Volume 5, Number 1, 2010

CHILD SEXUAL EXPLOITATION PROGRAM

UPDATE

What Every Prosecutor Should Know About Peer-to-Peer Investigations

By Sergeant Josh Moulin,[1] CFCE, DFCP, ACE, CEECS

Peer-to-Peer (P2P) file sharing is one of the fastest and easiest ways for individuals around the world to obtain and trade images and videos of child sexual exploitation. As of October 2007, the Wyoming Internet Crimes Against Children (ICAC) Task Force has captured 377,044 unique computers sharing image and movie files containing child sexual exploitation using the Gnutella network.[2] If you investigate and prosecute technology-facilitated crimes against children it is important to understand P2P and the way that sexual predators use it to exploit children.

P2P first received notoriety in 2001, in the *A&M Records v. Napster* case, when the 9th Circuit Court of Appeals ruled that Napster was liable for contributory infringement of record companies' copyrights and had to pay music creators and copyright owners 26 million dollars.[3] The court focused on Napster's role as a centralized P2P network; meaning that the data was stored on a centralized server owned by Napster. [4] The court found that because Napster had a direct role in distributing files to users who were searching for music, they were liable for copyright infringement.[5] The Napster case taught P2P users and developers a major lesson and lead to the evolution of the decentralized model of P2P networks that exists today.

As a result of the decision in the Napster case, today's P2P networks do not utilize a centralized server. This means that no one person is responsible for the content of what is being shared on P2P networks. Additionally, unlike Napster where the company's server could simply be shut down and the copyright infringement stopped, files are now shared and kept on multiple individual computers. While multiple P2P networks exist, the Gnutella network is a decentralized network that is by far the most popular and will be the focus of this article. Some other less popular but functionally similar P2P networks include FastTrack, BitTorrent, eDonkey, and Freenet.

How the Network Works

Since no centralized server hosts files, the Gnutella network depends on each client computer to share information about the files available for other users. When a computer connects to the Gnutella network it is designated as either an ultrapeer (sometimes referred to as a supernode) or a leaf. Ultrapeers are the backbone of the Gnutella network and act as traffic cops, directing incoming search requests to the appropriate computers that contain files or information that a user wants. Leaves are computers that do not have the capacity to support the Gnutella network and that connect to ultrapeers to share files.

Determining whether a computer is an ultrapeer or leaf is based on a number of factors. A computer with a fast Internet connection, not behind a firewall, and with a historically stable connection to the Gnutella network may be designated an ultrapeer. The average Gnutella user does not know if his or her computer is an ultrapeer or a leaf.

Each ultrapeer may be responsible for up to 30 individual leaves and for maintaining connections with other ultrapeers. Ultrapeers help to speed up the Gnutella network by allowing slower computers (leaves) to connect to an ultrapeer rather than the entire Gnutella network. While an individual leaf may have slower downloads because of connectivity issues, that particular leaf will not slow down the other computers connected to the various ultrapeers. Additionally, ultrapeers maintain lists of active Internet Protocol (IP) addresses connected to them and an index of files being shared. All leaves on the network periodically send an index of the files being shared to the ultrapeer they are connected to without any user interaction.

One way ultrapeers keep track of files available on their leaves is by maintaining a Distributed Hash Table (DHT) of the names of all their shared files. Normally hash values are a mathematical algorithm that produce a huge combination of letters and numbers that are the equivalent of a digital fingerprint for a particular file (such as MD5 or SHA-1). However, when referring to DHT, the hash value created is only for the filename and has nothing to do with the actual content of the file.

For example, let's say the words "Phil" and "Collins" create a hash value of three. The words "Phil," "Collins," "In," "The," "Air," "Tonight" create a hash value of nine. When a user searches for the term "Phil Collins" it translates to the hash of three and the ultrapeers begin looking for computers with a three in their DHT. Once the table is searched and matching hash values are found, the information is populated on a list on the computer that initiated the search. For example, if the above search was run, a list of all the possible files with the hash value of three would appear on the monitor of the computer where the search was entered.

Once the user receives the search results, he or she must then make a purposeful action to begin the download process. Depending on the software used to access the Gnutella network, this step can be accomplished several ways: double clicking the file, right-clicking it and selecting "download," highlighting several files and clicking on a download button, and/or any combination of these methods. What is important to those involved in investigating or prosecuting individuals possessing child pornography is showing that the user initiated the download. This action helps establish the element of intent required in such cases.

Gnutella also uses hash values to identify files independent of the filename and to confirm a file has been completely downloaded. This type of hashing is similar to what a forensic computer examiner uses, as opposed to the DHT values mentioned above. Gnutella uses a base 32 SHA-1 value to identify its files. This is important for forensic examiners to understand because it is different from the traditional base 16 hash values normally used in forensics. If a forensic examiner attempts to "hash" a file with traditional forensic software he or she will come up with a completely different hash value from the base 32 SHA-1 value Gnutella uses. Free software applications are available to convert the base 32 SHA-1 to a base 16 SHA-1.

Returning to our example, when a user attempts to download the song "In the Air Tonight," Gnutella will not look at filenames, but rather, for other computers sharing that file based upon its SHA-1 hash value. If, "In the Air Tonight" is being shared by multiple users then it is possible for the Gnutella network to obtain parts of the file from several users; instead of the user downloading the entire song from one computer, he or she will get a small piece from several different users sharing the same file. This allows a computer to simultaneously download different portions of the song, making the entire download process faster and more reliable.

Each portion of the file that is retrieved from the other leaves that are attached to the ultrapeer is then reassembled as one file on the computer that initiated the search. Once the computer that requested the file has completely downloaded the file, the software (LimeWire, BearShare, etc.) will hash the file and compare the downloaded file's hash with the hash value of the original file selected for download. If the hash values match, the download is complete and the file is moved into the folder used to store downloaded files. If the file is not completely downloaded, then it is moved into a folder

UNSAFE HAVENS I: INVESTIGATION AND PROSECUTION OF TECHNOLOGY-FACILITATED CHILD SEXUAL EXPLOITATON
APRIL 19-23, 2010
BOCA RATON, FLORIDA
This five-day course is designed to familiarize prosecutors with the various stages of an investigation, pre-trial and case preparation in a technology-facilitated child sexual exploitation case.

EQUAL JUSTICE FOR CHILDREN: INVESTIGATION AND PROSECUTION OF CHILD ABUSE
MAY 17-21, 2010
CHARLESTON, SOUTH CAROLINA
Equal Justice is a five-day comprehensive course designed to meet the needs of frontline child abuse professionals, including prosecutors, law enforcement, medical and mental health professionals, SANE/SART teams, emergency response teams, child protection and child advocacy center staff, social workers and victim witness advocates. Conference participants will be taught by and have the opportunity to interact with many of the nation's foremost child abuse experts.

UNSAFE HAVENS II: PROSECUTING TECHNOLOGY-FACILITATED CRIMES AGAINST CHILDREN
AUGUST 23-27, 2010
NATIONAL ADVOCACY CENTER
COLUMBIA, SOUTH CAROLINA
An advanced trial advocacy course focusing on prosecuting technology-facilitated child sexual exploitation cases from opening statements through sentencing.

SAFETYNET: MULTIDISCIPLINARY INVESTIGATION AND PROSECUTION OF TECHNOLOGY-FACILITATED CHILD SEXUAL EXPLOITATION
SEPTEMBER 27-OCTOBER 1, 2010
STONEHILL COLLEGE
EASTON, MASSACHUSETTS
This intensive five-day course is intended for prosecutors, investigators and computer forensic examiners investigating technology-facilitated crimes against children. This course includes hands on computer lab exercises.

For more information about these conferences, see our Web site at www.ndaa.org

UPDATE is published by the National District Attorneys Association's National Center for Prosecution of Child Abuse. Items may be reprinted if attributed to NDAA's National Center for Prosecution of Child Abuse. Please provide copies to NCPCA. Contact us if you have inquiries or article suggestions at 703.549.9222

used to store incomplete files. Forensically, two files that contain the exact same hash value have the same contents within the file.

How the User Software Works

Before an individual can access a P2P network he or she needs to install client software on the computer. Once again the purposeful act of downloading the software helps demonstrate the suspect's intent. Some P2P applications like LimeWire have a free version and a paid version. The paid version offers technical support and faster downloads. The installation of the P2P software allows the user to customize his or her settings or, as is common in most situations, utilize the default settings to allow sharing of the downloaded files. Tracking how the software is set up on the computer may again demonstrate not only intent but also the user's sophistication. For example, an investigation may find evidence that only certain types of files are shared or that the user has set up an additional file structure to store different types of images or movies he or she collects. These actions demonstrate the user's knowledge and intent.

Normally, the default settings create two folders to house the files that are downloaded and shared by the P2P client. These may be called, "Shared" and "Incomplete." However, the user can change the names, point the downloaded files to another existing folder or use a completely different file path. Any modification to the default settings is another point that a prosecutor could use to demonstrate the user's knowledge and intent. During the installation of most P2P software programs the software asks the user if he or she want to share files already existing on the computer and will advise him or her that files downloaded from Gnutella will be automatically shared. For prosecutors who are considering whether to file the most serious charge of dissemination of child pornography rather than possession of child pornography, this information is vital. Users can also select to share partial downloads, meaning that if he or she is in the process of downloading a file but have not completely downloaded it, the chunks the computer has received can be shared with others looking for that same file.

The P2P software maintains a library of files that have been downloaded or made available for sharing and the user can preview movies, music, or pictures right from within the P2P software. In some software programs it is possible to find evidence of the user previewing the file as it was being downloaded. For example, in LimeWire if a user previews a file as it is downloaded, it will create a file named Preview-T (file size in bytes) (filename) and save that file in the incomplete folder. A forensic examiner can search for "preview-t" and find hits in both allocated and unallocated space. This information provides valuable evidence of what a user was searching for and viewing. If the file is still in allocated space, the file may be viewable if enough of it was downloaded and dates and times should be associated with it. Users can also choose to have all audio, videos, images and documents currently on their hard drive automatically added to their library for sharing.

P2P Investigations

Many investigators around the world conduct undercover online investigations using P2P. ICAC task forces and other law enforcement agencies have specialized software to search for and identify individuals involved in the sharing of child pornography. These programs allow law enforcement officers to use common search terms for child pornography to locate images and videos of child sexual abuse on the Gnutella network. Once a list of files returns based upon the investigator's search term, the SHA-1 hash value is compared to known or suspected child pornography files in the ICAC database. Any files actively being shared that match a file in this database are viewable by the investigator. The investigator is also able to select a user who is actively sharing content and determine his or her Internet Protocol (IP) address and obtain a rough geographical location of the suspect so the investigator can focus on that jurisdiction.

Another feature contained in most P2P software applications is the ability to directly connect to another user's computer. It is common for this feature to be enabled by the default settings, however, it may be deactivated by the user. The direct connect feature is turned on in the default settings based on the theory that if a user finds someone sharing one item he or she is interested in, then odds are he or she may have other files of interest. By browsing the host the user makes a direct connection with the other computer and can list all shared files on his or her computer and select any of them for download. Investigators performing P2P operations should attempt to make a direct connection with a computer that has been identified as sharing child pornography to see if there are additional child pornography files. However, for someone to browse the host computer/hard drive through the direct connect feature the computer sharing the files must be actively online.

During an investigation, an investigator will usually run a NETSTAT command on their computer, which shows all active connections, i.e., which other computers are connected through the P2P software. The IP address of the suspect's computer should be listed in the NETSTAT response showing that the investigative computer and the suspect computer were directly connected. This evidence helps to demonstrate to a judge or jury that the child pornography that forms the basis of the charge came from the suspect. While this is great evidence to have, it may not be available as the suspect computer may have turned this functionality off on their computer or their computer is behind a firewall.

Another area of information available through the ICAC network is the Globally Unique Identifier (GUID) and the IP history of a particular computer. A GUID is a randomly assigned serial number to the P2P client. The GUID is captured by ICAC tools and entered into a database. While IP addresses can change due to dynamically assigned IP addresses, GUIDs are less likely to change and provide an investigator the ability to see how many times a particular GUID has come up in investigations done by other investigators. Investigators can also determine how many times a particular IP address has been captured sharing child pornography during other investigations.

When a P2P suspect's computer is sent to the forensic lab for analysis, there is generally an enormous amount of information available to the examiner. In addition to all of the normal computer forensic evidence the examiner should report on, a computer forensic examiner should also be able to tell you: what P2P software is installed on the computer; how long it's been on the machine; the file paths for the shared and incomplete folders; whether sharing was enabled on the computer; approximately how many times the P2P software has been used; and, the GUID of the P2P software. In situations where an investigator has a P2P case that did not start as the result of an ICAC investigation, the GUID should be checked against the ICAC database to see if it was ever captured during another investigation. This will allow law enforcement to establish that the suspect has been involved in distributing pornographic images of children as well as the length of time images have been on the machine.

Additional evidence comes from recovering search terms used in P2P programs by using forensic programs or by creating a virtual computer from the forensic image of the suspect's computer. By creating a virtual machine (VM) of the suspect's computer and booting it up, the examiner (and the jury) may view the computer exactly like the suspect would have seen it but in a forensically sound manner. Screenshots of the defendant's sharing preferences, his or her library, and other evidentiary items like their desktop background, folder structures and registry information are powerful courtroom exhibits. Several P2P software programs can store past search terms within the search bar; an examiner looking at a VM of the suspect's machine can drop down the past search terms and take screenshots of that as well. Having and presenting this type of evidence makes it difficult for a defendant to claim accidental download or unknowing possession when search terms related to child sexual exploitation are found in the P2P software's search history or even Internet search history.

When done correctly, a P2P investigation and forensic computer examination will reveal multiple layers of evidence to help prove possession, and possibly dissemination, of child pornography. It is important for a prosecutor to understand the components of P2P software so that they can explain this evidence to a jury. Armed with this information investigators and prosecutors can initiate successful cases against individuals who use P2P platforms to harm children through the possession and distribution of child pornography.

[1] Sergeant Josh Moulin is the Commander of the Southern Oregon High-Tech Crimes Task Force that provides digital evidence forensics and cyber crime investigations to over thirty federal, state, and local law enforcement agencies. Sgt. Moulin teaches computer forensic topics for the NDAA/NCPCA in courses such as Unsafe Havens.

[2] Waters, Flint. Prepared for House Judiciary Committee. *Child Sex Crimes on the Internet*. Oct. 3, 2007. Available at: judiciary.house.gov/hearings/pdf/Waters071017.pdf; Accessed: 3/9/10.

[3] *See A&M Records v. Napster Inc.*, 284 F.3d 1091 (9th Cir. 2002).

[4] *Id.*

[5] *Id.*

The National Center for Prosecution of Child Abuse is a program of the National District Attorneys Association. This publication was prepared under Grant No. 2007-JL-FX-K005 from the Bureau of Justice Assistance, DOJ Office of Justice Programs. This information is offered for educational purposes only and is not legal advice. Points of view in this publication are those of the authors and do not necessarily represent the official position of the U.S. Department of Justice and NDAA.

National District Attorneys Association
National Center for Prosecution of Child Abuse
44 Canal Center Plaza, Suite 110
Alexandria, Virginia 22314
www.ndaa.org