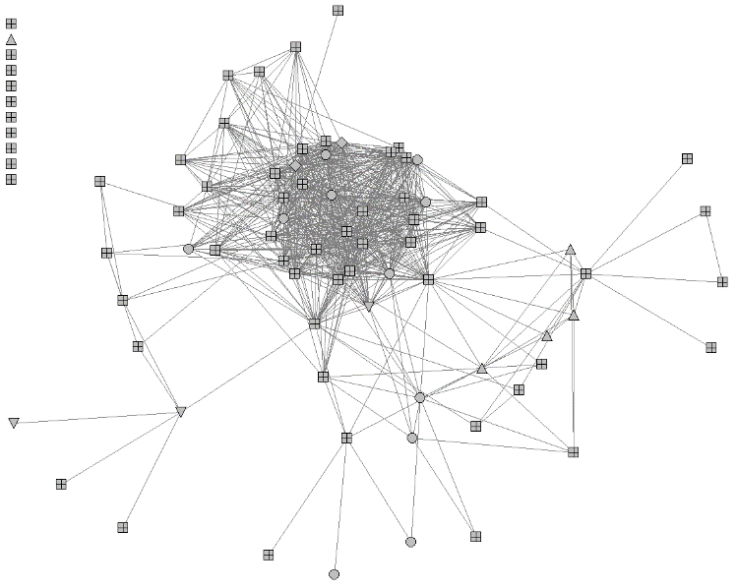

Figure 2 – Domain Network

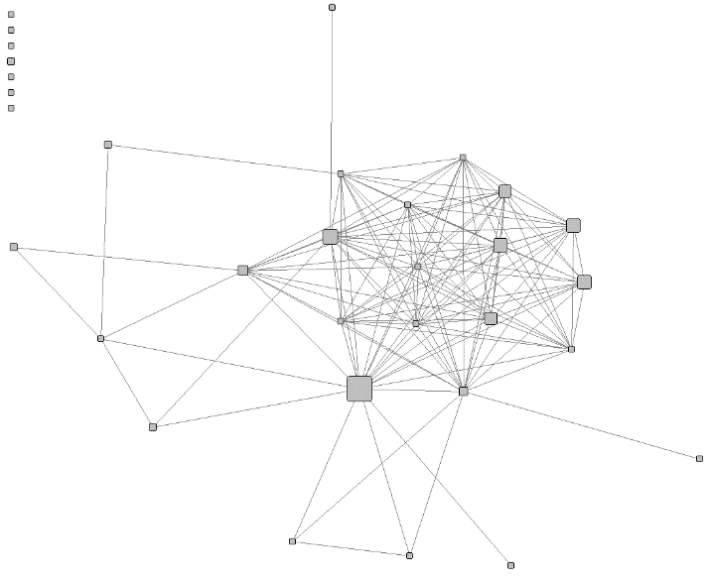

Figure 3 – Registrant Network

where $\sigma_{v,w}(u)$ represents the total number of shortest paths between each pair of nodes such as *v* and *w* that pass through node *u* and $\sigma_{v,w}$ denotes the total number of shortest paths from *v* to *w* [23].

Alternatively, hub attacks identify and target nodes with high degree centrality scores, which have the largest number of incoming and outgoing links to other nodes in the network [1]. In other words, a domain or actor is viewed as a key player based on the total number of direct connections. Therefore, this measure focuses on the most visible nodes in the network. A node with a high degree is in direct relationship or is neighbor to many other nodes. Nodes should be recognized by other nodes as main channels of information spreading, indeed, a crucial cog in the network occupying a central position [24]. In contrast, actors with low degrees are more peripheral in the network and these nodes are not as active in the connection process. Degree centrality of an actor *u* is calculated as follows:

$$C_D(u) = \frac{d_u}{N-1}$$

where d_u is the number of first level neighbors of u, and N is the total number of actors in the network [24]. In-degree centrality for a node is defined as the fraction of nodes its incoming edges are connected to; in this study the fraction of domains that have links to a given domain. This is in contrast to the out-degree centrality, which is defined similarly but based on the nodes' outgoing edges; i.e. the fraction of domains that a given domain has links to, out of all the possible links.

D. Attacking the Network

Three different attacks for removing nodes within the network were employed on both the domain network and the registrant network: hub-attack (in-degree measure), hub-attack (out-degree centrality measure), and bridge attack (betweenness centrality measure). It should be noted that for the purposes of this research, all disruption attacks were done in simulation, and at no time were any domains attacked or brought offline.

Given this information, the research presented in this paper used multiple disruption strategies to provide a number of flexible possibilities. The attacks had two subsets. First, prioritizing the attack using a "Domain Prioritization Strategy" where the characteristics of the domains were analyzed to determine removal order in the Domain Network. Second, prioritizing the attack using a "Domain Prioritization Strategy" where the characteristics of the registrants (in the Registrant Network) were used to determine removal order in the Domain Network. The attack strategy followed this process. First, in order to evaluate the success of three different disruption strategies on the domain network, the desired attribute of the registrant network was used to prioritize targets in the domain network. As domain groups were removed based on the value of their registrant, an equal amount of nodes were removed from the domain network prioritized on the attribute of each individual. After each removal, the corresponding analysis was re-run to prioritize the next target. The process was repeated until the next removal would exceed the top quartile of nodes - the top quartile encapsulated the most central (or important) nodes while still being parsimonious enough to measure the impact across multiple strategies In this instance the top-quartile was selected in order to evaluate the impact the different strategies had on their respective key players. In application by law enforcement, where target selection is restricted by practical and legal considerations, the domain or user count would be even smaller.

E. Outcome measures

Density is one of the key measures used to determine network topography, both before and after removal of perceived key-players. Density is defined as the total number of ties present within a network divided by the total number of possible ties

			Measure			
	Attack Strategy	Attack Removed (total)	Density	Ties	Distance-cohesion	Average distance
Wave 1	Registrant Prioritized	1 registrant	0.123	780	0.209	1.894
	Domain Prioritized	3 nodes	0.114	720	0.2	2.135
	Difference	---	*7.59%*	*8.00%*	*4.40%*	*11.96%*
Wave 2	Registrant Prioritized	2 registrants	0.149	639	0.207	1.479
	Domain Prioritized	17 nodes	0.082	352	0.116	1.615
	Difference	---	*58.01%*	*57.92%*	*56.35%*	*8.79%*
Wave 3	Registrant Prioritized	3 registrants	0.159	602	0.217	1.443
	Domain Prioritized	21 nodes	0.06	249	0.09	1.899
	Difference	---	*90.41%*	*82.96%*	*82.74%*	*27.29%*

Table 2 – Bridge Attack

within a network. In this case ties represent the links between the domains as explained in section II.C. By assessing the density both before and after any form of disruption, we can determine the change in relationships and ties between actors [32]. This measure assesses how connected each actor is and selected to indicate how easy it is to travel from one node to another in a cluster.

In-degree centrality was chosen as a measure as it suggests that actors who receive many incoming ties are considered prominent within the network [20]. Similarly, out-degree centrality suggests that actors within the network with a large number of outgoing ties are capable of sharing information with others within the network, and are therefore considered influential within the network [20]. In both respects, a key-player is identified by their high number of incoming, or outgoing ties within the network.

Average distance refers to measuring the average length of all the shortest paths between all connected nodes within the network. Average distance makes for a useful measure as it can point to the speed with which resources (in this case information or CEM material) diffuses within the network [35]. This measure was selected to represent changes in total network accessibility over particular individuals or speed of access by actors in the network.

Distance based cohesion is the aggregate measure of connectivity that includes the sum distances of the ties. It is utilized to illustrate that the disruption strategies have different impacts depending upon the end goal of the research. This allows for observation of how wide reaching the network is as opposed to the centrality measures which are much more focused. All outcome measures were considered in the three attack strategies when considering impact.

Measure	**CEM network**
Nodes	82
Ties	882
Graph Centralization (IN)	2.12%
Graph Centralization (OUT)	2.61%
Average Distance	1.957
Distance-based cohesion	0.235

Table 1 - Network Topography

III. RESULTS

A. Topography

Data collection for the purposes of this research took place on Dec 26, 2014 and ended on Jan 10, 2015. In total, LECEN retrieved 54,156 web-pages and all images located on them. There was a total of 4,937 different category 1 and 2 images (1,569 category 1 images and 3,368 category 2 images). These images were spread across 1,166 different webpages and 63 servers/domains.

Furthermore, the network's overall topography was assessed. The extracted child exploitation network contained a total of 82 nodes and 882 ties. The networks average distance was 1.957, and the overall density was 0.13. A summary is shown in Table 1, with Figure 2 showing the domain network, and Figure 3 showing the registrant network as visualized in UCINET [33].

B. Bridge Attacks

Registrant Prioritized and Domain Prioritized attack strategies were first used to demonstrate the difference in network structure when interrupting the connections between domains using a node's betweenness centrality score (Table 2). The stark contrast between the two measures indicates the superiority of the Domain Prioritization's impact on segmenting networks using betweenness as an attack measure. This is demonstrated with three waves of attack which resulted in the removal of the top 25% (quartile) of the nodes. The result indicated the Domain Prioritization attack left a widely dispersed and lightly connected network with a higher average path length (indicating less efficiency in a user's ability to travel from one domain to another).

Through all three waves, the Registrant Prioritized attack strategy showed only minor reduction in distance-based cohesion and ties – in fact becoming denser as many periphery nodes were removed (Table 2). The average-distance of path length remained lower than that in the Domain Prioritized attack, indicating it is easier to travel from one domain to another on average, and so a longer average is desirable in terms of disruption.

In the first wave, three nodes were removed premised upon the registrant's ownership of nodes with the highest

			Measure					
	Attack Strategy	Attack Removed (total)	In Centrality	Out Centrality	Density	Ties	Distance-cohesion	Average distance
Wave 1	Registrant Prioritized	1 registrant	3.848	2.337	0.104	624	0.204	2.111
	Domain Prioritized	5 nodes	3.307	3.307	0.1132	680	0.21	2.041
	Difference	---	*15.12%*	*34.37%*	*8.47%*	*8.59%*	*2.90%*	*3.37%*
Wave 2	Registrant Prioritized	2 registrants	3.03	2.48	0.08	407	0.164	2.473
	Domain Prioritized	11 nodes	3.03	2.48	0.095	486	0.197	2.187
	Difference	---	-	-	*17.14%*	*17.69%*	*18.28%*	*12.27%*
Wave 3	Registrant Prioritized	3 registrants	7.909	1.338	0.063	261	0.133	2.562
	Domain Prioritized	18 nodes	2.844	2.822	0.097	418	0.188	2.017
	Difference	---	*94.21%*	*71.35%*	*42.50%*	*46.24%*	*34.27%*	*23.80%*

Table 3 – In-Degree Hub Attack

betweenness centrality score in the registrant prioritized network. Although the number of ties was reduced (882 to 720), distance-based cohesion remained relatively similar after this attack (0.235 to 0.2). However, average distance was reduced by 3.2% in the registrant prioritized network (1.957 to 1.894), indicating that this method of attack was primarily targeting and removing the more peripheral players while simultaneously confining the central group. Conversely, a 9.1% higher average path length (1.957 to 2.135) was observed following the attack, indicating a less efficient network overall.

In the final wave a total of 21 nodes had been removed from the network, with the overall density of the network when prioritizing domains (0.06) being much lower than that of the registrant prioritized attack strategy (0.159), with substantially lower ties (249 compared to 602) and distance-based cohesion (0.09 compared to 0.217).

C. In-Degree Hub Attacks

As previously discussed, in-degree hub attacks demonstrate the impact of removing the central hubs of attention (those domains that a lot of other webpages link to). By comparing the disruption of this attack between the two strategies, we were able to evince the trade-off in merely focusing on the in-degree criteria. At the cost of a more in-degree centralized network, other network measures are reduced using the Registrant Prioritization. In keeping with the premise established in the bridge attack, the hub-attack (based on in-degree centrality) removed nodes in three sequential waves (Table 3).

Barring the disparity between the attack strategies in terms of out-degree centrality - wave 1 and 2 had similar reductions across all measures. This was due to a large overlap in the nodes which were removed. During wave 3 (with an additional 7 nodes removed) a more dramatic departure was noted from the two attack strategies. Domain prioritization showed a large disparity between in- (7.909) and out-degree (1.338) centrality, leaving a highly centralized dense cluster, but with the main nodes which were sending users to other hubs having been reduced to more marginal numbers (0.063).

While the Domain Prioritized strategy had greater reductions in in-degree centralization, the Registrant strategy had greater reductions in all other outcome measures. This result was due to the inclusion of nodes that would have failed to be included when focusing on the Domains only. The average path length was higher in the Registrant Prioritization, indicating a less efficient network (2.565 to 2.017).

D. Out-Degree Hub Attacks

The out-degree hub attack focus on those nodes that link to a lot of other nodes. This attack demonstrated close parallels in outcome measure reductions in all cases except degree centrality. This result demonstrates (across multiple waves) that only a small amount of efficiency of the attack is lost by emphasizing the registrant importance over the domain's importance – and, as Table 4 indicates, out-degree centrality is reduced further by the Registrant Prioritization attack. In keeping with the hub based attacks, this disruption strategy focused on out-degree centrality, resulting in four waves (Table 4). In wave 1, when comparing network measures between the Domain Prioritization and Registrant Prioritization, there was a small percentage difference in all categories (from 0.48% to 3.29%), which favored the domain prioritization. The same similarities between strategies were observed in waves 2 and 3, including a large increase in wave 2.

The Registrant Prioritization demonstrated less centralization than the Domain Prioritized attack until wave 4, where the disparity narrowed in out-degree (3.09 and 4.296) and it increased in in-degree centralization to 7.226. Wave 4 provided the most interesting result when considering all attack strategies and all changes accrued within each wave. It was noted that with wave 4, aside from the variance in the in-degree centralization, the cost of using the Registrant Prioritization was by far the lowest. Except for in-degree, every measure provided a level of disruption and was clearly comparable with the domain prioritization strategies. Similar to the out-degree based attacks, the cost of using the Registrant Prioritized strategy is the reduction in centrality in the network, but with a nearly equal or superlative reduction in accessibility and navigability of the network with all other measures only having a difference between 1.5% and 4.6% (Table 4).

			Measure					
	Attack Strategy	Attack Removed (total)	In Centrality	Out Centrality	Density	Ties	Distance-cohesion	Average distance
Wave 1	Registrant Prioritized	1 registrant	2.181	2.395	0.11	658	0.211	2.054
	Domain Prioritized	5 nodes	3.307	3.307	0.1132	680	0.21	2.041
	Difference	---	*41.03%*	*31.99%*	*2.87%*	*3.29%*	*0.48%*	*0.63%*
Wave 2	Registrant Prioritized	2 registrants	3.536	4.949	0.111	647	0.213	2.052
	Domain Prioritized	6 nodes	6.28	4.19	0.111	651	0.209	2.055
	Difference	---	*55.91%*	*16.61%*	-	*0.62%*	*1.90%*	*0.15%*
Wave 3	Registrant Prioritized	3 registrants	1.973	2.154	0.089	440	0.173	2.088
	Domain Prioritized	12 nodes	4.49	4.13	0.102	506	0.189	2.095
	Difference	---	*77.89%*	*62.89%*	*13.61%*	*13.95%*	*8.84%*	*0.33%*
Wave 4	Registrant Prioritized	4 registrants	7.226	3.091	0.089	381	0.162	2.103
	Domain Prioritized	17 nodes	4.737	4.296	0.085	366	0.168	2.135
	Difference	---	*41.61%*	*32.62%*	*4.60%*	*4.02%*	*3.64%*	*1.51%*

Table 4 – Out-Degree Hub Attack

IV. DISCUSSION

The first objective of this research was to test the impact three different disruption strategies would have on an extracted child exploitation-related network. The strategies chosen to disrupt the network were a bridge attack, a hub-attack focusing on out-degree centrality, and a hub- attack focusing on in-degree centrality. The element which makes the study unique, is that it compares the effects of these attacks on two different types of network: the website domain, and the registrants of these domains. Network studies typically consider a single type of node, and research on the networks of child exploitation websites had identified the consideration of registrants as a potentially important development for the field [5, 6]. Considering registrants is important because 1) the same individuals may be responsible for multiple websites, which is important information when contemplating disruption; and 2) it allows law enforcement to consider punitive measures on real individuals facilitating the existence of child exploitation on the public internet as opposed to a potentially more diffuse impact on internet domains.

The results suggested that some disruption strategies favored attacking the domain on its own merits, while others were more successful focusing on the registrant. Bridge attacks were most impactful when applied to the domain network, as opposed to the registrant network. Bridge attacks prioritizes nodes which broker access to other clusters within the network. Targeting these websites may lead to a decrease in bridges within the network, which reduces an individual's ability to navigate to other child exploitation sites. Hub attacks, on the other hand, were most impactful on the registrant network. This was found for both in-degree, and out-degree attacks.

It is possible that given the choice between theoretically disrupting the online network by targeting prominent domains, or targeting prominent owners of these domains in an off-line setting, law enforcement may favor penalizing individuals directly if the results of the disruption are comparable to those targeting the domains. The notion that arresting one individual could disrupt the network substantially by taking out multiple prominent domains within the network affords law enforcement flexibility in using limited resources to combat child exploitation online.

Future research should adopt longitudinal research designs, in order to explore the life-span of child exploitation domains. This article takes a snap-shot over a short period of time, and little is known regarding the long-term functionality of these illicit domains. It is difficult to determine the efficacy of attack strategies over time, as well as the changes in the network due to such attacks outside of the simulations performed for the purpose of this study.

The current design of LECEN permits navigating only the public internet for the purposes of finding target information. Although CEM continues to persist on the public internet, adjusting LECEN's parameters in order to allow it to explore the Tor network may provide larger sample sizes and unveil the presence of previously unidentified CEM.

There are a number of other limitations to this research that should be considered in interpreting these results. First, as mentioned before, although all websites are at least indirectly associated to child exploitation material via hyperlinks, there is the possibility that some websites captured by LECEN do not contain child exploitation content. However, this potential hurdle was limited as only domains which contained a category 1 or category 2 hash value were included in the network for the purposes of analysis. It should be noted that LECEN does not capture webpage protected by entry passwords. Although this may limit the amount of material and webpages captured for analysis, it reflects material which is publically accessible without contravening website and forum password protection. Additionally a significant number of individuals (over 50%) utilized a paid service to hide the registrant data associated with their domain. These registrants were still included for the purposes of analysis, but it must be considered that many

domain hosts obfuscate their true identities behind protection services and pseudonyms.

V. Conclusions

This research assessed the impact of various disruption strategies on a child exploitation network by way of an automated internet web-crawler called LECEN. A particularity of the research is that it considered a network of website domains, but also created a network of the legal owners of these domains, in order to account for the possibility that multiple websites are run by the same individuals. This kind of overlap is especially troublesome in the context of planning interventions on website domains.

It was found that bridge attacks favored disruption at the domain priority level, while hub attacks showed results which would favor disruption directly at the registrant level. This research thus offers a unique perspective upon disrupting online child exploitation networks with practical implications for law enforcement agencies tasked with combatting them.

VI. Acknowledgements

The authors of this paper would like to thank the Canadian Internet Registration Authority who, through their Community Investment Program, funded the extension of CENE into LECEN, and thus the research presented in this paper. The authors would further like to thank *whoisxmlapi.com* for graciously providing the Whois lookups for this research free of charge.

VII. References

1) Joffres, K., Bouchard, M., Frank, R., & Westlake, B. (2011). Strategies to disrupt online child pornography networks. Proceedings of the 11th ACM SIGKDD Workshop on Intelligence and Security Informatics.

2) Wortley, R., & Smallbone, S. Child Pornography on the internet. (2006). Washington, DC: Office Of the Community oriented Policing Services.

3) Jewkes, Y., & Andrews, C. (2007). Internet child pornography: International responses. In Y. Jewkes (Ed.), Crime Online (pp. 60-81). Portland, Ore: Willan Publishing.

4) Crimes against Children. (n.d.). Accessed August 3, 2014 from the INTERPOL website: http://www.interpol.int/Crime-areas/Crimes-against-children/Crimes-against-children

5) Westlake, B., Bouchard, M., & Frank, R. (2011). Finding the Key Players in Online Child Exploitation Networks. *Policy & Internet*, 3(2), 104.

6) Westlake, B., Bouchard, M., Frank, R. (2012). Comparing methods for detecting child exploitation content online. *Proceedings of the European Intelligence and Security Informatics* 2012.

7) Frank, R., Westlake, B., & Bouchard, M. (2010). The structure and content of online child exploitation networks. Proceedings of the tenth ACM SIGKDD Workshop on Intelligence and Security Informatics '04.

8) Internet Usage Statistics. (2014). Accessed March 16, 2015 from http://www.internetworldstats.com/stats.htm

9) Broughton, D.D. (2009). Child exploitation in the 21st century. *Paediatrics and Child Health*, 19, S197-S201. doi: 10.1016/j.paed.2009.08.006

10) Cohen-Almagor, R. (2013). Online Child Sex Offenders: Challenges and Counter-Measures. The Howard Journal of Criminal Justice, 52(2), 190-215. doi: 10.1111/hojo.12006

11) Beech, A. R., Elliott, I. A., Birgden, A., & Findlater, D. (2008). The internet and child sexual offending: A criminological review. Aggression and Violent Behavior, 13(3), 216-228. doi:10.1016/j.avb.2008.03.007

12) Engeler, E. (2009, September 16). UN expert: Child porn on internet increases. The Associated Press. Retrieved from: http://abcnews.go.com/Technology/wireStory?id=8591118.

13) Microsoft Collaborates with Global Police to Develop Child Exploitation Tracking System for Law Enforcement Agencies. (2005, April 7). Accessed August 3, 2014 from the Child Exploitation and Online Protection website: https://www.microsoft.com/en-us/news/press/2005/apr05/04-07cetspr.aspx

14) Protecting our Children. (2011, October 3). Accessed August 3, 2014 from the Federal Bureau of Investigation website: http://www.fbi.gov/news/stories/2011/october/cyber_100311/cyber_100311\

15) Shiels, M. (2008). Google tackles child pornography. Accessed August 3, 2014 from the BBC website: http://news.bbc.co.uk/2/hi/7347476.stm

16) Microsoft and NetClean provide PhotoDNA Technology to help law enforcement fight online child sexual exploitation. (2012, March 19). Accessed August 3, 2014 from the Netclean website: https://www.netclean.com/en/press/microsoft-and-netclean-provide-photodna-technology-to-help-law-enforcement-fight-online-child-sexual-exploitation/

17) J. McLaughlin, "Cyber child sex offender typology", 2004. Available at: http://www.ci.keen.nh.us/police/typology.html

18) Monk, B., Allsup, R., Frank, R. "LECENing places to hide: Geo-Mapping Child Exploitation Material", Intelligence and Security Informatics Conference 2015, Phoenix, AZ. Forthcoming

19) Wasserman, S. and Faust, K. Social Network Analysis: Methods and Applications . Cambridge University Press, Cambridge, 1994.

20) R. Hanneman, and M. Riddle, Introduction to Social Network Methods. Riverside, CA: University of California, Riverside, 2005.

21) V.E. Krebs, "Mapping networks of terrorist cells," Connections, vol. 24, pp. 43-52, 2002

22) S. Milgram, "The small world problem," Psych. Tod., vol. 1, pp. 61-67, May 1967.

23) Carrington, P., Scott, J. and Wasserman, S. (2005). Models and methods in Social Network Analysis. Cambridge University Press, Cambridge, 2005.

24) Wasserman, S. and Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge University Press.

25) Westlake, B., Bouchard, M. (Forthcoming). Criminal Careers in Cyberspace: Examining Website Failure within Child Exploitation Networks. *Justice Quarterly*.

26) Joffres, K., Bouchard, M. (2015). Vulnerabilities in online child pornography networks.pp. 153-175 In Aili Malm and Gisela Bichler (Eds), *Using Network Analysis to Prevent Crime*. Crime Prevention Studies, Monsey, NY: Criminal Justice Press.

27) Tremblay, P. (2006). Convergence settings for nonpredatory 'Boy Lovers'. In R. Wortley & S. Smallbone (Eds.), Situational prevention of child sexual abuse (pp.145-168). Monsey, NY: Criminal Justice Press.

28) Deepakumara, J., Heys, H. M., & Venkatesan, R. (2001). FPGA implementation of MD5 hash algorithm. In Electrical and Computer Engineering, 2001. Canadian Conference on (Vol. 2, pp. 919-924). IEEE.

29) Minor, T. (2012). Attacking the Nodes of Terrorist Networks. Global Security Studies, 3(2), 1-12.

30) Holme, P., Kim, B. J., Yoon, C. N., & Han, S. K. (2002). Attack vulnerability of complex networks. *Physical Review E*, *65*(5), 056109.

31) Xu, J., & Chen, H. (2008). The topology of dark networks. Communications of the ACM, 51(10), 58-65.

32) Everton, S. F. (2012). Disrupting dark networks (Vol. 34). Cambridge University Press.

33) Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

34) Bouchard, M., Nash, R. (2015). Researching Terrorism and Counter-Terrorism through a Network Lens. In M. Bouchard (Ed.). *Social network, terrorism and counter-terrorism: Radical and connected*. New York; Routledge.

35) Morselli, C. (2009). *Inside criminal networks.* New York: Springer.

36) ICANN WHOIS. (2015). Notice, Disclaimers, and terms of use. Retrieved from http://whois.icann.org/en