A New Generation of File Sharing Tools

Excessive file sharing can have serious effects on a variety of organizations. These can range from lost revenue to lost productivity and wasted resources. Napster demonstrated that there is a huge demand for this material and it will only increase as more feature films appear online. IT security professionals need to know what the risks are and what techniques the authors of files sharing tools are using. Downloading a few ripped DVDs can bring your network down just as quickly as a targeted denial of service attack a...

A New Generation of File Sharing Tools
Dan Klinedinst
GSEC Practical Requirements (v.1.4b)

1. Introduction

1.1. Executive Summary

Despite the demise of Napster, online file sharing tools are only growing more and more popular, aided by the increasing number of Internet users who have broadband connections and the increasing level of sophistication of those users. Setting aside the debates over the ethics of file sharing and where to draw the line on copyright protection, we can assert that these tools can be real problems for IT departments, ISPs and content providers. After initially being far behind the technology curve, these groups have made many strides recently in countering the ill effects of file sharing. However, the authors of file sharing tools have not been idle either. Several tools have been released and promoted recently that attempt to thwart or circumvent the defensive measures put in place by the defenders. Certainly, some of these tools can be used for legitimate purposes, and the technology being implemented is quite impressive. However, it behooves today's information security professional to know the details of how these new tools work and what the threat is.

1.2. What is peer-to-peer

The basis of file sharing tools is peer-to-peer networking. Peer-to-peer can be defined as "a communications model in which each party has the same capabilities and either party can initiate a communication session". [searchNetworking.com] This is in contrast to client/server networking, where a client initiates a communication session, but the server generally has more capabilities such as access control and, usually, more computing power and resources. In partial peer-to-peer networking, a central server or group of servers is used initially by clients to obtain directory information, such as the IP address of another client or the location of certain files. The other client is then contacted directly and the server is no longer involved in the communication session. Examples of partial peer-to-peer networks include Napster (where central servers recorded the individual machines where each file could be found) and instant messaging (where central servers track the IP address of each user who is logged in.)

Full peer-to-peer networking does away with the requirement for a server altogether. In this case, a client obtains directory information either by probing the network for other clients, or by getting information from other known clients (similar to how RIP, the Routing Information Protocol,

disseminates information between routers.) The advantage to peer-to-peer networking is the lack of a requirement for large, expensive servers to dish out information. The biggest disadvantage is that without such centralized hubs, it is much more difficult to control the information on the network.

1.3. History of file sharing

The first peer-to-peer program to be widely used on the Internet was the ICQ instant messaging client. Released in July 1996 [Einstein], ICQ had grown to the point where download.com reported it as having been downloaded 200 million times in May, 2002. [Cunningham] While it was originally used only for instant messaging and chat, ICQ also had the ability to easily send and receive files between users. However, the need to have the transfer initiated by the possessor of the file and the lack of a directory listing available files limited its use for file distribution.

In May, 1999, Shawn Fanning released the Napster file sharing client. [DeJesus] Napster had several major advantages over sharing files with ICQ. It could contact another user running Napster and download any files they had made available without their cooperation (or even knowledge); it could automatically search a new user's computer and catalog all files that might be of interest to the Napster community (typically MP3 music files); and it maintained a directory of files so that you could search on file names to find files (MP3s) you wanted. These advantages made it an almost instant hit and it quickly grew to a user base of 60 million people.

However, the vast majority of files being shared on Napster were copyrighted music that was being disseminated without the permission of the copyright holder. The copyright holders, in the form of the RIAA, sued Napster and, since the company maintained a database of all files available and therefore was clearly party to this illegal copying, managed to shut down the service.

Shortly, other file sharing clients started to be released, many based on the Gnutella protocol. These programs did not have any kind of central directory and contacted other users only through information gleaned from other clients. While these clients currently enjoy widespread popularity, they suffer from several flaws which, although good for the defenders of network integrity and copyrights, are sufficient grounds for the most ardent users of these systems to design newer, better file sharing clients.

1.4. Who is effected

As previously mentioned, some of the groups that suffer the most detrimental effects of these programs are IT departments, ISPs and copyright holders. Each has different ways of responding to the problems that plague them.

IT departments are typically concerned about three things. One is the utilization of their organization's resources for purposes not related to the goals of the organization. Most organizations do not mind if computer and Internet users use a small amount of disk space or network bandwidth for personal files or emails, but file sharing programs can quickly eat up hundreds of gigabytes of disk space and/or saturate a network with music files and, increasingly, video files. Another problem is that many of these files, if downloaded to an organization's computers, become illegal copies of copyrighted material. The organization could be held responsible for these criminal activities. Finally, when users are spending time searching out new music or video files, they are not working and thus negatively impacting productivity.

The IT departments use a variety of methods to prevent the use of these programs within the organization. The most important is a strong policy prohibiting their use, complete with enforceable ramifications for non-compliance. Others include scanning network traffic for suspicious patterns (such as ".mp3" file extensions), blocking TCP ports known to be used for these purposes, and monitoring computers for known software clients.

ISPs are mostly concerned with the network utilization these programs create. They are beginning to respond to this threat by charging premium fees for "power users" who use substantially more bandwidth than other users. [Kingsbury]

Finally, content providers and copyright holders are concerned with the potential for lost revenue because would-be customers download their intellectual property for free rather than paying for it through a certified distributor (record store, video rental store.) They have tried various ways of protecting their content, including encrypting digital content so it will only be playable on certain devices, making the digital files "un-copyable", and introducing fake files into the file sharing networks themselves, hoping to frustrate users trying to download the most popular offerings. There is even a bill currently being considered by Congress that would allow copyright holders to "hack" (gain unauthorized access to) the computers of anyone suspected of possessing illegal copies of intellectual property. [McCullagh]

1.5. Goals of authors

The authors of file sharing software have a variety of motivations. Some of them are looking for recognition in the open source and hacking communities. Some want to solve specific problems that impede their abilities to share files. Some want to strike a blow for ideals related to privacy or freedom of speech. Others just want to annoy corporations and lawyers. Despite their varied motives, they are all essentially pursuing the same specific goals. We can sum these up as six major characteristics of the next generation of file sharing tools: deniability, anonymity, resource conservation, stealth, immunity to attack and scalability.

Deniability is the ability to refute any accusation of wrongdoing. This is mostly achieved by making it impossible to prove said wrongdoing.

Anonymity is the ability to engage in an activity, in this case file sharing, without anyone being able to determine your true identity.

Resource Conservation is the goal of getting the maximum efficiency out of resources such as storage, bandwidth and processing power.

Stealth is the ability for the file sharing activities to happen undetected by those who are not participating.

Immunity to attack describes defenses against several forms of attack that file sharing networks are vulnerable to.

Scalability means the file sharing network can grow to a very large (tens of millions of users) size without a negative impact on any of the other goals.

1.6. Implementations

There are a variety of new tools that attempt to overcome shortcomings both in the original Napster model and the newer Gnutella-based tools such as Kazaa, Morpheus and Limewire. The two most popular and most technologically advanced are Freenet and GnuNet. GnuNet may be slightly more technologically advanced, but it suffers from the problem of only running on Linux, Solaris and BSD at this time. [ovmj.org] Freenet, on the other hand, runs on Windows, Linux and MacOS. [freenetproject.org]

1.7. List of major advances

These are some of the key ways in which clients like Freenet and GnuNet have advanced. We will go into a few of the key ones in more detail later in the paper.

1.7.1. Encryption

Any security professional knows that encryption is not only one of his or her key tools in enforcing confidentiality and integrity. It can also be the bane of your existence when you are trying to monitor traffic for bad patterns such as viruses or prohibited material. This can obviously benefit users who are illicitly sharing files since there is no way to tell copyrighted material, porn or company secrets from any legitimate traffic that is being transferred over the network.

1.7.2. File splitting

File splitting is exactly what it sounds like. The file sharing client takes a file and splits it into chunks, which are then dispersed across different nodes of the network. This has many advantages, which are detailed below.

1.7.3. Economy

GnuNet is the only client that uses the concept of economy. This essentially means that a user gets more out of the system if they contribute more. For example, someone who contributes an inordinately large number of MP3 files may get a higher percentage of the available bandwidth than someone who contributes little.

1.7.4. Integrity Checks

Integrity checks are used in two ways. One is to ensure that a file comes from a certain user. That way, if it's a user that you trusts (or the network suggest you trust, due to their position in the economy), you can be reasonably sure the file is what it claims to be. This is increasingly important as record and movie companies are currently releasing fake files onto the file sharing networks in the hopes of drowning out good copies in a sea of fakes. Integrity checks can also be used to verify that a file really is what it claims to be based on the fact that it matches a known good signature (much the way professionals use MD5 sums or PGP keys to verify downloaded software.)

1.7.5. File system attributes

It should be noted that a peer-to-peer network used to share files is essentially just one big way to store information. Looked at from this point of view, it is obvious that such a system will need many of the same attributes that a traditional filesystem would need.

1.7.5.1. File storage vs. file sharing

The first key point is that the first generation of file sharing tools were just that – file sharing. You made the files on your computer accessible to others on the network. You didn't store anything for anyone else and the space taken up on your machine was exactly equal to the space taken up by

the specific files you wanted to store. The new generation of tools emulates a real networked file system. If you contribute a file to the network, it may get dispersed across a dozen computers. There may not even be a copy on your local machine. GnuNet takes this even farther by having the least requested files get duplicated in fewer and more remote files. They can, in fact, be automatically recycled (deleted) if they are not requested for long enough.

1.7.5.2. Indirect referencing

FreeNet actually uses a hierarchical "file system" so you can structure your documents. For example, if you (illegally) contributed a variety of MP3 songs by pop artist Britney Spears, you might arrange them in groups like /pop/britney/album1/song1, /pop/britney/album1/song2, etc. You can also use indirect references, so /pop/britney/album1/song1 could just be a pointer to the actual binary. This is important because files in FreeNet are located by their hash key; so if you changed the binary of "song1" (say you re-ripped it at better quality), users would no longer be able to find it unless there was a pointer to it. [Clarke, et. al., "Protecting Free Expression Online with Freenet"]

1.7.5.3. Search issues

Searching is a major area of research for these clients. If everything is encrypted and spread out all over the world, and no one person is supposed to know where it is, how can anyone ever find anything? Freenet answers this by using the hierarchical method described above (known as a "Signed Subspace" because it is digitally signed by the (anonymous) user who maintains it.) GnuNet resolved the problem by allowing you to attach encrypted keywords to files you contribute. This makes it easier to find what you're looking for, but it opens up the possibility of known-plaintext attacks against the encryption scheme. [Bennett, et. al., "Efficient Sharing of Encrypted Data"] Presumably, the authors aren't particularly worried about record companies having a lot of skilled cryptanalysts on hand.

1.7.6. Traffic Shaping

Traffic shaping is another security technique being co-opted by the hackers. In the case of GnuNet, data is broken down into 1K packets. [Bennett, et. al., "GNET"] Real data transfers are then interleaved with fake data (complete with fake routes that will lead analysts awry) and the client pumps out a continuous stream of

uniform data. It is virtually impossible to tell when files are being transferred or to where.

1.7.7. WebProxies

There are also tools to allow Web sites to be shared over the file sharing networks. At this point these are limited to relatively static Web sites, but they do offer users a crude ability to "surf the Web" without sacrificing the anonymity of the file sharing network. Sharing files from the peer-to-peer network onto the WWW would obviously be counterproductive.

2. Encryption

Encryption is one of the keys (no pun intended) to the workings of these networks. Encryption is used to evade content monitoring; to verify binaries; to prove identity; and to generate search terms. The precis of the GnuNet authors' paper on encryption indicates how key the technology is to their work:

"This paper describes the encryption of content for the file-sharing layer of GNUnet. We describe a new technique to encode content such that it can be easily distributed, searched for and retrieved. The encryption scheme allows users to insert the same content under multiple keys; yet multiple keys lead to practically identical copies in the system, reducing storage requirements. Keys can be chosen from natural language and can be combined to boolean queries. Queries and content can not be decrypted by intermediaries without guessing the key. The encoding of the content produces many small GBlocks, which can be easily distributed over several hosts. This allows the network to balance load. Single hosts are never hit with requests that take a long time to process." ["Efficient Sharing of Encrypted Data", Krista Bennett, Christian Grothoff, Tzvetan Horozov and Ioana Patrascu, ACISP 2002]

3. File Splitting

As mentioned previously, file splitting has a variety of uses, all aimed clearly at the goals of the designers mentioned above. One of the major reasons for splitting the files is to preserve deniability. Can a court of law find you guilty of possessing illegal copies of copyrighted material if your computer has one file which, when decrypted with a key you don't have, is a single chunk of a copyrighted file and useless without the rest of the chunks? This, obviously, has yet to be tested in court.

Another reason for file splitting is load balancing. Anyone who has used a file sharing network has experienced the frustration of finding the one particular file you want, only to discover it's stored on the computer of a user who is connected at 56K. File splitting increases the chances that you will never have to download more than a small part of the requested file from a slow

connection. Of course, there is always the chance that every chunk of the file will end up on slow nodes; see the next section on Economy for ways that GnuNet overcomes this.

Finally, file splitting aids in anonymizing the traffic. If, for example, an employee of a record label logged on to the network and downloaded a song from one of their artists, it would be more difficult to determine who was providing the illegal content since they would need to trace down the sources of numerous different chunks rather than just a single session.

4. Economy
The concept of economy is one of the more interesting aspects of GnuNet. In essence, what happens is that a node that does a better job of supplying a good selection of quality files at high speeds will get preference when downloading files from other nodes. Also, files which are requested more often get duplicated in more places and on faster nodes. Thus, a popular files is likely to take less time to download than an unpopular one because it is more likely to be located close to you.

   The point of an economic system (in which nodes automatically rate all other nodes they are in contact with) is to minimize the damage a freeloader or attacker can do. Since nodes which are well behaved get priority, an antagonistic node can only do damage in proportion to the amount of surplus resources that are made available to it. If the node does manage to begin to increase network utilization, that node will shortly find that it no longer is allowed to participate, as all available bandwidth will be allocated to nodes with better standing. These concepts are explained in much greater detail in Grothoff's "An Excess Based Economy".

5. Traffic Shaping
Traffic shaping has a variety of uses in data communications. Traffic shaping essentially means dividing up the data that is being transmitted and, if necessary, adding extraneous data so that the amount of data stays consistent over given time periods rather than having bursts and lulls at different times.

   For security purposes, this is used to confuse attackers who are watching for patterns. For example, if a VPN connection between a retail store and its corporate headquarters shows heavy traffic every night just after 9:00 p.m. but light usage at all other times, an attacker could surmise that daily sales information is being transmitted at that time. By limiting their decryption efforts to just the data being sent at that time, they greatly increase their chances of being able to break the VPN's encryption.

   For economic purposes, traffic shaping is used to ensure that an organization uses its allocated bandwidth most efficiently. Technically, what network

engineers do is try to use the full bandwidth of their committed information rate (CIR) without going over that rate and incurring additional charges. In some cases, low priority information like an email may be delayed if more important information is using a large percentage of the network. The email will then be sent when the level of utilization has dropped.

For file sharing networks, anonymity is more important than pure speed. For this reason, traffic shaping is used to prevent network administrators from noticing bursts indicating possible downloads of unacceptable material. A user who attempts to download the a video file of a feature film will certainly cause a spike in network utilization, but if they spread the download over many hours, it will just look like they are doing regular online work.

In order to minimize the waste caused by traffic shaping (since additional packets sometimes need to be transmitted to fill the quota for a particular time period), the file sharing tools use these packets to transmit fake queries, redirects and other misinformation that will theoretically cause forensic analysts massive headaches if they try to determine what a user is really doing.

6. Legitimate Uses

These file sharing tools seem very antagonistic to law abiding citizens and IT organizations. One IT veteran commented that the techniques in these tools were the equivalent of techniques he had learned in the military for obfuscating communications. However, there are legitimate uses for these tools.

One is as a resource efficient distributed file server. These tools could be used to build a storage network that did not rely on central servers and massive disk arrays, but rather on the excess capacity that is available in most of today's desktop machines. It has shortcomings – no access control, for example, but it could theoretically replace a large part of many corporate intranets. It could have some other advantages, too, such as built in encryption and the economic model that ensured that the most frequently used files were available the fastest.

Another legitimate use would be for intelligence agencies. These tools could be useful for foreign service agents who become entirely reliant on networks controlled by a possible adversary. They may not be as secure as, say, a direct phone call placed through a military satellite with high grade encryption, but they have the advantage of allowing your communications to go unnoticed among regular traffic.

The Internet community at large may also want to start using these just as a way to bypass the World Wide Web – allowing users to share content of their

own creation without paying for Web hosting or suffering from the incessant banner ads of the "free" Web site providers.

Finally, some would argue that it is legitimate for those who live in areas where human rights are scarce and censorship is prevalent to use these tools to get uncensored information and news. For example, few Americans would have had a moral problem with a woman who lived in Afghanistan under the Taliban accessing information because she wanted to learn something but was forbidden to go to school. However, those same people might take a more disapproving stance of users in France using these tools to access information on Nazism. The question of whether the use of these tools is ever morally correct even though it is illegal needs to be answered by the philosophers and theologians rather than by the technologists.

Conclusion

Excessive file sharing can have serious effects on a variety of organizations. These can range from lost revenue to lost productivity and wasted resources. Napster demonstrated that there is a huge demand for this material and it will only increase as more feature films appear online. IT security professionals need to know what the risks are and what techniques the authors of files sharing tools are using. Downloading a few ripped DVDs can bring your network down just as quickly as a targeted denial of service attack and downloading copyrighted materials is theft of proprietary information, whether or not someone actually hacked into your servers to get it.

Bibliography

Bennett, Krista; Grothoff, Christian; Horozov, Tzvetan; and Patrascu, Ioana. "Efficient Sharing of Encrypted Data", Australasian Conference on Information Security and Privacy, 2002. http://www.ovmj.org/GNUnet/download/esed.ps (Nov 11, 2002)

Bennett, Krista; Grothoff, Christian; Horozov, Tzvetan; Patrascu, Ioana; and Stef, Tiberius. "Gnet". http://www.ovmj.org/GNUnet/download/main.pdf (Nov 11, 2002)

Clarke, Iain; Miller, Scott G.; Hong, Theodore W.; Sandberg, Oskar; and Wiley, Brandon. "Protecting Free Expression Online with Freenet", IEEE Internet Computing 6(1), 40-49 (2002)

Cunningham, Wayne. "ICQ Forever!", C|net Download.com, May 21, 2002. http://www.icq.com/press/cnet200.html (Nov 11, 2002)

DeJesus, Ian. "Technology Timeline", Napster Overview. http://www.personal.psu.edu/users/j/i/jid102/assig7b.html (Nov 11, 2002)

Einstein, David. "First Movers and Creators of the Space", San Francisco Chronicle, Feb 1998. http://www.icq.com/company/about.html (Nov 11, 2002)

Freenetproject.org. "Download FreeNet 0.5", the free network project. http://freenetproject.org/cgi-bin/twiki/view/Main/Download (Nov 11, 2002)

Grothoff, Christian. "An Excess Based Economy". http://www.ovmj.org/GNUnet/download/ebe.ps (Nov 11, 2002)

Kingsbury, Peter. "The Ethic of broadband". http://matrixcubed.cjb.net/writings/ethicofbroadband.html (Nov 11, 2002)

McCullagh, Declan. "Could Hollywood Hack Your PC?", news.com, Jul 23, 2002. http://news.com.com/2100-1023-945923.html?tag=fd_lede (Nov 11, 2002)

Ovmj.org. "GNUNet – Download". http://www.ovmj.org/GNUnet/download.php3 (Nov 11, 2002)

SearchNetworking.com. "peer-to-peer – a searchNetworking definition", Aug 12, 2001. http://searchnetworking.techtarget.com/sDefinition/0,,sid7_gci212769,00.html (Nov 11, 2002)

Last Updated: August 20th, 2018

Upcoming SANS Training

Click here to view a list of all SANS Courses

SANS Copenhagen August 2018	Copenhagen, DK	Aug 27, 2018 - Sep 01, 2018	Live Event
SANS SEC504 @ Bangalore 2018	Bangalore, IN	Aug 27, 2018 - Sep 01, 2018	Live Event
SANS Tokyo Autumn 2018	Tokyo, JP	Sep 03, 2018 - Sep 15, 2018	Live Event
SANS Amsterdam September 2018	Amsterdam, NL	Sep 03, 2018 - Sep 08, 2018	Live Event
SANS Wellington 2018	Wellington, NZ	Sep 03, 2018 - Sep 08, 2018	Live Event
SANS MGT516 Beta One 2018	Arlington, VAUS	Sep 04, 2018 - Sep 08, 2018	Live Event
SANS Tampa-Clearwater 2018	Tampa, FLUS	Sep 04, 2018 - Sep 09, 2018	Live Event
Threat Hunting & Incident Response Summit & Training 2018	New Orleans, LAUS	Sep 06, 2018 - Sep 13, 2018	Live Event
SANS Baltimore Fall 2018	Baltimore, MDUS	Sep 08, 2018 - Sep 15, 2018	Live Event
SANS Alaska Summit & Training 2018	Anchorage, AKUS	Sep 10, 2018 - Sep 15, 2018	Live Event
SANS Munich September 2018	Munich, DE	Sep 16, 2018 - Sep 22, 2018	Live Event
SANS London September 2018	London, GB	Sep 17, 2018 - Sep 22, 2018	Live Event
SANS Network Security 2018	Las Vegas, NVUS	Sep 23, 2018 - Sep 30, 2018	Live Event
SANS DFIR Prague Summit & Training 2018	Prague, CZ	Oct 01, 2018 - Oct 07, 2018	Live Event
Oil & Gas Cybersecurity Summit & Training 2018	Houston, TXUS	Oct 01, 2018 - Oct 06, 2018	Live Event
SANS Brussels October 2018	Brussels, BE	Oct 08, 2018 - Oct 13, 2018	Live Event
SANS Amsterdam October 2018	Amsterdam, NL	Oct 08, 2018 - Oct 13, 2018	Live Event
SANS Riyadh October 2018	Riyadh, SA	Oct 13, 2018 - Oct 18, 2018	Live Event
SANS Northern VA Fall- Tysons 2018	Tysons, VAUS	Oct 13, 2018 - Oct 20, 2018	Live Event
SANS October Singapore 2018	Singapore, SG	Oct 15, 2018 - Oct 27, 2018	Live Event
SANS London October 2018	London, GB	Oct 15, 2018 - Oct 20, 2018	Live Event
SANS Denver 2018	Denver, COUS	Oct 15, 2018 - Oct 20, 2018	Live Event
SANS Seattle Fall 2018	Seattle, WAUS	Oct 15, 2018 - Oct 20, 2018	Live Event
Secure DevOps Summit & Training 2018	Denver, COUS	Oct 22, 2018 - Oct 29, 2018	Live Event
SANS Houston 2018	Houston, TXUS	Oct 29, 2018 - Nov 03, 2018	Live Event
SANS Gulf Region 2018	Dubai, AE	Nov 03, 2018 - Nov 15, 2018	Live Event
SANS Sydney 2018	Sydney, AU	Nov 05, 2018 - Nov 17, 2018	Live Event
SANS Dallas Fall 2018	Dallas, TXUS	Nov 05, 2018 - Nov 10, 2018	Live Event
SANS London November 2018	London, GB	Nov 05, 2018 - Nov 10, 2018	Live Event
SANS DFIRCON Miami 2018	Miami, FLUS	Nov 05, 2018 - Nov 10, 2018	Live Event
Pen Test HackFest Summit & Training 2018	Bethesda, MDUS	Nov 12, 2018 - Nov 19, 2018	Live Event
SANS Osaka 2018	Osaka, JP	Nov 12, 2018 - Nov 17, 2018	Live Event
SANS San Francisco Summer 2018	OnlineCAUS	Aug 26, 2018 - Aug 31, 2018	Live Event
SANS OnDemand	Books & MP3s OnlyUS	Anytime	Self Paced