Child Abuse & Neglect xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Child Abuse & Neglect

Measuring a year of child pornography trafficking by U.S. computers on a peer-to-peer network☆

Janis Wolak[a,*], Marc Liberatore[b], Brian Neil Levine[b]

[a] Crimes against Children Research Center, University of New Hampshire, USA
[b] School of Computer Science, University of Massachusetts, Amherst, USA

ARTICLE INFO

Article history:
Received 22 July 2013
Received in revised form 16 October 2013
Accepted 24 October 2013
Available online xxx

Keywords:
Child pornography
Peer-to-peer
Internet
Child sexual exploitation

ABSTRACT

We used data gathered via investigative "RoundUp" software to measure a year of online child pornography (CP) trafficking activity by U.S. computers on the Gnutella peer-to-peer network. The data include millions of observations of Internet Protocol addresses sharing known CP files, identified as such in previous law enforcement investigations. We found that 244,920 U.S. computers shared 120,418 unique known CP files on Gnutella during the study year. More than 80% of these computers shared fewer than 10 such files during the study year or shared files for fewer than 10 days. However, less than 1% of computers ($n = 915$) made high annual contributions to the number of known CP files available on the network (100 or more files). If law enforcement arrested the operators of these high-contribution computers and took their files offline, the number of distinct known CP files available in the P2P network could be reduced by as much as 30%. Our findings indicate widespread low level CP trafficking by U.S. computers in one peer-to-peer network, while a small percentage of computers made high contributions to the problem. However, our measures were not comprehensive and should be considered lower bounds estimates. Nonetheless, our findings show that data can be systematically gathered and analyzed to develop an empirical grasp of the scope and characteristics of CP trafficking on peer-to-peer networks. Such measurements can be used to combat the problem. Further, investigative software tools can be used strategically to help law enforcement prioritize investigations.

Introduction

Online child pornography (CP) trafficking has become a serious crime problem in the United States and worldwide, fostered by the development of online and digital technologies (Beech, Elliott, Birgden, & Findlater, 2008; Jenkins, 2001; Wolak, Finkelhor, & Mitchell, 2011; Wortley & Smallbone, 2012). We use the term *trafficking* because it denotes an illegal trade, which accurately describes the online trade in child pornography. Photographs and videos that contain child pornography are contraband because they show actual children being sexually abused and exploited (Wolak, Finkelhor, & Mitchell, 2005a). The crimes that constitute online CP trafficking (i.e., using the Internet to distribute or acquire CP) are unusual because they are a form of child sexual exploitation that involves no direct interaction with a victim. However, CP trafficking is not a victimless crime; the children and adolescents pictured in the images are victims of exploitation and often abuse by the

☆ Janis Wolak was supported by Grant No. CNS-1016788 awarded by the National Science Foundation; Marc Liberatore and Brian N. Levine were supported in part by CNS-1018615 awarded by the National Science Foundation and in part by 2008-CE-CX-K005 awarded by the U.S. Department of Justice. The supporting agencies had no role in the study design, data collection or analysis, preparation of the manuscript or decision to publish. We are grateful to the law enforcement partners who assisted in this project.

* Corresponding author.

0145-2134/$ – see front matter
http://dx.doi.org/10.1016/j.chiabu.2013.10.018

producers of the images (Wolak, Finkelhor, & Mitchell, 2005b; Wolak, Finkelhor, Mitchell, & Jones, 2011). In addition, victims who realize their images are trafficked online may suffer from knowing their images are viewed for sexual purposes (Bazelon, 2013; Svedin & Back, 1996; Weiler, Haardt-Becker, & Schulte, 2010).

In this paper, we describe our progress, joint with law enforcement, in investigating and measuring online CP trafficking in peer-to-peer (P2P) file sharing networks. P2P networks are said to be major online locations of the illegal trade in CP (Koontz, 2005; U.S. Department of Justice, 2010). Increasing proportions of U.S. arrests for CP possession and distribution involve offenders who used P2P networks to acquire CP, from 4% of all such arrests in 2000 to 61% in 2009 (Wolak, Finkelhor, & Mitchell, 2012). Although any online venue that can be used to transmit or post photographs or videos can be used to distribute or acquire CP, P2P networks make especially large contributions to the problem because of their worldwide range, public nature, and the easy access they provide to child pornography (Latapy, Magnien, & Fournier, 2013; Steel, 2009).

Research on the extent to which CP trafficking occurs on P2P networks is limited, however, and systematic measurements of the size of the problem, which could help law enforcement and policy makers understand its scope and characteristics, are scarce. There has been some research about the amount of CP trafficking in P2P networks, but most studies have been narrow in scale. For example, several researchers have documented the presence of CP on P2P networks (Hughes, Walkerdine, Coulson, & Gibson, 2006; Latapy et al., 2013; Prichard, Watters, & Spiranovic, 2011) or analyzed searches by P2P users to determine the percentage that appear to be requests for CP (Rutgaizer, Shavitt, Vertman, & Zilberman, 2012; Steel, 2009).

In contrast, our work paints a broader and more detailed picture by characterizing online activities by computers trafficking in CP on a P2P network over a year's time. Two previous papers published for the computer science field described worldwide patterns of CP trafficking on the Gnutella and eMule P2P networks (Hurley et al., 2013; Liberatore, Levine, & Shields, 2010). The present study is aimed at informing researchers, practitioners, law enforcement, and policy makers in the child maltreatment field about how CP trafficking in P2P networks can be measured and how such measurement can lead to strategic responses to combat and possibly even reduce the amount of CP available online. We measure CP trafficking activity by U.S. computers located on one P2P network, Gnutella, and provide perspective about how U.S. activity contributes to CP trafficking worldwide.

P2P File Sharing Networks

P2P file sharing networks are vast global systems used by millions of people to acquire, for free, popular music, current television shows, movies, electronic books, and other digital material from network users who are willing to share items in their possession. Media attention has focused on unauthorized sharing of copyrighted music and video files, but pornography is also widely available. This includes both legal images featuring adults and child pornography.

In general, individuals become P2P users by downloading software that connects them to the computers of other users in a network (e.g., Gnutella, BitTorrent, Ares). These other users could be located anywhere in the world. The software allows users to log onto the P2P network and issue requests for and download files from other network users, called *peers*. Users create shared folders that are accessible to others in the network and use these folders to receive downloaded files and also to share files they possess. Procedures vary somewhat among networks, but in most, users search for electronic files by using keywords, which are broadcast to the network of participating peers. Certain keywords are specifically associated with CP, but network users do not have to know these to obtain CP. Users can locate and acquire files by employing search terms that, for example, describe sex acts and children's ages, which are often contained in CP file titles and tags (Steel, 2009).

When a search finds a relevant file, the network generates an automated response that identifies the network location of the computer with the file and information about the file (e.g., size, name). If the user requests to download the file, the file is transferred. Searches may locate duplicates of requested files in shared folders from multiple computers, and downloads can be made simultaneously from such multiple sources. When this happens, each source contributes a portion of the file, in which case the file can be downloaded more quickly.

There are several reasons P2P networks may be particularly attractive to CP traffickers. First, CP on P2P networks is free and publically accessible. Any person with access to the Internet can connect to a P2P network, and any P2P user who wants to obtain CP can do so. Second, P2P networks do not make use of servers, thus, users can transmit illegal material without oversight from electronic service providers. Third, P2P networks may be more anonymous than other means used to acquire or distribute CP online, most of which require more direct contact with others. For instance, to transmit CP via email or text message, senders must direct communications to specific addresses. To access CP on websites, image boards, social networking, and similar sites, persons generally must know the CP distributor, join an online group, or find and pay for access to an illegal website, which may require credit card or other information to allow for payment. In contrast, CP traffickers who use P2P networks do not have to risk personal contact with other individuals or reveal their identities. To acquire CP, they simply search for material and download what they find into a shared folder. To distribute CP, they simply upload files into a shared folder, which allows others to find the material when they search for it. Given that P2P networks have millions of users and CP can be acquired and distributed as easily as other content, CP traffickers who consider the risk of being singled out and pursued by law enforcement may feel it is quite low.

This easy access to child pornography on P2P networks may create distinct harms. Network users who are curious about CP can easily satisfy their curiosity, and finding it so easily may make viewing CP seem normal and acceptable (Quayle & Taylor, 2002). Individuals who might not have become CP traffickers may do so after encountering the material in P2P networks. Easy access to CP in P2P networks also may foster the proliferation of CP. Every time a CP file is downloaded, a new copy of the image is created. If left in shared folders, these new files add to the quantity of CP that is available on the network. Much of the rampant online circulation of CP is due to its transmission in P2P networks, because files are duplicated continuously as they are downloaded and shared among network users. Finally, as noted above, easy access to CP on P2P networks may increase the harm to victims of child pornography production when they have to live with the knowledge that their images are publically available and easily accessible to P2P users.

The Law Enforcement Response

The ways that P2P networks facilitate online CP trafficking have created a number of challenges for law enforcement and policy makers seeking effective ways to combat the proliferation of CP. These challenges include how to develop responses to CP trafficking on P2P networks, target the most egregious offenders, understand and measure the impact of policing efforts and, ultimately, reduce CP trafficking and the number of illegal images and videos of children that are circulating online.

On the other hand, the public nature of P2P networks has given law enforcement unique opportunities to proactively investigate CP trafficking. Law enforcement and computer scientists have developed sophisticated systems known by names such as RoundUp, Gridcop, and Ephex which allow police to detect illegal activity by specific Internet Protocol (IP) addresses on a variety of P2P networks (Liberatore, Erdely, Kerle, Levine, & Shields, 2010; U.S. Department of Justice, 2010). However, using these systems is just a first investigative step. It requires significant resources to make a criminal case against a suspect when the case begins with only an IP address observed trafficking in CP. In most cases, investigators will search a public database to determine which electronic service provider hosted the IP address and subpoena the service provider to acquire the name and address of the account holder for the IP address. Investigators then search public records and criminal databases to learn more about the account holder, premises and possible other occupants; apply for and execute a search warrant for the computer and premises; and execute the warrant by searching and seizing implicated computers and related contraband. They also question suspects and witnesses to tie individuals to the CP trafficking and conduct forensic analyses of computers to provide prosecutors with evidence of trafficking that can be used in court.

Although time-consuming and resource-intensive, investigations of P2P networks serve important purposes.

Investigations and arrests may deter CP trafficking. Arrests remove trafficking computers and their files from a network, which may reduce the amount of CP that is circulating online. Further, an unknown proportion of CP traffickers are also child molesters (Seto, Hanson, & Babchishin, 2011). About 10% of investigations of online CP trafficking that end in arrest detect child molesters who might not otherwise have been caught (Wolak et al., 2012).

A major question for law enforcement agencies is how to effectively prioritize investigations of CP trafficking on P2P networks, given the presumably large numbers of U.S. computers that are trafficking in CP and the sizeable effort required and limited resources available to investigate individual cases. One stated law enforcement goal is to target CP traffickers who also directly molest children (U.S. Department of Justice, 2010). However, at this time there is no way that we know of to single out such offenders based solely on how they behave on P2P file sharing networks. Reducing the number of CP files available on P2P networks may be a more realistic goal for the time being.

The Present Study

First, we describe a software tool called *RoundUp*, which is used for proactive investigations of CP trafficking in P2P networks. Second, we use data gathered through the use of the RoundUp tool to measure and characterize a year of CP trafficking activity on the Gnutella P2P file sharing network. Our findings are focused on computers located in the United States but provide perspective about how CP trafficking from U.S. computers contributes to the worldwide availability of CP files. Third, we use RoundUp's measurement capabilities to show how U.S. law enforcement might prioritize investigations to target high-contribution CP traffickers (i.e., those that contribute the most to the number of CP files available in a P2P network). Finally, we discuss the limitations of our data and make recommendations to improve data gathering efforts. Our data have significant limitations, but our findings show how progress in measurement is occurring and that it is possible to develop systematic measurements that can assist law enforcement and policy makers by providing an empirical grasp of the scope and characteristics of CP trafficking in P2P networks.

Method

The RoundUp Investigative Tool

The RoundUp software tool works by detecting CP in shared folders on a P2P network (Liberatore, Erdely et al., 2010). Such detection is authorized because courts have ruled that P2P users have no expectation of privacy regarding files in shared folders, which are publically accessible to other P2P users.

RoundUp investigative software was developed by computer scientists at the University of Massachusetts Amherst, including two of the authors of this paper, who worked in collaboration with law enforcement investigators. The U.S. Department of Justice funded the development of RoundUp. Its nationwide use is managed by a law enforcement agency. The developers have been given access to data for research purposes, but they have no financial interest in the software.

The first version of RoundUp, used on the Gnutella P2P network, was launched in 2009. In 2013, several versions of RoundUp were in daily use by more than 2,000 U.S. law enforcement investigators observing the Gnutella and other P2P networks. Since 2009, RoundUp investigations have resulted in over 4,000 arrests for CP possession and distribution.

In the Gnutella network, RoundUp works by detecting *known CP files* that have been identified as CP in previous law enforcement investigations. RoundUp software recognizes known CP files based on cryptographic hash algorithms, or *hash values*, which are unique numeric identifiers generated by computer algorithms based on the content of digital files. Duplicate files will have the same hash value but a file that is even slightly different, for example a photograph that has been cropped or a video in which one frame has been edited, will have a different hash value. P2P network software uses hash values because users frequently rename or edit electronic files that are shared. Hash values distinguish when files are identical despite such changes. The RoundUp system does not include actual CP photographs or videos, only the hash values associated with such files.

During the study year, the RoundUp system was programmed to recognize the hash values of approximately 384,000 known CP files, both photographs and videos. This was a convenience sample of known CP files that law enforcement personnel who managed the system gathered from various agencies. It included files from a variety of P2P networks, Web sites, and collections found on the computers of arrested suspects. Although RoundUp software searches for known CP files, investigations often identify previously unknown CP files that are found in suspects' possession, for example in shared network folders and on computers. When new CP files are discovered, their hash values are added to the list of known CP files. Because law enforcement personnel updated the list of known CP files during the study year, known files could be recognized only after their hash values were entered into the system. The system managers did not keep detailed records when hash values were uploaded, so we did not have exact information about how many known CP files were uploaded into RoundUp and when.

Once RoundUp recognizes that a known CP file is being shared, investigators can determine if the activity is within their jurisdiction via information supplied through a commercial geo-location database that provides town, state, and country based on IP address. The system also provides investigators with the dates and times a target has been observed online sharing known CP files and the hash values of other known CP files shared. RoundUp stores data describing the IP addresses sharing CP, dates and times of observations, numbers and hash values of known CP files in shared folders, and changes over time in these observations. RoundUp data include online activities only; there is no information about the characteristics of individual CP traffickers operating the observed computers. Also, no information about the content of known CP files was available for this paper.

Sample

Data for this paper include all RoundUp observations of known CP files shared on the Gnutella P2P network from October 1, 2010, to September 18, 2011. RoundUp was used almost continuously by law enforcement during that time and recorded over 870 million observations of known CP files being shared by computers in more than 100 countries. Each observation is a record of one computer sharing one known CP file in a shared folder at a particular time on the Gnutella P2P network.

Variables and Measures

Computer. We measured CP trafficking activity for individual computers that were connected to the Gnutella network during the study year. We are reasonably confident that we distinguished individual computers, based on IP addresses, other numeric identifiers, and specific files that were shared (Hurley et al., 2013). However, individual computers do not necessarily correspond to individual persons; we had no way of knowing how often one individual operated multiple computers.

Known CP file. A known CP file is a photograph or video associated with a specific hash value which was identified by law enforcement as child pornography in a previous investigation and was recognizable by the RoundUp system.

Duplicate known CP file. Duplicate files are identical copies of known CP files, which have the same hash values.

Sharing a known CP file. A computer shared a known CP file when RoundUp software observed the file in a shared folder on the network. Each shared folder was associated with a specific computer. Known CP files were in shared folders when a user either uploaded them from their computer into the shared folder or downloaded them from the network into the shared folder. Data are limited to known CP files in shared folders. We did not have information about CP that might have been stored on a computer's hard drive or other locations.

File availability. A known CP file was available on a given day if at least one computer was sharing it on that day. The more days that a file was shared the greater the file availability. Generally, duplicate files shared by multiple computers had greater availability than files shared by only one computer.

Annual contribution to file availability. We created a measure of each U.S. computer's contribution to file availability on the Gnutella network during the study year, taking into account when multiple computers shared duplicate known CP files. We focused on U.S. computers because we were interested in how RoundUp data could inform U.S. law enforcement activity. First, we counted the number of known CP files shared by each U.S. computer per day throughout the study period. Second, we divided each known CP file shared on a given day by the number of its duplicates shared on that day worldwide. For example, if 100 computers shared duplicates of a known CP file on a given day, each computer contributed one-hundredth of that file to the availability of that file on that day. If two computers shared duplicates of a known CP file, each contributed

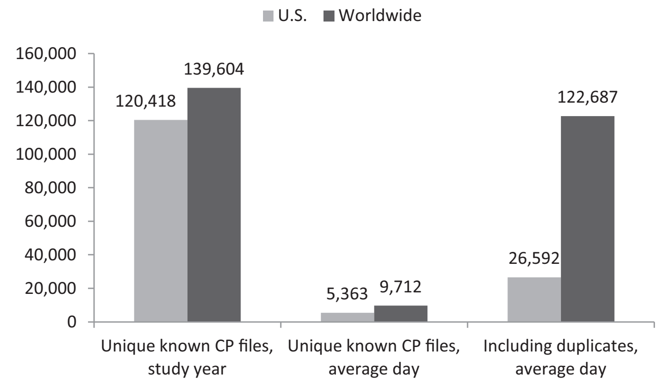

Fig. 1. Number of known CP files shared on the Gnutella P2P network from U.S. computers and worldwide during the study year. *Abbreviations*: CP, child pornography; P2P, peer-to-peer; U.S., United States. *Note*: Duplicates are identical copies of known CP files.

one-half to the availability of that file on that day. We included duplicates shared worldwide because P2P networks do not differentiate among shared files based on geographical boundaries. Finally, we calculated each U.S. computer's annual contribution to file availability as the sum of its file contributions across the year of the study.

High-contribution computers. After we calculated the annual contribution to file availability for each U.S. computer, we created a subgroup of high-contribution computers. This consisted of the computers in the highest category, those that had an annual contribution to file availability of 100 or more known CP files.

Statistical Analyses

We created a database of RoundUp observations and used Structured Query Language (SQL) to organize and analyze the data. The IP addresses were anonymized and the data were stored in a Microsoft SQL Server. We used a standard SQL normalization procedure to organize the data (e.g., eliminate duplicative data, create subsets of related data in separate tables, establish relationships between tables). We then used SQL statements that we wrote for this purpose to conduct descriptive analyses of RoundUp observations (e.g., number of days sharing known CP files per computer, known CP files shared per computer, known CP files shared on average day, and duplicates of such files shared). We also calculated the potential impact of subtracting known CP files belonging to high-contribution U.S. computers from the total number of known CP files shared worldwide during the study year. Results of our analyses were loaded into Excel for calculations such as percentages and averages.

Results

Number of U.S. Computers Sharing Known CP Files

Police using the RoundUp system observed more than three-quarters of a million computers (N = 775,941) located in more than 100 different countries sharing known CP files on Gnutella during the study year. Of these computers, 32% (n = 244,920) were seen exclusively at locations in the United States.

Number of Known CP Files Being Shared

Of the approximately 384,000 known CP files that RoundUp software could recognize, RoundUp observed 139,604 (36%) being shared worldwide on the Gnutella P2P network during the study year (Fig. 1). On an average day, RoundUp observed 9,712 known CP files shared worldwide, about 7% of the total. However, this number did not include duplicates of known CP files, which were often available. Including duplicates, on an average day, RoundUp observed 122,687 known CP files shared worldwide on the Gnutella P2P network.

Narrowing observations to computers seen exclusively in the United States, RoundUp saw a somewhat smaller number of known CP files being shared during the study year, 120,418, or 86% of the worldwide total. On an average day, 5,363 known files were shared from U.S. computers, about 4% of the U.S. total. Counting duplicates, RoundUp observed 26,592 known CP files shared from U.S. computers on an average day.

Activity Levels of U.S. Computers Sharing CP

A large percentage of U.S. computers trafficking in known CP files on Gnutella shared relatively few files and did so infrequently, based on RoundUp observations (Table 1). We found that many shared known CP files for only one day (47%,

Table 1
Computers located exclusively in the U.S. that were observed on the Gnutella P2P network sharing known CP files during the study year.

U.S. computers sharing known CP files during the study year (*n* = 244,920)	% (*n*)
Days sharing known CP files	
1 day	47% (115,397)
2–9 days	41% (101,593)
10–99 days	11% (27,347)
100–353 days	<1% (583)
Known CP files shared during study year per computer	
1 known CP file	41% (99,821)
2–9 known CP files	42% (102,295)
10–99 known CP files	16% (39,622)
100 or more known CP files	1% (3,182)
Computers sharing one known CP file for one day	29% (70,286)
Number of computers sharing known CP on a daily basis during study year	
Lowest day	<1% (1,176)
Average day	1% (3,396)
Highest day	3% (6,955)

Abbreviations: CP, child pornography; P2P, peer-to-peer; U.S., United States.

n = 115,397) or shared only one known CP file during the study year (41%, *n* = 99,821). Twenty-nine percent shared one known CP file for one day (*n* = 70,286). On an average day, RoundUp observed a relatively small number of the U.S. computers that trafficked in CP during the study year with known CP files (1%, *n* = 3,396). Further, more than 80% of U.S. computers trafficking in CP were observed with fewer than 10 known CP files (83%, *n* = 202,116) or sharing such files for fewer than 10 days (88%, *n* = 216,990) during the study year.

There are several possible reasons for why we found so much low-level CP trafficking. Some U.S. computers may have been sharing other CP that was not among the 384,000 files that RoundUp software was configured to recognize. Some may have quickly removed downloaded known CP files from shared folders to avoid detection or for other reasons. Some computers may have had small numbers of known CP files that were downloaded unintentionally incidental to other activities. For example, a user may have searched for and requested legal pornography files featuring adults and received a CP file without realizing it. Some computers that shared a small number of known CP files on Gnutella may have trafficked more actively in CP on other P2P networks or through other sources. In other cases, computers may have simply shared small numbers of known CP files for short periods of time.

At the same time, relatively small numbers of U.S. computers trafficking in CP were more active on Gnutella. Seventeen percent of the U.S. computers that RoundUp detected with known CP files shared 10 or more such files (*n* = 42,804) during the study year. One percent shared 100 or more known CP files (*n* = 3,182) and 0.04% (*n* = 100) shared 1,000 or more. RoundUp observed 11% of the U.S. computers that were trafficking in CP sharing known CP files for 10 or more days (*n* = 27,930) and less than 1% for 100 or more days (*n* = 583).

Similarly, although U.S. computers trafficking in CP shared 120,418 known CP files in total during the study year, most of these files were shared by relatively few U.S. computers and were shared on Gnutella for relatively few days (Table 2). For example, almost half of known CP files (45%, *n* = 54,640) were shared from only one U.S. computer. In other words, no other U.S. computers shared duplicates of those files. Most known CP files (79%, *n* = 95,017) were shared from any U.S. computer for fewer than 10 days, including 25% (*n* = 30,519) that were shared for only one day. Of the known CP files that RoundUp

Table 2
Known CP files shared on the Gnutella P2P network by U.S. computers during the study year.

Known CP files available during study year from US computers (*n* = 120,418)	% (*n*)
Number of known CP files shared by	
1 computer	45% (54,640)
2–9 computers	43% (52,429)
10–99 computers	9% (10,561)
100–999 computers	2% (2,344)
1,000 or more computers	<1% (444)
Number of known CP files available for	
1 day	25% (30,519)
2–9 days	54% (64,498)
10–99 days	17% (20,670)
100–353 days	4% (4,731)
Number of known CP files available on a daily basis during study year	
Lowest day	2% (2,159)
Average day	4% (5,363)
Highest day[a]	24% (29,386)

Abbreviations: CP, child pornography; P2P, peer-to-peer; U.S., United States.
[a] The daily high of 29,286 was due to one computer that was sharing more than 29,000 known CP files.

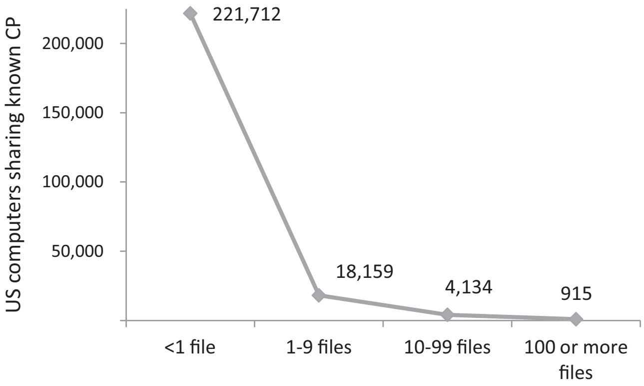

Fig. 2. Annual contributions of U.S. computers to the availability of known CP files on the Gnutella P2P network during study year. *Abbreviations*: CP, child pornography; P2P, peer-to-peer; U.S., United States.

observed being shared by U.S. computers, only 4% (n = 4,731) were available for 100 or more days. Files that were available for longer periods tended to have many duplicates, which were shared over multiple days.

Small Contributions to File Availability by Most U.S. Computers

We measured the annual contribution to the availability of known CP files by U.S. computers, which is the extent to which U.S. computers contributed to the number of known CP files available worldwide on Gnutella during the study year. Our measure took into account the number of duplicates of known CP files. We did this because if one computer is sharing files that have many identical copies across the network, removing that one computer from the network will have no impact on the number of known CP files available because so many duplicates exist. However, other known CP files are relatively rare; they have few or no duplicates. Removing these files from the network will do more to reduce the overall number of known CP files available than removing common files.

We found that the great majority of U.S. computers that trafficked in known CP files (91%, n = 221,712) contributed less than one known CP file to the number of known CP files available worldwide on Gnutella during the study year (Fig. 2). These computers only shared highly duplicated files that were widely shared by other computers. Removing all of the known CP files shared by these computers during the study year would have virtually no impact on the number of known CP files available worldwide on Gnutella. Only a small number of U.S. computers (<1%), had annual contributions of 100 or more known CP files (n = 915) to file availability on Gnutella. Among these high-contribution computers, contributions ranged from 100 to more than 44,000 known CP files, with a median annual contribution to file availability of 471 files.

We tested what effect removing the files shared by the high-contribution U.S. computers would have on the number of known CP files available worldwide on the Gnutella network during the study year. We performed this test to examine whether a law enforcement strategy of targeting high-contribution U.S. computers might be an effective way to reduce the number of known CP files available in a P2P network.

When we subtracted the files shared by the high-contribution computers from our dataset, it reduced the number of known CP files available worldwide on Gnutella by 30%, from 139,604 files to 97,691. We also examined the impact of removing the known CP files contributed by *all* U.S. computers that RoundUp observed trafficking in CP (n = 244,920). When we did this, the number of known CP files worldwide was reduced by 32%, from 139,604 known CP files to 95,305. This amount was just 2% greater than the 30% of files which could be removed by targeting the 915 high-contribution U.S. computers. In other words, the high-contribution computers, which comprised about 0.4% of all U.S. computers that were trafficking in CP on Gnutella, were responsible for almost all of the U.S. annual file contribution to the availability of distinct file content on the network. This imbalance suggests that a law enforcement strategy that targets high-contribution CP traffickers has potential for reducing the number of known CP files shared in P2P networks.

Limitations

Our findings probably underestimate CP trafficking on the Gnutella network during the study year for several reasons. First, the list of known CP files that RoundUp software used to recognize child pornography was not comprehensive. It did not include all such files previously identified by law enforcement, and it did not include CP files that were unknown to police. Second, the entire list of known CP files was not in the RoundUp system for the entire study year. For both of these reasons, computers could have shared CP files that were not recognized by the RoundUp system and thus not included in our measurements. Third, our findings may underestimate CP trafficking in known CP files worldwide and overestimate the U.S. annual file contribution because the RoundUp list of known CP files was compiled solely from U.S. sources. Fourth, the observations of CP trafficking were gathered pursuant to law enforcement activity by numerous police agencies. The

Please cite this article in press as: Wolak, J., et al. Measuring a year of child pornography trafficking by U.S. computers on a peer-to-peer network. *Child Abuse & Neglect* (2013), http://dx.doi.org/10.1016/j.chiabu.2013.10.018

researchers were not able to design the methods of data collection and did not have any control over how law enforcement used the RoundUp software. In addition, law enforcement agencies were actively pursuing cases. Some of the computers included in our data and the images they shared were probably taken offline by police during the study year, but we had no way of taking police activity into account in our analyses. Fifth, participation in the Gnutella network may have declined significantly at the beginning of the study year because of legal rulings related to violations of copyright infringement laws. The generally low levels of CP trafficking we observed by U.S. computers could have been related to this decline.

Discussion

We measured one year of CP trafficking by U.S. computers on the Gnutella P2P network with data gathered via RoundUp software, which is used by law enforcement to proactively investigate CP trafficking in P2P networks. Our data provide the most complete picture that we know of to date of CP trafficking from U.S. computers on a P2P network. We found that during the study year, 244,920 U.S. computers shared files that had been identified in previous law enforcement investigations as containing child pornography. While that number seems quite large, it should be seen in proportion to the almost 85 million U.S. households that had computers in 2010 (U.S. Census Bureau, 2012). If the U.S. computers we identified as trafficking in CP were located in 244,920 different households, this would amount to 0.3% of U.S. households with computers. However, this percentage is a lower bound estimate. The actual number of U.S. computers trafficking in CP on P2P networks could be considerably higher because our data could not comprehensively measure the number of CP files being shared on Gnutella, and because Gnutella is only one of several P2P networks where child pornography can be acquired and distributed. Further, P2P networks are global in scope and these 244,920 U.S. computers contributed about one-third of the known CP files that were shared worldwide on the Gnutella P2P network during the study timeframe.

We also found that most of the U.S. computers that shared known CP files on Gnutella were not very active. More than 80% were observed sharing fewer than ten known CP files or sharing such files for fewer than ten days during the study year. About 40% shared only a single known CP file during the study year, and close to half shared known CP files for only one day. Almost 30% shared only one known CP file for only one day. Most U.S. computers shared only files with many duplicates that were readily available in many shared folders on the network. Consequently, they contributed minimally to the number of known CP files available on Gnutella.

Unfortunately, we could not compare the P2P activities of CP traffickers to the activities of other P2P users. RoundUp did not collect such data and we could find no research that described overall patterns of file sharing in the Gnutella P2P network. Consequently, we cannot say whether our observations of low levels of file sharing by most CP traffickers are characteristic of P2P users in general, who might or might not tend to share small numbers of files, leave files in shared folders for short periods of time or pursue file sharing for relatively few days during a year.

When evaluating our findings, it is important to recognize that our data pertain to the overall population of P2P users who traffic in child pornography, rather than the subgroup of cases that are pursued by law enforcement. Law enforcement cases are probably more likely to involve the active CP traffickers who share large numbers of files. Also, our data have significant limitations and may considerably underestimate the number of computers trafficking in CP and the number of CP files being trafficked. On the other hand, while we consider our findings to be conservative lower bounds estimates, there is some support for low levels of file sharing among U.S. CP traffickers. Steel (2009) analyzed searches for child pornography on Gnutella and responses. He found that almost 30% of the searches for CP in his dataset originated in the U.S., but only 6% of the files that were located in response to searches were shared by U.S. computers. In other words, the demand for child pornography in the United States was greater than the supply, a finding that Steel suggests is attributable to strict U.S. laws and enforcement.

Prioritizing Law Enforcement Investigations

Clearly, with tens of thousands of U.S. computers making very small annual contributions to the number of known CP files available on Gnutella, indiscriminant arrests of CP traffickers may not be effective in reducing the scope of the problem (although arrests serve other purposes, such as deterrence). Our analyses suggest that law enforcement may be able to effectively and meaningfully reduce the number of known CP images circulating in a network by using tools such as RoundUp to identify and target high-contribution computers. Further, prioritizing law enforcement cases in this way could also help the victims depicted in CP images by giving them hope that the proliferation of CP in P2P networks can be abated.

Our findings also point out the importance of international cooperation in attacking this problem. Even if U.S. law enforcement could remove all U.S. computers that were trafficking in CP on Gnutella, most of the CP trade on Gnutella would be left intact and available to network users worldwide. At the same time, a concerted global response that consistently targeted high-contribution computers could have a substantial impact over time.

Implications

Despite limitations, our findings show real progress has been made in addressing the dilemmas posed by CP trafficking on P2P networks. RoundUp and similar investigative tools can accurately identify child pornography files and locate CP traffickers. Tools can be used strategically to prioritize investigative goals and the impact of strategic uses can be measured. Reducing the amount of child pornography circulating on P2P networks is possible.

We also have shown that online CP trafficking in P2P networks can be measured. We examined data derived from observations of the Gnutella network, but RoundUp and other tools are observing other networks too and RoundUp is developing the ability to track CP traffickers between networks (Hurley et al., 2013). With continued progress, law enforcement observations of CP trafficking in P2P networks may provide enough data so that aspects of the online CP trade can be more fully measured and tracked over time. Such information can guide law enforcement responses, show shifts in the dynamics of the online CP trade and gauge the effectiveness of interdiction and eradication efforts.

Most importantly, investigative tools can be improved and limitations of data overcome. Continued investment in upgrades of RoundUp and similar systems created by computer scientists in collaboration with law enforcement would allow the criminal justice system to keep up with advances in technology and changes in file sharing networks and prevent investigative tools from becoming obsolete. Further, long term research can provide evidence-based guidelines for investigations and improve the ability of tools to prioritize targets for police and measure results. We need a variety of data that is currently unavailable but feasible. These data include accurate baseline measures of P2P CP trafficking in different P2P networks followed by methodical periodic measurements of change, systematic information about the number and content of CP files being shared online, and information about the characteristics of offenders who are trafficking in CP on P2P networks.

Developing a systematic way of adding the hash values of known CP files to the RoundUp system would allow for more comprehensive measurements of CP trafficking across P2P networks. This system could also provide a means for tracking the emergence and spread of previously undistributed images. Periodic measurements of trafficking in different networks would provide maps of fluctuations in the P2P CP trade and assessments of the effectiveness of law enforcement responses.

In addition, our knowledge about CP trafficking could be extended by finding ways to access and include information about the content of known CP files circulated in P2P networks. An accurate, consistent system for coding the content of known CP files could help answer pressing questions such as how many new images enter online circulation annually, how many victims are represented, and what is the online trajectory and life of an image or series of images. Further, one law enforcement goal is to target the subgroup of CP traffickers who also molest children (U.S. Department of Justice, 2010). Some research suggests that CP traffickers who are also molesters could be distinguished by the content of the CP they possess (Long, Alison, & McManus, 2012). A system for coding the content of images would allow researchers to test hypotheses about whether sharing certain CP content is associated with committing molestation offenses. Verified hypotheses could be used to program law enforcement tools to target computers sharing such content, which could help to increase the rate at which child molesters are detected among CP traffickers.

Conclusion

Using data collected via RoundUp, a law enforcement investigative tool, we found relatively low levels of trafficking in known CP files among U.S. computers on the Gnutella P2P network during one year. However, our measures were not comprehensive and represent lower bounds estimates; the low level trafficking was widespread and some computers made large contributions to the problem. There is no easy way to curtail CP trafficking on P2P networks, but our findings show that investigative tools can be used strategically to this end. Further, data that are systematically gathered and carefully analyzed can be used to develop an empirical grasp of the scope and characteristics of CP trafficking on P2P networks, which can be used to combat the problem.

References

Bazelon, E. (2013, January 24). *The price of a stolen childhood*. Retrieved from http://www.nytimes.com

Beech, A. R., Elliott, I. A., Birgden, A., & Findlater, D. (2008). The Internet and child sexual offending: A criminological review. *Aggression & Violent Behavior*, *13*, 216–228.

Hughes, D., Walkerdine, J., Coulson, G., & Gibson, S. (2006). Peer-to-peer: Is deviant behavior the norm on P2P file-sharing networks? *IEEE Distributed Systems Online*, *7*(2), 1–11.

Hurley, R., Prusty, S., Soroush, H., Walls, R. J., Albrecht, J., Cecchet, E., Levine, B. N., Liberatore, M., Lynn, B., & Wolak, J. (2013). Measurement and analysis of child pornography trafficking on P2P networks. In *Paper presented at the international world wide web conference* Rio de Janeiro, Brazil.

Jenkins, P. (2001). *Beyond tolerance: Child pornography on the Internet*. New York: New York University Press.

Koontz, L. D. (2005). *File sharing programs: The use of peer-to-peer networks to access pornography*. Washington, DC: U.S. Government Accountability Office.

Latapy, M., Magnien, C., & Fournier, R. (2013). Quantifying paedophile activity in a large P2P system. *Information Processing and Management*, *49*, 248–263.

Liberatore, M., Erdely, R., Kerle, T., Levine, B. N., & Shields, C. (2010). Forensic investigation of peer-to-peer file sharing networks. *Digital Investigation*, *7*, 95–103.

Liberatore, M., Levine, B. N., & Shields, C. (2010). Strengthening forensic investigations of child pornography on P2P networks. In *Paper presented at the ACM conference on future networking technologies* Philadelphia, PA.

Long, M. L., Alison, L. A., & McManus, M. A. (2012). Child pornography and likelihood of contact abuse: A comparison between contact child sexual offenders and noncontact offenders. *Sexual Abuse: A Journal of Research and Treatment*, *25*, 1–26.

Prichard, J., Watters, P. A., & Spiranovic, C. (2011). Internet subcultures and pathways to the use of child pornography. *Computer Law & Security Review*, *27*, 585–600.

Quayle, E., & Taylor, M. (2002). Child pornography and the internet: Perpetuating a cycle of abuse. *Deviant Behavior*, *23*, 331–361.

Rutgaizer, M., Shavitt, Y., Vertman, O., & Zilberman, N. (2012). Detecting pedophile activity in BitTorrent networks. *Passive and Active Measurement*, *7192*, 106–115.

Please cite this article in press as: Wolak, J., et al. Measuring a year of child pornography trafficking by U.S. computers on a peer-to-peer network. *Child Abuse & Neglect* (2013), http://dx.doi.org/10.1016/j.chiabu.2013.10.018

Seto, M. C., Hanson, R. K., & Babchishin, K. M. (2011). Contact sexual offending by men with online sexual offenses. *Sexual Abuse: A Journal of Research and Treatment*, *23*, 124–145.

Steel, C. M. S. (2009). Child pornography in peer-to-peer networks. *Child Abuse & Neglect*, *33*, 560–568.

Svedin, C. G., & Back, K. (1996). *Children who don't speak out: About children being used in child pornography*. Stockholm, Sweden: Rädda Barnen.

U.S. Census Bureau. (2012). *Computer and Internet use in the United States: 2010* [data file]. Retrieved from http://www.census.gov

U.S. Department of Justice. (2010). *The national strategy for child exploitation prevention and interdiction: A report to Congress*. Washington, DC: Author.

von Weiler, J., Haardt-Becker, A., & Schulte, S. (2010). Care and treatment of child victims of child pornographic exploitation (CPE) in Germany. *Journal of Sexual Aggression*, *16*, 211–222.

Wolak, J., Finkelhor, D., & Mitchell, K. J. (2005a). *Child pornography possessors arrested in Internet-related crimes: Findings from the National Juvenile Online Victimization Study*. Washington, DC: National Center for Missing and Exploited Children.

Wolak, J., Finkelhor, D., & Mitchell, K. J. (2005b). The varieties of child pornography production. In M. Taylor, & E. Quayle (Eds.), *Viewing child pornography on the Internet: Understanding the offense, managing the offender, helping the victims* (pp. 31–48). Dorset, UK: Russell House.

Wolak, J., Finkelhor, D., & Mitchell, K. J. (2011). Child pornography possessors: Trends in offender and case characteristics. *Sexual Abuse: A Journal of Research and Treatment*, *23*, 22–42.

Wolak, J., Finkelhor, D., & Mitchell, K. J. (2012). *Trends in arrests for child pornography possession: The Third National Juvenile Online Victimization Study (NJOV-3)*. Durham, NH: Crimes against Children Research Center, University of New Hampshire.

Wolak, J., Finkelhor, D., Mitchell, K. J., & Jones, L. M. (2011). Arrests for child pornography production: Data at two time points from a national sample of U.S. law enforcement agencies. *Child Maltreatment*, *16*, 184–195.

Wortley, R., & Smallbone, S. (2012). *Child pornography on the Internet (Problem-Specific Guides Series, Problem-Oriented Guides for Police, No. 41)*. Washington, DC: Office of Community Oriented Policing Services, U.S. Department of Justice.