Forensics in Peer-to-Peer Sharing and Associated Litigation Challenges

Presented by
Mo Hamoudi, Seattle, WA (Mo_Hamoudi@fd.org)
Terry Lahman, Snoqualmie, WA (terry@eforensicspro.com)

1 Statement of Probable Cause

"Between April 08, 2016, and April 09, 2016, while acting in an undercover capacity, I used a law enforcement version of eMule, a commonly used P2P file sharing program for the eD2k file sharing network, to monitor for P2P users possessing and distributing image and video files depicting child pornography. I used the law enforcement version of eMule to download several files depicting child pornography from a P2P user at IP address <redacted> (the SUBJECT IP ADDRESS)."

The statement of probable cause implies the detective was sitting at a computer utilizing a special version of software to identify and download suspected child pornography. It also implies that the specialize software is running autonomously on the detective's computer. The law enforcement version of eMule runs automatically without user invention. And the law enforcement computer is only one component of a large scale network of computers.

Digging into the technical data of the law enforcement version of eMule required a deep investigation similar to peeling an onion, one layer at a time. The entire process required a number of discovery requests to peel back each layer. Each new discovery item was analyzed to dig deeper into the next layer.

PRACTICE POINT: The practitioner needs to use the Federal Rules of Criminal Procedure 16 or its state counterpart to obtain information beyond the affidavit. *United States v. Soto-Zuniga*, 837 F.3d 992 (9th Cir. 2016)

Note: An expert is necessary to make the appropriate requests.

PRACTICE POINT: The practitioner, depending on the circumstances, can attack the affidavit. *Franks v. Delaware*, 438 U.S. 154 (1978)

2 Technology Definitions

2.1 eDonkey 2000 Peer-to-Peer Network

Peer-to-Peer (P2P) file sharing is a method of communication available to Internet users through the use of special software programs. P2P file sharing programs allow groups of computers using the same file sharing network and protocols to transfer digital files from one computer system to another while connected to a network, usually on the Internet. There are multiple types of P2P file sharing networks on the Internet. To connect to a particular P2P file sharing network, a user first obtains a P2P client software program for a particular P2P file sharing network, which can be downloaded from the Internet.

In general, P2P client software allows the user to set up file(s) on a computer to be shared on a P2P file sharing network with other users running compatible P2P client software. This is accomplished when the P2P client software communicates with a centralized computer server that stores the shared file hash value, filename, and the IP address of P2P clients that are currently sharing the file.

A user can identify and download files using the P2P client software on the user's computer by conducting a keyword search for files that are of interest and currently being shared on the P2P file sharing network. The results of the keyword search are displayed to the user. The user then selects file(s) from the results that he wishes to download. The download of a file is achieved through a direct connection between the computer requesting the file and one or more computers on the same P2P network containing the file.

The advantage of P2P software over downloading from a central computer server is that the P2P software will download the files from multiple computers that are sharing the same file. This significantly improves the download speed of the file.

There is a protocol specification that defines how the P2P clients communicate with one another when sharing and downloading files.

PRACTICAL POINT: This is very similar to the Napster Network that was used to exchange music.

2.2 P2P Client Software

There are multiple client software programs that enable a user to share and download files on the eDonkey2000 network. eMule and Shareaza are two of the popular client software programs that enable a user to share and download files on the eDonkey2000 P2P network. The eMule program extends the eDonkey2000 protocol specification. The extended protocol enables the ability to track the user ID of the peer computer and the number of bytes downloaded or uploaded with the other peer.

2.3 eDonkey2000 Index Servers

Index servers store information about files that are shared on the eDonkey2000 P2P network. The actual file is not stored on the index server. Index server computers allow P2P client software to locate the IP addresses of computers that are sharing files. A P2P Client can execute a keyword search to identify file names and computer IP addresses that match the keyword search. The user can select the file to download and begin a download of the file from the computers that are sharing the file.

The index servers implement protection to prevent an excessive number of searches from any one IP address. This prevents a client program from fully automating searches to run continuously on the eDonkey2000 P2P network. This is very similar to Google detecting and preventing automated searches.

IMPORTANT POINT: The files are not on the index servers but on the client’s computer.

2.4 Internet Protocol (IP) address.

An Internet Protocol (EP) address is a numerical label assigned to devices communicating on the Internet. The Internet Assigned Numbers Authority (IANA) manages the IP address space allocations globally. An IP address provides the methodology for communication between devices on the Internet. It is a number that uniquely identifies a device on a computer network and, using transport protocols, moves information on the Internet. Every device directly connected to the Internet must have a unique IP address.

An IP address is typically comprised of four series of numbers separated by periods and is most commonly represented as a 32-bit number such as 71.227.252.216 (Internet Protocol Version 4) however, a newer version, IPv6, is currently being deployed as well and is represented as a 128-bit number such as 2001:db8:0:1234:0:567:8:1.

IP addresses are owned by the Internet Service Provider (ISP) and leased to a subscriber/customer for a period of time. They are public and visible to others as you surf the Internet.

IP addresses are similar to a license plate on a motor vehicle. They are the property of the issuer, and not the vehicle owner. Just as a license plate is visible as you cruise your city or town, your IP address is visible as you cruise the Internet. Your IP address is visible to the administrators of websites you visit, attached emails you send, and broadcast during most Internet file and information exchanges that occur on the Internet.

PRACTICE POINT: Although the IP address is the property of the issuer, the IP address's activities may be entitled to an expectation of privacy to the extent they are aggregated to obtain information otherwise unavailable through ordinary observations. *United States v. Jones*, 565 U.S. 400 (2012); *Carpenter v. United States*, 585 U.S. ___ (2018)

2.5 Hash Value

A unique numerical identifier that can be assigned to a computer file, a group of files, a folder and its contents, a computer hard drive, etc. Based on a standard mathematical algorithm. All of the data from the source is applied to the mathematical algorithm and the result is the hash value. There are different algorithms used for different purposes. Here is an example of a hash value generated with the MD5 hashing algorithm: ECD834C3A570A5336C798A8E2FC2F0A1.

PRACTICAL POINT: This is similar to DNA.

3 RoundUp Network of Computers

The government uses a large- scale network of computers to conduct surveillance on the eDonkey2000 peer-to-peer file sharing network, called RoundUp. It consists of thousands of computers at law enforcement agencies running RoundUp eMule, a centralized server that stores a database of hash file values related to child pornography, and a centralized server that stores a database of eDonkey2000 search results identifying the hash value and IP address of computers that were found to share child pornography related files.

DISCUSSION QUESTIONS: Is this a generalized warrantless search in violation of the Fourth Amendment even though it is occurring on a network accessible by anyone? What about standing? What about plain view?

3.1 Hash Value Server

A centralized server stores hash values of files related to child pornography. The hash values are used to search the eDonkey2000 P2P network for computers that are sharing the files. There are millions of hash values stored in the database. It would take over one year for one single computer to execute searches on the P2P network for all the hash values in the database. Instead, the hash values are distributed by the hash value server to the law enforcement computers running the RoundUp eMule program for execution on the P2P network. In this manner a search of all the hash values can be executed in a reasonably short period of time.

DISCUSSION QUESTION: The government is leveraging technology unavailable to the general public to conduct surveillance on millions of IP addresses. *Kyllo v. United States*, 533 U.S. 27 (2001); *Carpenter v. United States*, 585 U.S. ___ (2018)

3.2 RoundUp eMule

RoundUp eMule is the law enforcement version of the P2P client software. It was based on the eMule program and modified for use by law enforcement. RoundUp eMule is a version of the open source eMule P2P client tailored to the needs of law enforcement. The program is available only to law enforcement, and the underlying source code is closely guarded by the government.

RoundUp eMule consists of two separate programs executing on a law enforcement computer, the Scheduler, and the Downloader. The two programs were developed from scratch using some of the source code from the original eMule to speed development of the software and ensure compatibility with the eDonkey2000 P2P network.

3.2.1 RoundUp eMule Scheduler

The scheduler program is used to executed the distributed searches from the hash value server. It requests a block of hash values from the hash value server and executes the searches on the eDonkey2000 network to identify the IP address of computers that are sharing files with those particular hash values.

The results of the searches are stored in a database on the centralized search results database. There are no files downloaded during this process. The scheduler only executes the searches and identifies the IP address of computers that are sharing files with the same hash value.

The Scheduler throttles the searches on the network to prevent detection by the index servers. It can only execute a small number of searches every 5 minutes. If the program is detected it would be blocked from using the index server for a period of time. This prevents any one computer from executing too many searches on the network.

The government has bypassed (hacked) around the limitation by using the hash valuer server to distribute the hash value searches across thousands of law enforcement computers to prevent detection.

3.2.2 RoundUp eMule Downloader

The Downloader program retrieves search results from the centralized search results database server and identifies only the IP addresses that are within the jurisdiction of the law enforcement agency running the software. The search results retrieved are only the most recent results. The program attempts to download the child pornography related files that were shared by a target computer. The program is documented by the government as performing a single source download, i.e. it will only download the file from the target computer whose IP address was identified during the search of the P2P network.

The Downloader program generates a large number of log files to document the download attempts that identify the hash value, file name, IP address and other data.

PRACTICE POINT: The three components of RoundUp eMule; RoundUp Scheduler; RoundUp Downloader are not regularly subject to validation testing. *United States v. Budziak*, 697 F.3d 1105, 1112 (9th Cir. 2012).

PRACTICE POINT: There may be a basis to attack the software and opinions related to it under FRE 104, FRE 702, and *Daubert v. Merrell Dow Pharmaceuticals, Inc*., 509 U.S. 579 (1993)

3.2.3 Summary of the Process used by the Government

- RoundUp eMule Scheduler retrieves some hash values from the hash value database server. Thousands of law enforcement computers are doing the same thing.
- RoundUp eMule Scheduler executes a search on the P2P network for the hash values and stores the results in the search results database server. Thousands of law enforcement computers are doing the same thing.
- RoundUp eMule Downloader retrieves recent search results for their jurisdiction.
- RoundUp eMule Downloader attempts to download the child pornography related files from the IP address that was saved in the search results.
- RoundUp eMule Downloader generates log files to document the attempted download process. It also saves any files that were downloaded during the process.

3.3 Search Results Server

The search results server is a centralized server where hash value search results are stored in a database. The RoundUp eMule Downloader retrieves the most recent search results for IP addresses with the law enforcement agency jurisdiction.

There is a web based user interface where law enforcement can search and examine the data stored in the database. As an example, the user can enter an IP address and a report is generated of all the search results stored in the database for that IP address.

PRACTICAL POINT: The government is maintaining a database in perpetuity of all IP addresses that access files the government deems files of interest. The files are purported to be files involving contraband but the files are not necessarily verified.

4 RoundUp eMule log files

RoundUp eMule generates multiple files during a download attempt. There is a lot of information in the files that can be analyzed and utilized to assist in creating discovery requests while peeling back each layer of the RoundUp network to determine how it operates.

4.1 Netstat log file

Netstat is a Windows program that lists all the IP addresses the computer is connected to at the time the program is executed. RoundUp eMule saves the results in a log file every time a download is attempted from a target computer. The information is logged by RoundUp eMule to show that it was connected at the time of the attempted download to the IP address of the target computer. This log file was instrumental in identifying two IP addresses of computers in the Pennsylvania area. Through a discovery request the government documented the existence of the Hash Value Server and the Search Results Server.

4.2 Download log file

The download log file documents all the information related to an attempted or completed download process. The file logs the software and user hash value of the target machine. The government can use the user hash value and the name and version of the target machine to support identifying the target computer. (the user hash value is generated on a computer when the P2P client software is installed on the computer)

The download log file also documents how much of the file the target computer is currently sharing on the P2P network. A file is downloaded in parts. A large file is divided into parts to improve the download performance from multiple computers. The parts are not downloaded sequentially, they are downloaded as quickly as possible from the computers that are sharing the files. Many of the log files in a case involving RoundUp eMule document that many of the files RoundUp eMule attempted to download were only some of the parts of a file, not the entire file. It's an indication that there is a short delay between the time a RoundUp eMule Scheduler program identified an IP address during a search, logged it on the Search Results Server and the time a RoundUp eMule computer within the jurisdiction was able to attempt a download.

4.3 Identity Logging

This log file contains the identity information for the protocol extension implemented in the eMule P2P software. The information stored in the log file documents the user hash value and the public key used by both the RoundUp eMule computer and the target computer. This information documents the tracer tags that are sent to the target computer for storage. The eMule protocol extension is documented in the next section.

PRACTICAL POINT: The log files help in establishing a digital chain of custody and digital trace evidence.

5 eMule Protocol Extension

The standard eMule program contains an extension to the eDonkey2000 Network protocol to support a credit system. The credit system is used to reward users that share files. When a standard eMule program downloads files from another eMule program, the two computers conduct a challenge and response encryption protocol to authenticate the users. They exchange a user hash and an encryption key that are stored in a file on the computer. The number of bytes downloaded/uploaded are also stored. Future file transfers can identify the computer used in past transfers and reward the computer that is sharing and uploading files. From the specification:

“eMule supports a credit system in order to encourage users to share files. The more files a user uploads to other clients, the more credit it receives and the faster it will advance in their waiting queues.”

“The user ID plays an important part in the credit system, this provides motivation for ’hackers’ to impersonate to other users in order to receive the privileges granted by their credits. eMule supports an encryption scheme which is designed to prevent fraud and user impersonation. The implementation is a simple challenge-response exchange which relies on RSA public/private key encryption.”

The encryption scheme is called Secure User Identification. Secure User Identification is enabled by default on the standard eMule client program; however, there is a configuration option to disable the feature.

PRACTICAL POINT: The more your share files on the network, the better your experience on the network. The network is designed to promote sharing. The less you share the network, the worse your experience.

5.1 Tracer Tagging

Secure User Identification isn’t a useful feature to RoundUp eMule because the government never shares files for uploading by other peers. The government implemented the feature anyway to “tag” the target computer with its public key for use in identifying the target computer as the one that the file was downloaded from.

The intent is to place a tracer tag on the target computer that can be located when the computer is seized to document that this is the same computer used to download a file related to child pornography.

PRACTICE POINT: This may constitute a digital trespass under the Fourth Amendment. *United States v. Jones*, 565 U.S. 400, 409 (2012); *see also Taylor v. City of Saginaw*, 922 F.3d 328 (6th Cir. 2019)

6 Summary

The RoundUp Network of computers and software is a complex system with extensive interaction and communication between sub-components. It uses advanced technology and designs well beyond the capabilities of the typical user. The design avoids excessive search detection by distributing the searches across multiple law enforcement computers. The design employs a client/server model with distributed computing technology to collect search result data about users of the eDonkey2000 Network.

The government intentionally implemented the Secure User Identification protocol in the RoundUp eMule client program to place a tracer tag on the target computer for later retrieval to identify the target computer. Secure User Identification is an optional protocol and not required by RoundUp eMule to attempt file downloads from a target computer. The tracer tag is sent to the target computer upon initial communication and before any attempted file downloads. This is similar to law enforcement placing tracer tags, such as chalk marks, on the tires of parked cars and then later returning to identify cars that still have the tracer tags.