LECENing places to hide: Geo-Mapping Child Exploitation Material

Bryan Monk, Russell Allsup, Richard Frank
International CyberCrime Research Center
School of Criminology
Simon Fraser University
Burnaby, Canada
{bmm8, rallsup, rfrank}@sfu.ca

***Abstract* — The advent of the internet has unfortunately increased the scale and complexity of child exploitation material (CEM) with content increasingly moving online, forming online CEM networks through a series of websites that are hyperlinked to each other and lead consumers from one website to another. Extending on prior research focusing on examining network structure and network disruption strategies it was prudent to expand avenues to increase attack strategies. Geolocation and Whois data were utilized to map the prevalence of CEM globally. Differences in the Geolocation and Whois data were observed, suggesting both are critical pieces of information in generating accurate geo-mapping of CEM. These maps show how multi-jurisdictional attack strategies may be employed to attack these networks and remove this content.**

Keywords: child-exploitation, web-crawling, geolocation, Whois

I. INTRODUCTION

Despite the established harm inherent within child exploitation imagery and distribution online, current attempts to limit such content have been largely unsuccessful [1]. Law enforcement strategies intended to target child exploitation material (CEM) online have included chat-room stings [2], honey trap sites [3], injunctions issued against websites hosting child pornography [3], and traditional criminal investigations and investigatory techniques adapted for online use [4]. Strategies employed by law enforcement tend to view and investigate child exploitation websites and offenders in isolation, divorced from the larger and more important network of child exploitation websites surrounding and maintaining them [1]. This is in no way a slight against law enforcement, as challenges relating to jurisdiction, resources and manpower may arise and differ substantially between police agencies [5]. Furthermore, offenders are increasingly aware of common police investigatory techniques, and are themselves manipulating technology to elude prosecution [6].

It has become increasingly true that with the issues facing law enforcement in investigating CEM online, safeguarding of the web is also being undertaken by both governmental and private organizations. Groups such as INTERPOL continue to aid global police agencies by developing and maintaining the International Child Exploitation Image Database [7]. Private organizations such as Microsoft© have developed software tools such as the Child Exploitation Tracking System (CETS), which is used by police agencies worldwide primarily as a repository where police agencies internationally may share information and child exploitation images for the purposes of analysis in investigations [8].

Along with attempts to further understand child exploitation and its characteristics online, researchers have attempted to generate tools capable of automatically detecting and classifying images based on their content and visual properties. Three different technological methods have been utilized for locating and filtering pornographic content online: keyword based, exclusion, and content-based filtering [9]. Keyword based methods involve filtering images on the basis of specific words or text either contained within the name of an image or located on that image's web page [9]. This detection method can easily be bypassed by intentionally misspelling words or mislabeling the true content of images they distribute and host. Exclusion filtering involves the use of a blacklist, which is simply a constructed collection of websites blocked due to the content located on their site [10]. Image content can also be used to identify similar images, through the detection of similar skin color and texture [11]. Generally, this method has involved searching for images containing skin-color pixels and matching their geometry alongside that of the human anatomy in other images [12]. Bag-of-visual words (BOVW) involves feature extraction and has demonstrated success in scene retrieval [13], detection of nudity and general adult content in videos [14], as well as in still images [15].

Law enforcement also faces jurisdictional issues given the variability of legislation concerning CEM internationally, likely deterring or halting entirely attempts to investigate CEM online. Problems concerning lack of manpower, resources, and cooperation between provincial and federal levels of government have been previously cited as barriers to successful CEM investigations [4]. Because of these and other issues, law enforcement success has been limited in combatting child exploitation online [1]. Identifying CEM which is readily reachable through local legislation or jurisdiction would allow law enforcement to prioritize CEM, which could present the best result given the amount of resources available.

To that end, this research project extends work on previous research [16, 17, 1, 18] using a custom-written web-crawler called the Location Extraction of Child Exploitation Networks (LECEN) which seeks out, analyzes, and partially maps child exploitation material on the internet. For each CEM

encountered on the Internet, LECEN geolocates the domain hosting the webpage, the hosting service where the image resides, and also retrieves the Whois registrant information for that domain. This allows LECEN to literally create a map of the geographical distribution of CEM, allowing law enforcement to easily see content which is reachable through their local jurisdiction.

II. METHODS

Web-crawlers are the tools used by all search-engines to automatically navigate the Internet and collect information about each website and webpage. Given a starting webpage, they will recursively follow the links out of that webpage, until some user-specified termination conditions apply. During this process the web-crawler will keep track of all the links between other websites and (possibly) follow them and retrieve those as well. However, off-the-shelf web-crawlers do not usually have analysis capabilities. Building off a custom-written web-crawler CENE [16, 17, 1, 18], the goal of this project was to extend CENE to collect geolocation and registrant information during the crawling process. The end result is a web-crawler called the Location Extraction of Child Exploitation Networks (LECEN). LECEN enables geographically identifying both the location of the website server, the server hosting the image (where a webpage links to an image hosted by an external host for example), and the contact information of the organization who registered the domain. This would allow law enforcement to prioritize child exploitation material (CEM) based on which websites are reachable either through contacting the hosting provider, or one of the people responsible for the domain name.

LECEN starts by downloading seed webpages from the Internet (Phase 1). For each webpage, all the images within that webpage are also downloaded, their hash value computed and compared against the RCMP-supplied child exploitation image-hash database (Phase 2). Concurrently with the image analysis, the content of the website is analyzed for a list of keywords known to be related to CEM content (Phase 3). If certain requirements are met by the webpage, then a Whois service is queried for the domain registrant information (Phase 4), and a geolocation database is queried to determine the geographical longitude/latitude of the IP address associated to the domain (Phase 5). Once this process is done, all the results are stored in a database for later analysis (Phase 6). The specific components of Phases 1 through 6 are detailed in full below.

Phase 1 – CENE data collection

Seed websites are required in order to start the data collection (crawling) process. Those seed sites are downloaded, analyzed, and links from them are recursively followed out. The seed sites are usually found through a manual search for CEM using search engines. However, for this project no manual search for seeds was necessary since it was possible to use webpages containing CEM found through previous research [16, 17, 1, 18]. For each webpage, the source HTML was retrieved, the content analyzed, and if certain requirements were not met it was thrown away. If certain requirements were met (namely, that a confirmed CEM hash value was located, and/or a minimum of seven CEM relevant keywords were located), the page was considered 'on topic', stored in the database, and all links from it followed recursively.

Phase 2 – RCMP MD5 Image-Hash Database

A hash value is the result of a mathematical procedure whereby data is broken into a 24-hexidecimal code, and as a fingerprint**,** it is unlikely to be shared between files **[**17]. An MD5 hash value is simply a widely used cryptographic hash function. For each webpage, after the source HTML was retrieved, all images were retrieved and checked against the Royal Canadian Mounted Police's (RCMP) MD5 image-hash database. 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. 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.

Phase 3 – Keyword analysis

A collection of user-specified keywords were selected based upon their link to child pornography. In this case, 82 unique keywords and code words were selected based upon findings from previous research as relevant to CEM. To be captured by LECEN, each webpage contained a minimum of 7 unique keywords. Previously, seven was found to be the best threshold as to minimize false-positive and false-negative matches [18].

Phase 4 – Whois

The Internet's Whois service, originally referred to as Nicname, is a text-based query-response protocol which allows individuals to find out the registrant information on an internet domain [19]. This lookup allows IP addresses to be traced beyond the simple connection to the hosted site and provides details regarding the individual who owns an account linked to the domain in question. Such information adds an additional level of analysis which provides important implications to law enforcement as it allows another angle of attack: they might not be able to reach the service hosting the content but they may be able to reach the person owning the domain name. For this service *whoisxmlapi.com* was selected and integrated into LECEN because it was able to provide up-to-date Whois registrant information in an easily parsable XML format. *whoisxmlapi.com* derives their database directly from ICANN, implying that the information is both up to date and accurate.

The Whois information provides the registrant, administrative, and technical contact for the domain name. The registrant is the legal owner of the domain name. The administrative contact is the main contact regarding operation of the website. The technical contact maintains the website's operation and functionality. These contacts are often the same entity, but it is possible the registrant is a third-party company hired to provide these services, and could possibly be anonymous. Occasionally contact information may differ, which is important in jurisdictional considerations: three different contacts could yield three distinct geographical locations, which in turn provides three different jurisdictions where law enforcement can seek to intervene.

Category	Count	Domains	Definition
1	1,380	31	Child pornography
2	3,231	66	"Gray-area"
3	32,152	1,905	Adult pornography/Obscene materials/banners
Unknown	142,065	10,429	Indeterminate

Table 1 - Count of images found during this phase of the project

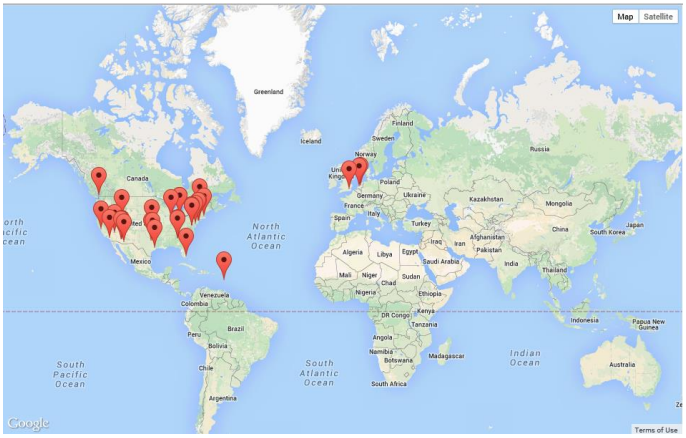

Figure 1 - Geolocation map of child exploitation material

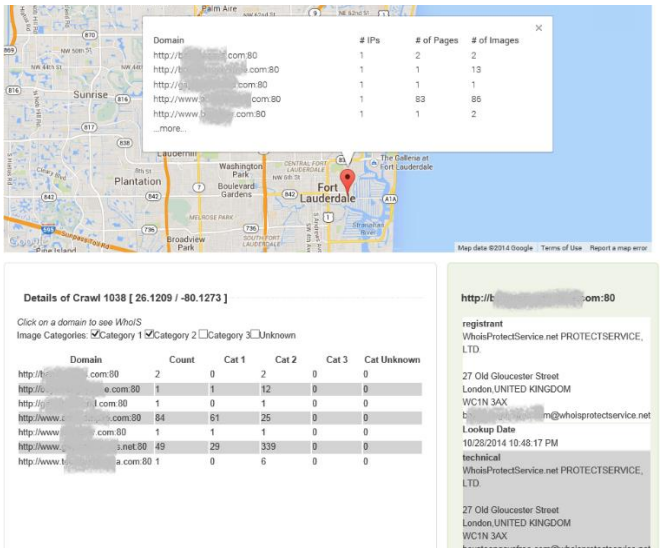

Figure 2 - Geolocation map with Whois information

Figure 3 - Image density

Phase 5 – Geolocation

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 [21]. The end 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.

MaxMind's GeoLite [20] is an offline database that is accurate, current and cost effective, and hence was integrated into LECEN. During the crawling process it allows LECEN to relatively accurately identify the location of any IPv4 address, as well as some IPv6 addresses. Its accuracy varies depending on the granularity level: 99.8% for Country level, 90% for State levels and 83% for city level [20]. Since CEM is a federal-level crime, the primary goal is identifying the correct country, which GeoLite supports with 99.8% accuracy. In the end, for each IP address encountered, a lookup is made to this database, and a longitude/latitude is associated to the IP address.

Phase 6 – Storage

Assuming that the webpage was retrieved successfully with all images associated to it, and at least one image was a Category 1 or 2 CEM image, or there were at least seven keywords on the webpage, the Whois and geolocation information is retrieved for the domain where the webpage is. All this information is then stored in a central database for later analysis. The images themselves are not stored, just their hash values.

III. Results

45,794 webpages containing CEM were discovered by CENE since 2010 [16, 17, 1, 18]. These webpages were used as the seed pages to start the crawling process in LECEN. Data capture started October 28th, 2014 and was stopped after 8 days. This yielded 207,957 webpages which met the requirements (either contained a CEM image, or at least 7 keywords), and 836,918 webpages were retrieved which did not meet the requirements. All webpages meeting the requirements were geolocated and Whois information for them retrieved. Overall,

Domain Country	Total Category 1 images		Total Category 2 images		Domains Hosting Category-1 images		Domains Hosting Category-2 images	
USA	1312	(95.1%)	2,708	(83.8%)	43	(74.1%)	74	(73.0%)
Netherlands	63	(4.6%)	487	(15.1%)	11	(19.0%)	16	(21.6%)
United Kingdom	3	(0.2%)	30	(0.9%)	2	(3.4%)	3	(4.1%)
Canada	1	(0.1%)	6	(0.2%)	1	(1.7%)	1	(1.4%)
Dominica	1	(0.1%)	0	(0.0%)	1	(1.7%)	0	(0.0%)
Grand Total	**1,380**	**(100.0%)**	**3,231**	**(100.0%)**	**58**	**(100.0%)**	**74**	**(100.0%)**

Table 2 - Total number of images by domain country

this crawling process located the presence of 1,380 unique Category 1 child exploitation images and 3,231 Category 2 'gray-area' images (Table 1).

Geolocation

The web-interface in Figure 1 shows a map which is produced from the results of the crawls. Each red marker represents the set of IP addresses hosting a CEM image which have been identified to that geo-coordinate. Multiple IP addresses which geolocate to the same longitude/latitude only show as one pin. By clicking on any pin, the list of domains which are geolocated at that exact same coordinate are displayed, along with statistics about each domain name, such as the number of Category 1, 2 and 3 images found and number of webpages analyzed on that specific domain. As can be seen in Figure 1, all but 6 of the geographical locations identified as hosting CEM were within the United States, with most of those clustered near the major city centers, presumably where major data-hosting centers are located.

For each domain, additional Whois information is displayed, such as the registrant, administrative, and technical contact details (see Figure 2). Unfortunately, there are a number of limitations with this information. For instance, although a website may provide specific administrative and registrant details, in many cases real names and addresses of the registrant are obfuscated behind anonymous protection services (as in the example shown on Figure 2). Even when contact details for specific persons are displayed, these may or may not be fictional accounts created for the purposes of hosting, given the extreme content which may be hosted.

Whereas in Figure 1 or Figure 2, the images are *aggregated* to the server level, Figure 3 clusters the distribution of the *individual* images by their location. In this case if a single server has 100 images, it will show as 100 in the map. Zooming out, high-level clusters are shown, which are broken apart to smaller clusters as the user zooms in. All the maps allow for zooming in order to provide more details.

Breaking down the geolocation information by country where the *images* are *hosted* (Table 2), it was found that the large majority of Category 1 images identified by LECEN were being hosted within the United States of America (95.1%). The Netherlands came in at a distant second (with 4.6%), but was considerably higher than all other subsequent countries.

Table 2 also shows the geolocation information broken down by the total number of *domains* by country of origin hosting both Category 1 and 2 imagery. In this case, the United States was yet again the primary offender, with it accounting for 43 of the 58 (74.1%) domains hosting Category 1 imagery and 54 of 74 (73.0%) domains hosting category 2 images in the United States.

Whois

The vast majority of images and domains are in jurisdictions (US & UK) with purportedly strict laws against CEM, meaning the content itself should be reachable by law enforcement. However, new hosts can be brought up behind the domain name, and the content replaced. Since a domain name registrant is considered the legal owner of a domain name registration, considering the presence of confirmed CEM material legally hosted by them, this information is of particular interest within this research, and to interested end-users such as law enforcement.

Through a Whois registrant lookup, contact details for the owners of the domain were collected. Table 3 shows a breakdown of the countries where the registrants were identified as being located (through their Whois contact details). In this case the large majority of *registrants* who have a domain name hosting Category 1 CEM images were located in both the United Kingdom (716 category 1 images) and the USA (414 category 1 images). This trend held true for the presence of Category 2 images in both the United Kingdom (801) and the United States (1,042). Again, it must be considered that many of the personal details (such as name or street-address) attributed to registrants, administrators, and technical contacts were obfuscated behind protection services. In these instances, the street-address of the protection service itself was used.

The domain level information only displays the involvement of four countries within this network (Table 2) as compared to the inclusion of registrant information (Table 3) which expands the list to nineteen. Comparing the domain level count by country (Table 2 – latter two columns) and the Category 1 & 2 Image counts by country (Table 2 – first two columns) with the registrant information (Table 3) reveals a previously concealed picture. Looking at only the domain information shows a United States centered network containing 95.1% Category 1 images, 83.8% of the Category 2 images, 74.1% of the domains hosting Category 1 images, and 73% of the domains hosting Category 2 images. Including the registrant data the United Kingdom and France emerge as key countries within this network. Registrants from France owning particular domains actually contain the largest percentage of the Category 2 images within this network at 35.7%. This is an important since France does not even appear on the domain level information by image or domain count. The United Kingdom (Table 2) appears to play a minor role in this network with just 0.2% of the Category 1 images but in reality the registrants from the UK have the majority of the Category 1 images at 51.9%. These large discrepancies show the value of

Registrant Country	Number of Images				Number of Domains			
	Cat-1 Images		Cat-2 Images		Cat-1 Images		Cat-2 Images	
United Kingdom	716	(51.9%)	801	(24.8%)	14	(24.1%)	15	(20.3%)
USA	414	(30.0%)	1042	(32.3%)	18	(31.0%)	25	(33.8%)
Australia	83	(6.0%)	31	(1.0%)	8	(13.8%)	4	(5.4%)
Panama	62	(4.5%)	35	(1.1%)	2	(3.4%)	5	(6.8%)
Argentina	58	(4.2%)	56	(1.7%)	1	(1.7%)	1	(1.4%)
Belize	19	(1.4%)	17	(0.5%)	1	(1.7%)	2	(2.7%)
France	14	(1.0%)	1154	(35.7%)	5	(8.6%)	5	(6.8%)
Ukraine	6	(0.4%)	5	(0.2%)	1	(1.7%)	1	(1.4%)
Estonia	2	(0.1%)	10	(0.3%)	1	(1.7%)	1	(1.4%)
Austria	1	(0.1%)	25	(0.8%)	1	(1.7%)	4	(5.4%)
Mexico	1	(0.1%)	10	(0.3%)	1	(1.7%)	1	(1.4%)
Cyprus	1	(0.1%)	8	(0.2%)	1	(1.7%)	3	(4.1%)
Federation of Saint Christopher and Nevis	1	(0.1%)	6	(0.2%)	1	(1.7%)	1	(1.4%)
Italy	1	(0.1%)	1	(0.0%)	1	(1.7%)	1	(1.4%)
Sweden	1	(0.1%)	0	(0.0%)	1	(1.7%)	0	(0.0%)
Israel	0	(0.0%)	15	(0.5%)	1	(1.7%)	1	(1.4%)
Luxembourg	0	(0.0%)	9	(0.3%)	0	(0.0%)	1	(1.4%)
Netherlands	0	(0.0%)	5	(0.2%)	0	(0.0%)	2	(2.7%)
Czech Republic	0	(0.0%)	1	(0.0%)	0	(0.0%)	1	(1.4%)
Grand Total	**1,380**	**(100.0%)**	**3231**	**(100.0%)**	**58**	**(100.0%)**	**74**	**100.0%**

Table 3 - Registrant country broken down by number of images and number of domains

Site	Boy	Teen	Twink	Gallery	Cum	Sex	Teen	Anal	Girl	Cock
A	37,215	27,508	3,637	2,386	2,150					
B	12,955	5,146	3,048	1,421		1,196				
C	33,190	20,001	4,921			5,303				3,300
D					1,516	2,221	1,176	882	588	
E	17,925	27,216	3,095		3,224	2,371				

Table 4 - Keyword count of top 5 Cat 1 Hosting Domains

Keyword	Count
Boy	427,235
Teen	233,196
Twink	134,199
Gallery	64,233
Sex	57,089

Table 5 - Top five keywords located

including the Whois data to support the domain level information.

One of the analyses undertaken was exploring the total number of keywords present within the websites extracted. Across all domains, the keyword *boy* was by far the most commonly appearing, more than any two other keywords combined. Furthermore, another "boy-centric" keyword *twink* was also very high in total counts, falling in third. These two results suggest that this set of sites could be focusing on boys.

One trend noticed was that the top 5 Category 1 domains were very focused on male content, given the large number of times the words *boy* and *twink* appeared (Table 4). This lends the observation that this online network is focused on more male, and less female content. For the top three domains, their results largely mirrored those in the overall top five keyword counts. Only one of the top five sites (identified as *Site D* in Table 4) in the network bucked this trend, having a large presence of keywords such as *girl* and *anal* compared to the other domains.

IV. Conclusions

Geolocation and Whois data provide valuable information regarding website ownership and domain location. Utilizing these services, this paper sought to expand the avenues available to law enforcement and third party organizations in attack or disruption strategies. Previous work using CENE [16, 17, 1, 18] centered upon Social Network Analysis and how it could aid in disrupting these networks more effectively. By expanding CENE into LECEN, traditional methods of manually examining location and registrant data are vastly improved. Upon finding a CEM website previously, manual extraction of information was required which is a time consuming process. Incorporating this functionality into the existing software allows law enforcement to focus time and effort into solving issues of jurisdiction and effective disruption.

Different jurisdictional and legal principles surrounding cyberspace continue to play a role in how law enforcement can police and enforce traditional concepts of physicality related to crimes. Identifying registrants and the locations of servers increases the possibility for law enforcement to attack these networks from multiple angles. A website may be hosted in the United States, registered to a person in the United Kingdom and maintained by a company in Canada which allows for all three countries to be involved in eliminating these CEM websites.

The results indicate a large majority of Child Exploitation content is hosted by, and registered to, servers and individuals within the United States and the United Kingdom. With the inclusion of Geolocation and Whois data the examination of previously unavailable variables such as jurisdictional and legal aspects of CE materials becomes possible. Comparing the United States and the United Kingdom, which both have strict laws governing CE content to other countries on the list with more relaxed legal systems highlights the ongoing struggles law

enforcement face when prosecuting these crimes. Even with a more proactive policing and punitive legal system these two countries host 81.9% of the Category 1 images found by registrant. This implicates another serious issue where 51.9% of the registered domains hosting category 1 images are from the UK but only 3.4% of the domains hosting category 1 images are located within the United Kingdom's borders. Possibly to avoid detection or to take advantage of the jurisdictional issues this suggests that users may be incorporating layering indicating a higher level of technological sophistication into the hosting of these child exploitation sites.

V. LIMITATIONS

In order to locate new domains hosting CEM, our crawler relied on previously identified seed sites hosting such material. Despite the need for seed sites in finding new domains, it may be true that the network extracted and analyzed may be skewed based on the nature of the selected seed sites. For example, if these sites dealt primarily with male-centric material, it would bias such results as being over-representative of male CEM. It is possible that manual verification of the nature of these seed sites and content contained within might provide a less skewed network.

VI. FUTURE RESEARCH

From this study two more are proposed using the LECEN software. The first is to exhaustively crawl all the webpages within the domains containing Category 1 and 2 images. Instead of having LECEN follow all links outside of each site, each domain would be examined in full to derive the proportionality of CE materials on each site. This would allow for a more comprehensive understanding of the information being hosted on these particular domains and how large each website truly is. The second proposed research study would utilize the domains located in this current research study as seed sites, in conjunction with an updated keyword list in order to expand our research criteria and potential results. Finally, there were a number of unknown images and videos located on domains hosting category 1 and 2 images. It is likely some of these unknown hash values are new or previously unidentified CEM. It may be possible to use LECEN to make basic predictions regarding websites hosting new CEM, allowing law enforcement to make informed decisions regarding key players within a network, and thus impact investigations both online and offline.

VII. 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.

VIII. 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) Stanley, J. (2001). Child abuse and the internet. *Chil.. Abu. Prev. Iss., vol. 15*, pp. 1-20.

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

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

5) Shipley, T.G., & Bowker, A. (2014). Investigating internet crimes: An Introduction to solving crimes in cyberspace (pp. 1- 412). Steven Elliot. doi: 10.1016/B978-0-12-407817-8.00020-5

6) Wolak, J., Liberatore, M., & Levine, B. N. (2014). Measuring a year of child pornography trafficking by us computers on a peer-to-peer network. *Child Abuse & Neglect*, 38, 347-365. http://dx.doi.org/10.1016/j.chiabu.2013.10.018

7) 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

8) 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

9) Zheng, Q., Zeng, W., Wen, G., Wang, W. (2006). Shape-based adult image detection. *International Journal of Image and Graphics*, 6(1), 115-124. doi: 10.1142/S0219467806002082

10) Ho, W. H., & Watters, P. A. (2004, October). Statistical and structural approaches to filtering internet pornography. *In Systems, Man and Cybernetics, 2004 IEEE International* Conference (Vol. 5, pp. 4792-4798). IEEE. doi: 10.1109/ICSMC.2004.1401289

11) Forsyth, D. A., & Fleck, M. M. (1996, December). Identifying nude pictures. In Applications of Computer Vision, 1996. WACV'96., Proceedings 3rd IEEE Workshop on (pp. 103-108). IEEE. doi: 10.1109/ACV.1996.572010

12) Lopes, A. P. B., de Avila, S. E. F., Peixoto, A. N., Oliveira, R. S., de M Coelho, M., & de A Araujo, A. (2009b). Nude detection in video using bag-of-visual-features. In Computer Graphics and Image Processing (SIBGRAPI), 2009 XXII Brazilian Symposium on (pp. 224-231). *IEEE*. doi: 10.1109/SIBGRAPI.2009.32

13) Sivic, J., & Zisserman, A. (2003). Video Google: A text retrieval approach to object matching in videos. In Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on (pp. 1-8). IEEE. doi: 10.1109/ICCV.2003.1238663

14) Lopes, A. P., de Avila, S. E., Peixoto, A. N., Oliveira, R. S., & Araújo, A. D. A. (2009a). A bag-of-features approach based on hue-SIFT descriptor for nude detection. In Proceedings of the 17th European Signal Processing Conference, Glasgow, Scotland (pp. 1552-1556).

15) Deselaers, T., Pimenidis, L., & Ney, H. (2008). Bag-of-visual-words models for adult image classification and filtering. In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on (pp. 1-4). IEEE. doi: 10.1109/ICPR.2008.4761366

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

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

18) 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.

19) Sullivan, A., & Kucherawy, M. S. (2012). Revisiting Whois: Coming to REST. *IEEE internet computing*, 16(3), 65-69.

20) Mueller, M., & Chango, M. (2008). Disrupting global governance: the Internet whois service, ICANN, and privacy. *Journal of Information Technology & Politics*, 5(3), 303-325.

20) MaxMind GeoLite (2014). Accessed on August 15, 2014 through https://www.maxmind.com/en/geoip2-databases

21) Poese, I., Uhlig, S., Kaafar, M. A., Donnet, B., & Gueye, B. (2011). IP geolocation databases: Unreliable?. *ACM SIGCOMM Computer Communication Review*, 41(2), 53-56.