Topic: P2P Digital Forensics

Executive Summary:

P2P networks are becoming a larger and larger part of the web connected world. As such, they are a perfect way to share files illegally. A DF investigator needs to be able to quickly determine what music and video on a computer system is illegal, and what is legal. Along with this it is important to find out where the illegal files came from in order to prosecute the distributor. The main sources of this information include Digital Watermarking, Digital Fingerprinting, and Live Monitoring.

DF Purpose:

Determine if a song or video is an illegal copy, and if it is, who was the source of the copy. Determine if illegal media traffic is moving over the network.

State of Practice:

The main tools at the moment in this field are active monitoring software. This includes firewalls, packet shapers, and Trojan clients. Firewalls can be set up in one of two main "wizard" modes. The first being "block everything", and the system admin deal with unblocking ports as users complain. This works good for company firewalls, since they don't often have to deal with people needing ports open, since only business programs should be running on the inside of the firewall anyway. The other being "block known" where the system admin provides a list of known programs to block and leaves the rest of the ports open. This is often the case at an ISP, where maximum service is needed for all programs, and as such ports can not often be blocked. This is a weakness, since it leaves ports open for P2P clients that can dynamically adjust to avoid the firewall.

To combat this, there are packet shapers that look for some indication of the type of traffic that is coming across the network, and blocks traffic that looks like P2P activity.

But, even if you are blocking a P2P system, some data will likely get through. Protocols like bit-torrent include encrypted modes that keep a packet shaper from seeing the traffic as P2P traffic. For cases like this, and for cases where you are hunting down a distributor, there is also the less forensically sound "Trojan Client" technique. This involves gaining access to the P2P network, and using a modified client to log the P2P traffic in hopes of finding out information about the hosts of a known illegal file.

Advanced practices in this field include Digital Fingerprinting and Digital Watermarking. These tools allow distributors of digital media to identify their content, sometimes even after it has been recompressed or sampled. Thus files downloaded over a P2P network can be tagged to the person who owned the original source file.

The simple version of media recognition is the robust watermark. This is some data added to the media that can be identified as unique to the media file. That data has to follow two main rules. The first is that the watermark should not be visible in the output media. The second is that, the watermark should be hard to remove (this can include the ability to withstand compression and re-encoding). A watermark then allows someone to check a media file against the watermark database to see who recorded and distributed the original file. The problem is that because the watermarks are part of the original distribution of the media file, and because there is no one standard for watermarking, there is no way to know without inspection what watermark might be in a media file and how to extract it.

The more complex data recognition technique is a Digital Fingerprint. This is a technique where the computer attempts to create a hash value for a file based on the semantic content of the file. If a human can recognize that a movie is a movie whether it is highly compressed, or original source, then the Digital fingerprint should match, or almost match from one extreme to the other. This gives an investigator some indication that a media file matches a known copy righted work with some degree of certainty. This can then be used to prove that the media file is either a legal copy, or an illegal rip based on the circumstances.

Gaps in Technology:

Private P2P networks, for the moment investigators need to gain access to someone on the network to find out what is on the network, or that the network even exists. That in mind, the associated encrypted P2P traffic is often useless other than an indication of possible illegal activity.

There is minimal standardization of watermarks used by different companies. While this is good for keeping the watermarks safe from hackers, it makes an investigator's job much harder. Centralized databases of companies, files, and the associated watermark readers are defiantly lacking.

There are few active Digital Fingerprint solutions. Most the working ones are designed around identifying audio files only (though this could be adapted to identify a movie via the audio tracks).

State of Research:

There is a tone of active research into passive ways to mark up digital media files so that either a computer can recognize the file, or so that the authenticity of a file can be confirmed.

Future Of The Practice:

With the absolutely vast quantity of malformed file metadata and number of home edited files, the need for continuation of automated media recognition will increase. There is also a need for centralized databases of Digital Fingerprints with associated Digital Watermarks. Though with any system, the issue is security of the information so that where possible media pirates are unable to obtain the required information to fool fingerprinting software or remove watermarks.

REFRENCES

JungHee Seo, HungBog Park, "Data Protection of Multimedia Contents Using Scalable Digital Watermarking," *icis* , pp. 376-380, 2005.

Zhu, Y., Zou, W., and Zhu, X. 2006. Collusion secure convolutional fingerprinting information codes. In *Proceedings of the 2006 ACM Symposium on information, Computer and Communications Security* (Taipei, Taiwan, March 21 - 24, 2006). ASIACCS '06. ACM Press, New York, NY, 266-274. DOI= http://doi.acm.org/10.1145/1128817.1128856

Links:
http://www.audiblemagic.com/products-services/registration/