The Evolution of the Peer-to-Peer File Sharing Industry and the Security Risks for Users

M. Eric Johnson, Dan McGuire, Nicholas D. Willey
Center for Digital Strategies
Tuck School of Business
Dartmouth College, Hanover NH 03755
{M.Eric.Johnson}@darmouth.edu

Abstract. Peer-to-peer file sharing is a growing security risk for firms and individuals. Users who participate in these networks to share music, pictures, and video are subject to many security risks including inadvertent publishing of private information, exposure to viruses and worms, and the consequences of spyware. In this paper, we examine the peer-to-peer file sharing phenomena, including an overview of the industry, its business models, and evolution. We describe the information security risks users' face including personal identification disclosure and leakage of proprietary business information. We illustrate those risks through honey-pot experiments and discuss how peer-to-peer industry dynamics are contributing to the security problem.[*]

Keywords: Security, Peer-to-peer, File sharing.

1 Introduction

Peer-to-peer (P2P) file sharing has seen both tremendous popularity and seemingly endless controversy. For many, P2P software clients have become part of the standard suite of PC applications. With millions of users world-wide sharing music, video, software, and pictures, file movement on these networks represent a significant percentage of internet traffic. Beyond the much discussed copyright infringement issues, P2P networks threaten both corporate and individual security. Our research shows that confidential and potentially damaging documents have made their way onto these networks and continue to do so. The research also shows that criminals trawl P2P networks and opportunistically exploit information that they find [20,21].

P2P file sharing represents a growing security threat because of the evolution of these networks. Internet service providers (ISPs), firms, and copyright holders have responded to the rise of P2P both technically (site blocking, traffic filtering and content poisoning) and legally. These challenges have prompted P2P developers to create decentralized, encrypted, anonymous networks that can punch through corporate and residential firewalls. These networks are almost impossible to track, are designed to accommodate large numbers of clients, and are capable of transferring vast amounts of data illicitly. This infrastructure, along with the computer science concepts underlying it, is well documented, easily available, and continues to be enhanced. This existing body of work represents a pre-made foundation that could be exploited for more malicious purposes.

In this paper, we provide an overview of the file-sharing "industry," discussing the popular software clients, business models, and the user base. We discuss the evolution of the industry and analyze the P2P security issues, establishing the vulnerabilities these software clients represent. Then we present experimental evidence of the risk through honey-pot experiments that expose personal financial information and track the resulting consequences. This analysis and experimental results clearly show the security risk of P2P file sharing networks. Finally, we conclude with observations about the industry, the policy and legal pressures, and future security issues.

2 Peer-to-Peer File Sharing

Peer-to-peer file-sharing networks enable users to "publish" or "share" files – any file from music to video to spreadsheets. P2P networks provide a ready-made sharing infrastructure that is difficult to block and even harder to track, providing cover for espionage and criminal activity. They encourage users to leave their computers on and connected to the internet at all times, running software that heavily uses their network, disk, and processor. Recent legal battles being won by the content industry (RIAA/MPAA) seem to have done little to really reduce file sharing, but have rather pushed users onto new clients and networks that are even harder track.

[*] We are grateful for the assistance of Tiversa Inc and Scott Dynes of the Center for Digital Strategies at the Tuck School. Experiments described in this paper were conducted in collaboration with Tiversa who has developed a patent pending technology that, in real-time, monitors global P2P file sharing networks. This work was supported under Award number 2000-DT-CX-K001 from the Office for Domestic Preparedness, Department of Homeland Security. Points of view in this document are those of the authors and do not necessarily represent the official position of the U.S. Department of Homeland Security.

2.1 A New Network on the Internet: The P2P Overlay

Peer-to-peer file sharing came of age during the dot.com boom and the rise of Napster. Between its debut in 1999 and its eventual failure in 2001, Napster enabled tens of millions of users to easily share MP3-formatted song files with each other. However, Napster was not a pure peer-to-peer network since it maintained lists on its central servers of connected systems and the files they had available to share. While the movement of M3P files was conducted directly between user computers, Napster controlled the search process. It was the maintenance of these centralized lists that made the system very easy and effective and which became the key legal issue that led to a court injunction against the company for copyright violation. After being shut down by court order, Napster reemerged as a small shadow of itself, legally selling songs. Its success and failure paved the way for many new P2P file-sharing networks such as Gnutella, FastTrack, e-donkey, and Bittorrent, with related software clients such as Limewire, KaZaA, Morpheus, eMule, and BearShare. This next breed of sharing systems has proven far more difficult to control and a much larger security threat.

The concept of peer-to-peer networking actually extends back almost four decades to the U.S. Department of Defense development of ARPANET. TCP/IP, introduced in 1973, cemented the notion of direct host-to-host communication with the network handling the mechanics of guiding the packets to their destination. Most of the protocols created since then (HTTP, SMTP, DNS, etc.) build on the idea that a host that needs data connects directly to the host that has it, and that it is the network's task to enable this. The techniques used by P2P file-sharing networking systems are simply an evolution of these principals.

Directory services, such as the P2P version of the Internet Domain Name Service system (DNS) that translates names to IP addresses, are provided by elected or designated hosts on the network. File transfer services are often provided using a tweaked version of the HTTP protocol underlying the World Wide Web. All of this inter-peer communication is accomplished with encrypted traffic that runs over the normal internet. These characteristics, along with other features, increase resiliency [43] and make the networks difficult to track.

2.2 Hiding Traffic

A number of internet service providers (ISPs) block or throttle traffic associated with P2P systems using a simple, fast approach known as port filtering. In response, P2P clients responded by using ports associated with other services (web traffic, email traffic, etc.) to exchange data. The P2P traffic then blends in with other traffic. Indeed, recent traffic studies suggest that P2P connections are now distributed across all ports with concentrations at a few preferred points [23].

Some providers use advanced filtering to search for P2P data. By examining the payload of each packet as it goes by, the ISP can determine the application that generated it and then impose bandwidth limits or even block it altogether. Unfortunately, the filtering heuristics tend to be fragile (false positives and false negatives can both impact quality of service), hard to maintain, and computationally expensive to implement. ISPs who decide to proceed anyway find themselves in an arms race. Some clients allow the user to encrypt P2P traffic, obscuring the signatures that the heuristics use to identify P2P data.

2.3 High Availability

Modern peer-to-peer networks are highly decentralized. Although the techniques used to support this vary from protocol to protocol, the approaches are generally the same:

- Replication. Popular files are spread widely across the network, so taking even a large portion of the network offline might do little to dent the number of files being distributed.
- Redundancy. Clients establish connections to many peers on the network. If one peer is removed, the client can fall back on the others. The other peers can even suggest alternatives to replace the lost peer.

2.4 Large Diverse Network Landscape

While Napster popularized P2P file sharing, its legal downfall spawned many new networks and clients. Today P2P traffic levels are still growing, but no single powerhouse application is driving it [23]. The aggregate numbers suggest that usage has more than doubled in the past three years, from less than 4 million to nearly ten million simultaneous users [28]. This does not include Bittorrent traffic, which is one of the most popular P2P applications for video and is more difficult to monitor. It also doesn't include users on private networks. Private networks, sometimes called dark networks (or darknets), are typically accessed through invitations from other users. Such networks, like the popular, OinkMe, may include millions of users.

Many users shift from network to network based on features and popularity. For example, the FastTrack network (used by KaZaA) has seen declines over the past three years while others like Gnutella have grown. Semi-successful attempts by content holders to disrupt access, coupled with KaZaA developers' efforts to increase revenue, quickly drove users to other networks, and even fostered the creation of new networks. This suggests low barriers to entry and also suggests that P2P networks serve a very mobile, well-informed user base that is willing to explore new alternatives as they arise.

3 Business Models

With the constant introduction of new file sharing systems, one might wonder what is driving the innovation. While there have been some astounding attempts to sell the computational services of the user network, the typical business models of the software client developers are fairly simple, either community-driven open source or advertising supported.

3.1 Open Source - LimeWire (gnutella network), eMule (eDonkey network),

Open source clients are developed with the same principals that govern development of the Linux kernel or the Apache web server. "We do this for fun and knowledge, not for money," say the developers of eMule[12]. Typically, one or two primary developers coordinate an army of contributors who do everything from submit new features to translate the program into new languages. The system works: eMule, for instance, is available in 30 languages, including Turkish and Basque. Funding comes from donations from users (usually in the $5 to $20 range) and from sales of merchandise such as shirts and message bags branded with the project logo. Some developers sell "Professional" versions of their clients for around $20, offering a few additional features and technical support.

3.2 Advertising Supported - eDonkey2000 (eDonkey network), KaZaA (FastTrack Network)

Firms with professional, full-time development teams have developed approaches that provide reoccurring revenue streams. Prospective KaZaA or eDonkey users have had the option of paying $20 to $30 to purchasing the application, or paying nothing to download a bundle generally containing the application, third party software that displays advertising on the user's computer ("adware")[11], other advertising-supported software, and trial versions of commercial software. Now-defunct MetaMachine (creator of eDonkey2000) managed campaigns for advertisers. Sharman Networks (creator of KaZaA) has outsourced its campaign management to Cydoor, an internet advertising firm[39]. Both firms directly negotiated deals with software partners.

Sharman Networks and MetaMachine have supplemented this revenue by harnessing their users' bandwidth to distribute DRM-protected audio, video, and software. Participating users hold the content on their own machine and serve it to other users as needed [25]. Users who volunteer their computers as servers are rewarded with "points" they can spend to get access to content. The content itself, along with the payment process, is managed by a third party, Altnet, a subsidiary of Brilliant Digital Entertainment [2]. Its customers, which include music from labels such as Artemis Records and software firms like Atari, pay distribution fees based on the number of downloads or the number of sales [5].

The advertising-supported business model has not proved a runaway success. The only publicly traded firm, Brilliant Digital Entertainment (BDE), was delisted from the American Stock Exchange in early 2004 and terminated the registration of its common stock altogether in April of 2006. Revenues in 2005 were $5.97MM, a 32% drop on the year before, and the firm is not profitable. Sharman Networks' financial situation is deliberately obscure. Officially, it is headquartered on the island nation of Vanuatu [49], a tax haven known for its corporate secrecy laws, with its employees provided on long term contract by an Australian firm. MetaMachine recently settled (September 13, 2006) copyright infringement lawsuits with RIAA agreeing to pay $30 million dollars to avoid further copyright infringement penalties. The also agreed to discontinue distribution of eDonkey2000. While appearing like just another in a string of victories for the music industry, the real impact on sharing and the eDonkey network is small because users had long ago migrated away from the eDonkey2000 client. Open source clients like eMule and Shareaza had already become the dominant clients (representing over 90% of the traffic on the eDonkey network).

Both Sharman and MetaMachine seem to be placing their hopes on the success of BDE's Altnet, but it may prove difficult to shift their business models to legitimate content distribution. Online content distribution is already a crowded market dominated by Apple's iTunes software. Further, as BDE notes in its SEC filings, major content producers that it would like to recruit as clients for its Altnet distribution system are among the firms suing it and its staff for their links with Sharman Networks.

Facing stiff competition from freely downloadable products with the same (or more) functionality and without adware, commercial firms don't seem to have a bright future. Independent developers have patched the KaZaA client to function without adware and are distributing the result free of charge - an effort that has continued despite significant legal pressure from Sharman. Also potentially hurting others looking to operate on advertising revenue is a new online service offering ad-supported, free, legal, and direct downloads. Universal Records has very recently partnered with the Spiral Frog service, a startup that hopes to grab users from P2P networks looking for free music. If individuals are willing to watch an advertisement while the music file downloads they are given the ability to download any song from the Universal catalog. These files contain Digital Rights Management (DRM) technologies that limit their use (no CD burning or file swapping for example), and the files expire after 6 months [38]. The service has not yet officially launched and thus its potential success is unclear.

3.3 Other Developing Models

Since the Internet Service Providers provide the bandwidth and therefore bear the cost of the heavy load that P2P subjects their networks to, many are interested in using their position as content distributors to bring in revenue. Currently traffic traded over P2P networks and even from legal downloading services does not result in any income for ISPs. An English company, PlayLouder MSP, has developed their network to support file swapping. The service is unique in that it allows users to legally swap copyrighted files. By signing licensing agreements with record labels such as Sony BMG, Playlouder allows all of its customers to download content belonging to its partners from any P2P network or other source in any form. For a price of £26.99 ($51.44), a PlayLouder customer receives a broadband internet connection and the right to download copyrighted content belonging to their partners without the fear of lawsuit [42]. However, there is a catch - PlayLouder will block all outgoing P2P traffic that is destined outside of its network [36].

3.4 Legitimate Downloads

The content industry has been trumpeting successes in converting users of P2P networks to legitimate downloads services such as Apple's iTunes music store, Real Rhapsody, and the reformed Napster. Certainly, the legal download market continues to grow. According to the RIAA digital signal, downloads increased by a whopping 163.3% from 2004 to 2005 for a total of 366.9 million downloads or monetarily, $363.3M. However, this is still far less than the CD market, which generated $10.52 billion in sales [37].

Apple's iTunes music store claims to have sold over 1 billion songs [19]. According to figures compiled from quarterly reports [3], by the third quarter of 2006, Apple had sold 58,672,000 iPods. The calculations indicate that Apple has sold approximating 17 songs for each iPod sold, or the equivalent of little more than a single full-length CD. Apple's marketing campaign for the original, now low capacity (5 GB), iPod touted its ability to put "1,000 songs in your pocket." [47] The 60 GB model now sold by Apple holds 15,000 songs. This leads us to question, "Where do the other 983 songs come from?" and "Why hasn't Apple sold 59 billion songs?" Clearly, MP3s ripped from previously purchased CDs filled mush of this void. However, recent studies have shown that, while owners of iPod are more likely to buy digital music than other those of other MP3 players, in general owners of all players (iPods included) don't buy much digital music [30]. In a related study of European users, free online music consumption significantly outweighs paid file-sharing networks. So makes of I3P players may owe much of their success to P2P (Figure 1).

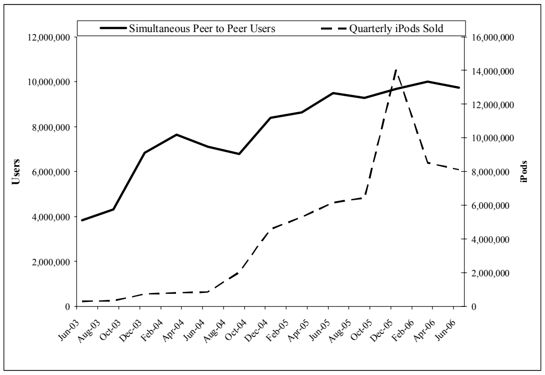


Fig. 1: Comparison of iPod sales growth with Big Champagne reported Simultaneous Global P2P Users

3.5 Content Creators Response

Content creators have responded with a barrage of lawsuits against individuals and firms, lobbying for new laws in Congress, and pressuring the World Trade Organization to clamp down on offending countries (e.g., Russia). They also attacked the networks themselves, by "poisoning" the content indexes. Even on a clean network, users looking for files will find many different versions of the same song or video. Each version may come from a different source (CD, radio, tape) or be recorded at different quality levels (high fidelity but large, low fidelity but small). Content creators use this to their advantage and hired firms to slip invalid versions of media files (invalid music files typically contain white noise, advertisements, or warnings about the illegality of file sharing) on to the network, which take their place among the legitimate files. The objective is to frustrate file sharers: the corrupted files take as long to download from the network as good files, and they hoped that the inconvenience of the process would encourage users to switch to more legitimate approaches [27]. Studies on the FastTrack network have found that as much as 50% of popular songs are polluted. In some cases, artists themselves joined the fight. Madonna employed a controversial scare tactic for the release of her album "An American Life." Users downloading what they thought were new songs were not so pleasantly interrogated by Madonna's voice questioning them, "What the f--- do you think you are doing?" [31] The P2P users responded by compiling a collection of remixes of the clip and marketing a CD known as "The Madonna Remix Project" [18]

Beyond parity and ridicule, the file sharing community has also responded with multiple solutions to combat file pollution. Advanced ratings systems have been developed that help users determine which versions of a file are high quality [10]. Regularly updated "block lists" of IP addresses employed by polluters have been distributed to file-sharing users [35]. By blocking results and requests from these IP addresses it is possible to block a large portion of the polluted content. In the end, much of the content creators' actions have seemed only to drive innovation in file sharing and shift users from one network to another.

4 P2P Demographic

While P2P may have once been exclusively for the technologically elite, P2P adoption is now widespread. One study found that 27% of adult Americans admit to sharing files from their computer with others [33]. Originally, some observers expected that lower income individuals would be more likely to be involved in file sharing, but this is not the case. In fact, income, race, and sex seem to play little role in determining whether an individual will engage in file sharing [34]. Age is by far the largest signal of an inclination to share: Students are almost twice as likely to share as non-students (35% vs. 18% respectively). It seems likely that a higher level of technological expertise and a stronger interest in music is driving this phenomenon.

An important predictor of file sharing is the individual's connection speed. Broadband users are more than 50% more likely to share than those on a dialup connection (30% to 19%). Therefore, as connection speeds continue to increase, it seems likely that more and more users will flock to P2P networks and file-sharing applications. The great magnitude of the problem may make it too difficult for content providers to eliminate. The lawsuits have been a successful deterrent to some extent, but the phenomenon still persists [4]. As the phenomenon perpetuates, it will only become more accepted.

Table 1. Pew Internet Activities Trends Report May-June 2005 [32]

Age Group	Percent Sharing
12 – 17	37 %
18 – 29	39 %
30 – 49	24 %
50 – 64	22 %
65+	14%

5 Traffic

P2P networks are dominated by what could arguably be called their "intended traffic" – a recent study across four major networks showed that audio and video represented more than 70% of the traffic generated by P2P [6]. (As with much of the research on activity within P2P, the exact estimates are noisy. For example, an analysis of traffic by researchers in Israel suggests that audio files represented 40% of all shared volume while video represented 52%). Given the large file size of video and audio, traffic estimates don't provide a clear picture of the number files shared of different types. Studies on usage patterns also paint an uneven picture of user willingness to both download and upload files. Estimates of users free-riding (downloading only) on P2P networks vary, but recent studies of Gnutella place the number of free riders at 15% of participants, a drop from 25% on 2002 [50]. In 2000, Adar and Huberman reported that 66% of users share no files at all and 73% share 10 files or less [1]. Similarly, a study in 2005 by Fessant *et al.* found that 68% of users on the eDonkey network were free riders [13].

By the end of 2004, over 60% of all web traffic was P2P-related [7]. Bittorrent alone accounted for 30% of all internet traffic [8]. At this time, the firm CacheLogic estimated that almost 10% of broadband users were involved in P2P trading at any given moment [9]. For content providers, stopping, observing, and regulating the river of data flowing through P2P is a monumental task. For the service providers, it is also a driving force behind home broadband adoption. Thus, they may be unenthusiastic about tampering with an unspoken selling point of their service – even if the heavy bandwidth P2P users drive up their cost of operation.

6 P2P Security Issues

Current P2P networks are designed "publish" or to "share" data. The user configures the client software to share items in a particular folder, and directs the client to move particular files and deposit them in that folder. In normal operation, a P2P client simply writes files to disk as it downloads them and reads files from disk as it uploads them. There are several routes for confidential data to get on to the network: 1) a user accidentally shares folders containing the information; 2) a user stores music and other data in the same folder that is shared; 3) a piece of software the user has chosen to download and execute surreptitiously shares it (malware); 4) client software bugs result in unintentional sharing of file directories

6.1 How Does Sensitive Information Become Exposed?

It is often not necessary for a worm or virus to expose personal or sensitive documents because many users will expose these documents unknowingly. Multiple reasons exist why users might expose personal or sensitive information:

- *Misplaced Files* – If a file is dropped accidentally into the wrong folder. Users may simply forget about the letter they wrote to the bank, or the documents from work they brought home one night. Similarly, teenagers using P2P may not know what Dad keeps on his Desktop.

- *Confusing Interface Design* – Users may be unaware of what folders are being shared or even that they are sharing files. For example, in a user study, Good and Krekelberg found that the KaZaA interface design contributed to user confusion over what files were being shared [15].

- *Incentives to Share a Large Number of Files* – Certain programs reward users for making files available or uploading more files. Some users may believe they can gain an advantage by sharing their entire hard drives.
- *General Laziness on the Part of the User* – If a user has a folder such as “My Documents” with many media folders inside, they may share My Documents rather than selecting each media folder individually to share, thus exposing all the other types of documents and folders contained within.
- *Wizards designed to determine media folders* – Some sharing clients come with wizards that scan an individual’s computer and recommend folders containing media to share. If there is an MP3 or image file in a folder with important documents, that entire folder could be exposed by such a wizard.
- *Poor Organization Habits* – Certain people may not take the time to organize their computers. MP3s, videos, letters, papers, passwords, and family pictures may all be kept in the same folder.

Many of these reasons point to the interface design and features of P2P clients that facilitate inadvertent sharing [44].

In many ways, the security risk of P2P clients is similar to Trojan horses, malware, and phishing scams: security breeches that depend on human intervention, abetted by a carelessness or lack of proper security education among users. The remedies are also similar: P2P network monitoring, user education, proper controls on corporate information, site blocking, periodic tests. We believe that the vast majority of information leaks are the result of such accidentally shared data rather than the result of malicious outsiders extracting data from an organization. However, there are many other trends that are driving more security concerns.

6.2 Growing Usage and Network Heterogeneity Means More Leaks

The amount of data being shared will likely continue to increase as P2P network usage grows. Assuming that current usage patterns persist, more and more confidential information will find its way on to these networks. Despite the significant positive network effects associated with using a particular P2P client (the larger the network, the more diverse the content, the greater the reliability, and the greater the speed), P2P networks are far more heterogeneous and faster moving than operating systems. With many networks and clients, users are not likely to grasp the security issues and P2P developers will likely not focus on security.

6.3 Set and Forget Increases Losses

The P2P usage model is very different than the standard internet model, and potentially reduces security. P2P traffic load, like most internet traffic, is cyclical. P2P traffic peaks at 10:00 p.m. EST. Significantly, studies have shown that peak traffic is only double the traffic at the slowest time of day. In comparison, peak web traffic levels tend to be five times greater than minimum web traffic levels. If we use web traffic as a proxy for a user’s presence at the keyboard, this indicates that P2P programs are being left running unattended. Indeed, research indicates that P2P clients tend to be “set and forget” applications that run in the background and while the user is not at the computer [14]. This suggests that the user is not carefully tracking the activities of the P2P client, increasing the opportunity for abuse. Further, even benign file sharing programs consume a good amount of processor time and network bandwidth, conditioning the P2P user to tolerate sluggish performance that, for others, might be a first sign that a system has been compromised. For typical internet applications, the vast majority of consumer traffic is downloads (web pages and email, complemented by the occasional outgoing message). Indeed, most ISPs have traditionally made the download pipe larger than the upload pipe (ADSL, cable modems). P2P generates significant uploading volume. Previously, significant outbound data on a consumer computer would be a signal to IT staff that a computer is compromised. Now it is par for the course, capable of cloaking a real compromise.

6.4 No Borders Result in Global Losses

Geography is largely irrelevant in P2P networks, meaning no particular country or region is safer than another. A computer logging on in Bombay or Brussels becomes part of the same network as a computer in Pittsburgh. Some studies, based on networking logs, have shown that hosts that are physically close to each other are only slightly more likely to be connected and sharing files than hosts that are physically distant. In addition, the network retains these characteristics if we dig into a particular region [26]. On the other hand, Fessant *et al.* argue that the opposite is true on the eDonkey network. They claim that “for 60% of the files, 80% of the replicas are located in the main country.” This supports some level of geographical clustering. Some of this may be explained by the popularity of content: “peers requesting a given video file may in a large proportion of cases download it from peers in their own country, thus achieving low latency and network usage compared to downloading it from a randomly chosen peer.” In any case, files certainly migrate globally and threats can come from any corner of the globe.

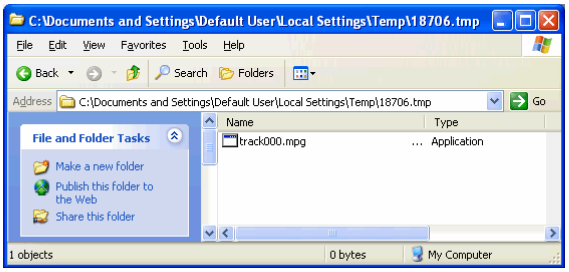
6.5 Digital Wind Spreads Files

Second generation P2P networks typically create file indexes using the names of files and metadata associated with them (the MS Word user who created it, for instance, or the company the product is registered to). As a result, malicious users are often searching opportunistically for files with key words or phrases, such as “credit card”.

Searching for "Wachovia" may turn up customers' records of their discussions with the bank (where "Wachovia" is a useful way to separate a bank conversation from a health insurance conversation). It also could snare Wachovia's internal documents because the bank name may appear in the company metadata tag of the file.

6.6 Malware

While the overwhelming majority of traffic on P2P networks is entertainment content (games, movies, music, etc.), also lurking on P2P networks are files that pose severe security risks [22,40]. Viruses and worms that exist in email and other programs also have variants that exist in peer-to-peer networks. A particularly severe virus known as Antinny, appeared on the Japanese-based Winny network that led to the disclosure of a large amount of private data including, U.S. military base security codes, and documents belonging to a police investigator involving a major investigation and 1,500 individuals [17]. Antinny propagates by making copies of itself, disguised to match filenames in a user's share folder or common names on the network (Figure 2). When another user downloads this file and opens it, they are greeted by a common error screen, indicating the file is unreadable. (Some variants even launch Windows Media Player.) In reality, the virus has been launched and opened a backdoor, often making every file on the user's computer available on the network. It may also take screenshots, note the user name, organization, and email address stored by Windows, and make changes to the registry. The file hides its true nature (an "exe" file extension) by cleverly inserting a large number of spaces after the name and the extension of the file it is masquerading as. If a user does not notice the ellipsis indicating a filename longer than can be displayed, they will likely launch the file [45].



C:\Documents and Settings\Default User\Local Settings\Temp\18706.tmp

Address: C:\Documents and Settings\Default User\Local Settings\Temp\18706.tmp

Name	Type
track000.mpg ...	Application

1 objects | 0 bytes | My Computer

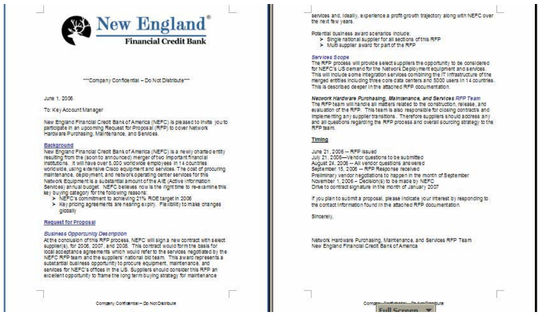
Fig. 2: Fake Winny file with long file name [46].

7 Experimental Results Illustrating Threat

To illustrate the threat, we ran a set of experiments in conjunction with Tiversa, Inc. We first posted the text of an email message containing an active VISA (debit) number and AT&T phone card in a music directory that was shared via Limewire[1]. The file was simply named "credit card and phone card numbers.doc" as a user who would title an email subject or file to reflect the message contents. With the help of a Tiversa, we observed both the activity of the file on our client and further tracked the file's movement across the P2P network. The file was quickly taken and retaken by a number of different clients. By the end of a week (1/10-1/17), the VISA card was used and balance depleted. We observed its use through the account's transactions statement posted by VISA on the web. Not knowing the exact balance of the card, the taker(s) used Paypal and Nochex (both processors of on-line payments) to drain funds from the card. It appears that two takers of the card were able to obtain funds as the activity was split into two groups and because one taker used Paypal, which is more US centric, while the other used Nochex, which is UK centric. Within another week the calling card was also depleted. Examining the call records of the card, all of the calls were made from outside of the US to two US area codes - 347 (Bronx, NY) and 253 (Tacoma, WA) clearly illustrating the P2P threat both within and outside of the US. Even more interesting, long after we stopped sharing the file, we observed the file continuing to move to new clients as some of the original takers leaked the file to others.

Next we developed an experiment that was more closely focused on the threat to firms. We created and shared three mock business documents. The first was a request for proposal (RFP) for a fictional bank that was looking for IT services to support the integration of a yet-to-be announced merger (Figure 3). Such a document represents strategic business information that could be valuable in many ways, including exploiting the information in illegal stock trades. The second was simply as a (publicly available) press release from Citibank announcing the completion of a merger. It would again represent business information that the takers might find valuable. The last was a draft of a fictional patent application for a new nanotechnology. This intellectual property is far more specialized, requiring a more sophisticated thief who could sell it to someone who could, in turn, exploit its value. Again, we placed the files in a music directory that was shared over a seven day period via Limewire. With the help of a Tiversa, our objective was to see both the file movement and the actions of those who took the file. We hypothesized that professional thieves who took the document would be careful not to share it while amateurs might take the documents and reshare.

Over the week, the two banking documents were taken twelve times – eight for the Citibank document and four for the fictional bank. The patent application was not taken during the week. We also observed that some of the takers immediately hid the



New England®
Financial Credit Bank

Company Confidential – Do Not Distribute

June 1, 2006

To: Key Account Manager

New England Financial Credit Bank of America (NEFC) is pleased to invite you to participate in an upcoming Request for Proposal (RFP) to cover Network Hardware Purchasing, Maintenance, and Services.

Background
New England Financial Credit Bank of America (NEFC) is a newly charted entity resulting from the (soon to announced) merger of two important financial institutions. It will have over 5,000 worldwide employees in 14 countries worldwide, using extensive Cisco equipment and services. The cost of procuring maintenance, deployment, and network operating center services for this Network Equipment is a substantial amount of the AIE (Active Information Services) annual budget. NEFC believes now is the right time to re-examine this key buying category for the following reasons:

- NEFC's commitment to achieving 21% ROE target in 2006
- Key pricing agreements are nearing expiry. Flexibility to make changes globally

Request for Proposal

Business Opportunity Description
At the conclusion of this RFP process, NEFC will sign a new contract with select supplier(s), for 2006, 2007, and 2008. This contract would form the basis for local acceptance agreements which would refer to the services negotiated by the NEFC RFP team and the suppliers' national bid team. This award represents a substantial business opportunity to procure equipment, maintenance, and services for NEFC's offices in the US. Suppliers should consider this RFP an excellent opportunity to frame the long term buying strategy for maintenance

Company Confidential – Do Not Distribute

services and, ideally, experience a profit growth trajectory along with NEFC over the next few years.

Potential business award scenarios include:

- Single national supplier for all sections of this RFP
- Multi supplier award for part of the RFP

Services Scope
The RFP process will provide select suppliers the opportunity to be considered for NEFC's US demand for the Network Deployment equipment and services. This will include some integration services combining the IT infrastructure of the merged entities including three core data centers and 5000 users in 14 countries. This is described deeper in the attached RFP documentation.

Network Hardware Purchasing, Maintenance, and Services *RFP Team*
The RFP team will handle all matters related to the construction, release, and evaluation of the RFP. This team is also responsible for closing contract/s and implementing any supplier transitions. Therefore suppliers should address any and all questions regarding the RFP process and overall sourcing strategy to the RFP team.

Timing

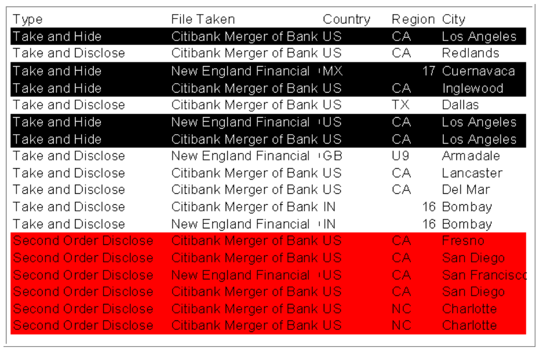
June 21, 2006 -- RFP issued
July 21, 2006—Vendor questions to be submitted
August 24, 2006 -- All vendor questions answered
September 15, 2006 -- RFP Response received
Preliminary vendor negotiations to happen in the month of September
November 1, 2006 – Decision(s) to be made by NEFC
Drive to contract signature in the month of January 2007

If you plan to submit a proposal, please indicate your interest by responding to the contact information found in the attached RFP documentation.

Sincerely,

Network Hardware Purchasing, Maintenance, and Services RFP Team
New England Financial Credit Bank of America

Fig. 3. Example Business Document.



Type	File Taken	Country	Region	City
Take and Hide	Citibank Merger of Bank	US	CA	Los Angeles
Take and Disclose	Citibank Merger of Bank	US	CA	Redlands
Take and Hide	New England Financial	MX	17	Cuernavaca
Take and Hide	Citibank Merger of Bank	US	CA	Inglewood
Take and Disclose	Citibank Merger of Bank	US	TX	Dallas
Take and Hide	New England Financial	US	CA	Los Angeles
Take and Hide	Citibank Merger of Bank	US	CA	Los Angeles
Take and Disclose	New England Financial	GB	U9	Armadale
Take and Disclose	Citibank Merger of Bank	US	CA	Lancaster
Take and Disclose	Citibank Merger of Bank	US	CA	Del Mar
Take and Disclose	Citibank Merger of Bank	IN	16	Bombay
Take and Disclose	New England Financial	IN	16	Bombay
Second Order Disclose	Citibank Merger of Bank	US	CA	Fresno
Second Order Disclose	Citibank Merger of Bank	US	CA	San Diego
Second Order Disclose	New England Financial	US	CA	San Francisc
Second Order Disclose	Citibank Merger of Bank	US	CA	San Diego
Second Order Disclose	Citibank Merger of Bank	US	NC	Charlotte
Second Order Disclose	Citibank Merger of Bank	US	NC	Charlotte

Fig. 4. Movement of Business Documents.

document after taking it – saving it into a directory that was not shared (Figure 4). Others continued to share the documents leading to another six secondary disclosures with Citibank drawing more interest, 5 to 1. Again, our experiment illustrated the risk of disclosure. Obviously, in this experiment, the risk appears much higher for financial documents than specific intellectual property like our patent application. While some of the takes may have taken the documents hoping to commit identity theft with personal consumer information, it appears likely that others were looking for business related documents. Whatever their motives, these business documents were taken and retaken. They also were taken by purposeful individuals who were quickly hiding their finds. As we discussed, many factors can drive the spread of files including the file naming conventions. A firm that has the unfortunate circumstance of sharing a name with a popular performer or song will experience far more activity. For example, the group Death Cab for Cutie recently recorded a popular song State Street Residential, which may increase the threat for documents from State Street Bank. While most takers looking for the song may have no malicious intent for the bank, the business files will spread, increasing the likelihood that they will be found by someone who is looking for them.

8 Evolving Threats Landscape

Due to the aggressive legal tactics used by the content industry against individuals, the next evolution in P2P will further protect, hide, and shield P2P users. These changes will likely create more security issues. Three such defenses are anonymization, trust networking, and distributed downloads.

8.1 Anonymization

Many of the newer clients provide some form of encryption to hide the identities and transmissions of their users. The Japanese Winny network is one of the earliest to do so. Today the most advanced such clients use multiple layers of encryption to hide the identity of their users. The I2P (Invisible Internet Protocol) network layer is used by a modified version of a Gnutella client known as I2Phex. The network hides both the identity of the sender and the recipient from one another and users are identified only by "cryptographic router keys." These keys are used to identify and produce a temporary virtual tunnel that data can be sent through to reach another user. A virtual tunnel employs multiple intermediary routers (other users), which only have knowledge of where the data is to be sent to next. The intermediaries do not know whether the next router in the tunnel is the intended recipient or another intermediary [48]. Any data that is transmitted over the network is done so with multilayered encryption making each kind of data sent over the network indistinguishable. In other words, email messages and MP3s sent over I2P look the same to outsiders; there is no way to tell what the content is without the keys [16]. As prosecution without physical evidence is difficult without knowledge of who sent and received the data, the intermediaries cannot be held liable because they were oblivious to what information they were sending.

8.2 F2F and Darknets

Friend to Friend (F2F) networking and Darknets rely upon secrecy and a web of trust. The idea is very simple, if users share files only with those that he or she trusts and knows then the user can be assured that he is not being monitored by prying eyes or entrapped. This is in stark contrast to older networks where joining the network was as easy as downloading and installing a client. With F2F networks one needs a willing insider to extend an invite. Currently, various implementations of the idea exist in many forms. The Firefox extension "AllPeers" is built similarly to an instant messaging client, allowing transfers to and from individuals on a "buddy list." The Soulseek network gives users the option of keeping a list of users allowed to download files while blocking other users. Other networks use the interconnectedness of social networks to share files. Instead of building a network of only your close friends, your friend may let you know about the files that his friends and his friends' friends have as well [41]. When files need to be transferred they could be handled in a similar method to the tunneling described earlier. Friends could be used to create anonymous connections between two friends who do not know each other. Of course this can be extended to multiple levels, and the result is a large well-connected network.

F2F networks fall under the category of darknets, but there are also other forms of darknets. For instance, there are many file-sharing websites appearing that are invite only. On these sites, people freely exchange media (often leaked content) using Bittorrent, and online file storage services. Gaining access to some networks may not be extremely difficult, but it serves as a hindrance to those that are curious about what is going on inside such sites. Examples of sites trading full length albums are: oink.me.uk, and www.indietorrents.com.

8.3 Distributed Downloads

Distributed downloads provide both protection and a more robust downloading experience. The idea is to split the file up into many parts from multiple destinations. This can allow for cooperation amongst downloaders and is the central idea behind Bittorrent: a small part of a file is useless by itself and is unplayable, requiring other parts and information on how to put the pieces together to reproduce the file. Some argue that sharing and downloading these files does not violate copyright because they are not readable on their own [29]. The Freenet network takes this idea to a higher level. Instead of sharing complete files or deciding what files to share from their computer, each user donates a portion of their hard drive for use by everyone on the Freenet network. Large files are broken up, encrypted, and stored in many places across the network. Even the individual hosting the files does not know what is on his or her hard disk. There is no central server or search directory to locate files on the Freenet network, the only way they can be found is if one user points another in the correct direction by providing him with a key. The key allows other users to find the pieces of the file, reassemble the file, and unencrypt it to view its contents. For an individual to find what is on the shared portion of his or her hard drive, he or she would need to find the key belonging to those files. It is difficult to hold someone responsible for sharing a small portion of a file that they know nothing about.

Taken together, the rapidly evolving P2P landscape seems likely to further perpetuate new and challenging security issues.

9 Conclusion

The popularity of peer-to-peer (P2P) file sharing has created many new security risks for individuals and organizations. In this paper, we have presented an analysis of the security vulnerability in P2P networks and provided accompanying evidence of the threat. There is little doubt that P2P presents a real security risk to both individuals and organizations. Certainly many individuals have been victims of identity theft as a result of their participation in these networks. Ironically, many of those victims may never realize the source of their misfortune. Rather than reducing the problem, we see many of the current trends further increasing the problem. While it is possible to use P2P sharing networks safely, the many evolving security threats mean that the best security advice is to avoid these networks altogether. In ongoing work, we are examining the implications for financial services firms. With thousands of employees, contractors, suppliers, and customers, spread over many countries, we believe large firms face significant risk from information leakage into P2P networks.

References

1. 33 Adar E, Huberman B. “Free Riding on Gnutella,” *First Monday Peer-Reviewed Journal on the Internet* [Online.] September 19, 2000. http://www.firstmonday.dk/issues/issue5_10/adar/index.html
2. 11. Altnet (Homepage) [Online] http://www.altnet.com/
3. 19 Apple – Investor Relations. [Online.] http://www.apple.com/investor/
4. 29 Blackburn D. “Online Piracy and Recorded Music Sales”. Harvard University. December, 2004. pdf available from: http://www.katallaxi.se/grejer/blackburn/blackburn_fs.pdf
5. 12. Brilliant Digital Entertainment, Inc. 10-KSB filed 12/31/05 [Online]. http://yahoo.brand.edgar-online.com/fetchFilingFrameset.aspx?dcn=0001170918-06-000350&Type=HTML
6. 31 CacheLogic Research: P2P in 2005. [Online]. http://www.cachelogic.com/home/pages/studies/filetype_02.php
7. 35 CacheLogic Research: P2P in 2005. [Online]. http://www.cachelogic.com/home/pages/studies/2005_07.php
8. 36 CacheLogic Research: P2P in 2005. [Online]. http://www.cachelogic.com/home/pages/studies/2005_06.php
9. 37 CacheLogic: Research: True Picture of P2P File sharing [Online]. http://www.cachelogic.com/home/pages/studies/2004_11.php
10. 25 Credence “Thwarting P2P Pollution:” **E**MİN **G**ÜN **S**İRER. http://www.cs.cornell.edu/People/egs/credence/
11. 8. eDonkey2000 (Advertise on Overnet and eDonkey2000) [Online]. http://www.edonkey2000.com/advertise.html
12. 7. eMule Project.net (Homepage) [Online]. http://www.emule-project.net/home/perl/general.cgi?l=1
13. 34 Fessant F., Handurukande S., Kermarrec A, Massoulie L.. “Clustering in Peer-to-Peer File Sharing Workloads,” Microsoft Research Cambirdge UK.
14. 40 Gerber, A., J. Houle, H. Nguyen, M. Roughan, and S. Sen. P2P The Gorilla in the Cable. In *National Cable & TelecommunicationsAssociation(NCTA) 2003 National Show*, Chicago, IL, June 8-11, 2003.
15. 38 Good N.S., and A. Krekelberg, “Usability and privacy: a study of Kazaa P2P file-sharing,” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Ft. Lauderdale, Florida, April 05-10, 2003.
16. 48 I2P, Intro http://www.i2p.net/how_intro
17. 44 Ingram, M. “66,000 Names and Personal Details Leak on P2P” April 29,2006. [Online.] http://www.slyck.com/news.php?story=1169
18. 24 Irixx, “The Madonna Remix Project” [Online]. http://www.irixx.org/madonna/madonnaremix.html
19. 18 iTunes Music Store Downloads Top One Billion Songs February 23, 2006. [Online.] http://www.apple.com/pr/library/2006/feb/23itms.html

20. 1. Johnson, M. Eric, McGuire, Dan, and Nicholas D. Willey, “Why File Sharing Networks Are Dangerous,” forthcoming in *Communications of the ACM*, 2007.

21. 2. Johnson, M. Eric and Scott Dynes, Inadvertent Disclosure: Information Leads in the Extended Enterprise, *Proceedings of the Sixth Workshop on the Economics of Information Security*, Carnegie Mellon University, June. 2007.

22. 42 Kalafut, A., A. Acharya, M. Gupta, “A Study of Malware in Peer-to-peer Networks,” *Proceedings of the Internet Measurement Conference*, ACM 2006.

23. 4. Karagiannis, T., A. Broido, N. Brownlee, K. Claffy, M. Faloutsos, “File sharing in the Internet: A characterization of P2P traffic in the backbone”, Technical Report, UC Riverside, 2003.

24. 5. Karagiannis, T. A. Broido, M. Faloutsos, and K. Claffy, 2004.Transport Layer Identification of P2P Traffic in *Proceedings of the 4th ACM SIGCOMM conference on Internet Measurement*. Taorima, Siciliy, Italy, 121-134.

25. 10. KaZaA (Download) [Online]. http://www.kazaa.com/us/products/downloadKMD.htm

26. 41 Leibowitz, N., Bergman, A., Ben-Shaul, R., and Shavit, A. Are File Swapping Networks Cacheable? Characterizing P2P Traffic, in *Proc. 7th Int. Workshop Web Content Caching and Distribution*, 2002.

27. 22 Liang, J. R. Kumar, Y. Xi, and K. Ross. Pollution in p2p file sharing systems. In *IEEE Infocom*, Miami, FL, USA, March 2005.

28. 6. Mennecke, T. “Slyck News – P2P Population Continues Climb” June 14, 2006. [Online]. http://www.slyck.com/news.php?story=1220

29. 50 Metcalfe, B. http://benmetcalfe.com/blog/index.php/2006/05/14/bit-torrent-legal-loophole/

30. 21 Mulligan, Mark. “Portable Media Player Owners,” Understanding iPod Owners' Music-Buying Habits, Juniper Research, September 14, 2006

31. 23 News.com, Hackers, Madonna mix it up. http://news.com.com/2100-1025-997856.html

32. 30 Pew Internet *Teen Content Creators and Consumers* (http://www.pewinternet.org/pdfs/PIP_Teens_Content_Creation.pdf) Nov. 2005

33. 27 Pew Internet Activities and Trends Report – June 05. Survey Question: Ever Share files from your own computer such as music, video, or picture files, or computer games with others online?

34. 28 Pew Internet Project Data Memo. [Online]. July 2003. http://www.pewinternet.org/pdfs/PIP_Copyright_Memo.pdf

35. 26 Phoenix Labs List http://test.blocklist.org/

36. 16 PlayLouder MSP FAQs [Online]. http://www.playloudermsp.com/faq.html

37. 17 RIAA, Recording Industry Association of America 2005 Year-End Statistics [Online] http://www.riaa.com/news/newsletter/pdf/2005yrEndStats.pdf

38. 14 Samson, T. SprialFrog’s “free” tunes are pricey. August, 30, 2006. [Online]. http://weblog.infoworld.com/techwatch/archives/007710.html

39. 9. Sharman Networks (Advertising) [Online]. http://www.sharmannetworks.com/content/view/full/251/

40. 43 Shin, S. J. Jung, H. Balakrishnan, “Malware Prevalence in the KaZaA File-Sharing Network,” *Proceedings of the Internet Measurement Conference*, ACM 2006.

41 49 Shirky, C. “File sharing goes Social” October 12, 2003

42. 15 Slyck News - Sony Frees Entire Music Catalog on PlayLouder ISP. [Online]. http://www.slyck.com/news.php?story=888

43. 3. Stutzbach, D. and R. Rejaie, “Understanding Churn in Peer-to-Peer Networks, *Proceedings of the Internet Measurement Conference*, ACM 2006.

44. 39 Sydnor II, T. D., J Knight, and L.A. Hollaar. “Filesharing Programs and Technological Features to Induce Users to Share,” A Report to the United States Patent and Trademark Office from the Office of International Relations, November, 2006.

45. 45 Symantec W32.Antinny.Q – Symantec.com http://www.symantec.com/security_response/writeup.jsp?docid=2004-053016-5101-99&tabid=2

46. 46 Symantec, Image from: http://www.symantec.com/content/en/us/global/images/writeups/w32.antinny.q.1.gif

47. 20 Tien. E. PULSE: P.S; 1,000 Songs in Your Pocket. Nov. 11,2001. The New York Times

48. 47 Tunnel Routing http://www.i2p.net/how_tunnelrouting

49. 13. Woody, T. “The Race to Kill Kazaa” in *Wired* Issue 2, Volume 11 (Feb. 2003). Available at http://www.wired.com/wired/archive/11.02/kazaa.html

50. 32 Zhao, S., D. Stutzbach, and R. Rejaie. Characterizing Files in the Modern Gnutella Network: A Measurement Study. In *Multimedia Computing and Networking*, (eds: S.Chandra and C Griwodz), 2006.