DOI: 10.1145/1461928.1461962

BY M. ERIC JOHNSON, DAN MCGUIRE, AND NICHOLAS D. WILLEY

Why File Sharing Networks Are Dangerous?

PEER-TO-PEER (P2P) SOFTWARE CLIENTS have become part of the standard suite of PC applications for many users. With millions of users worldwide sharing music, video, software, and pictures,[15] 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.

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 [1]) and legally. These challenges have prompted P2P developers to create decentralized, encrypted, anonymous networks that are difficult to track, are designed to accommodate large numbers of clients, and are capable of transferring vast amounts of data.

We 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 both business and personal financial information and track the resulting consequences. This analysis and experimental results clearly show the security risk of P2P file sharing networks.

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 to track.

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, 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.

A number of firms and 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, among others) 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.[8]

Today P2P traffic levels are still growing, but no single powerhouse application is driving it.[9] The aggregate numbers suggest that usage between 2003 and 2007 more than doubled, from less than 4 million to nearly ten million simultaneous users.[10] 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 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 (Figure 1). 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 for new file sharing systems and also suggests that P2P networks serve a very mobile, well-informed user base that is willing to explore new alternatives as they arise.

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.

P2P may have once been exclusively for the technologically elite, but today P2P adoption is widespread. One study found that 27% of adult Americans admit to sharing files from their computer with others.[11] Income, race, and sex seem to play little role in determining whether an individual will engage in file sharing.[12] Age is by far the largest signal of an inclination to share: Students are almost twice as likely to share as non-students.

P2P Security — How Does Sensitive Information Get Exposed?

Current P2P clients allow users to share items in a particular folder and often direct users to move files to 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: a user accidentally shares folders containing the information; a user stores music and other data in the same folder that is shared; a user downloads malware that, when executed, exposes files; or the client software has bugs that result in unintentional sharing of file directories. Of course it is not necessary for a worm or virus to expose personal or sensitive documents because many users will unknowingly expose these documents for many reasons:

▸ *Misplaced file.* If a file is dropped accidentally into the wrong folder.

▸ *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.[4]

▸ *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.

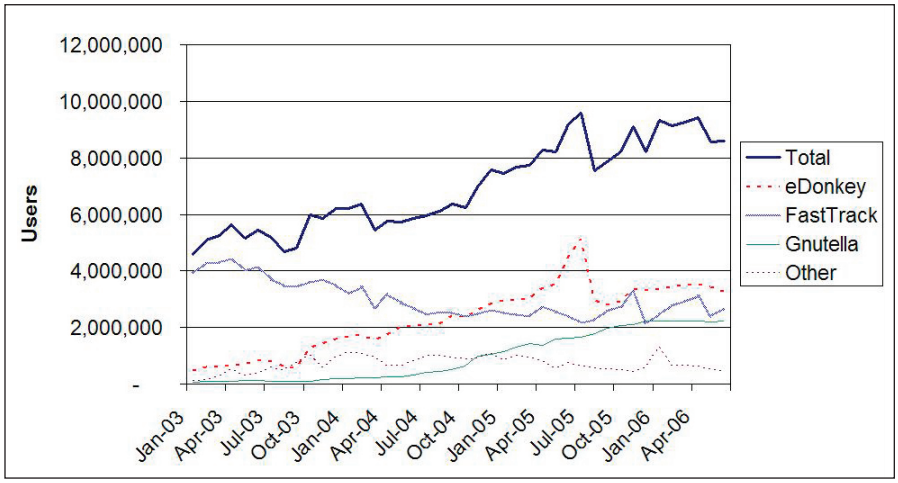
▸ *General laziness on the part of the user.* If a user has a folder such as "My Documents" with many media folders inside, they may share My Documents rather than selecting each media folder individually to share, thus exposing all the other types of documents and folders contained within.

▸ *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.

▸ *Unaware or forgetful of what is stored on the computer* and may simply forget about the letter they wrote to the bank, or the documents they brought home from work. Similarly, teenagers using P2P may not know what their parents keep on the Desktop.

▸ *Poor Organization Habits* – Certain people may not take the time to organize their files. MP3s, videos, letters, papers, passwords, and family pictures may all be kept in the same folder.

Figure 1. P2P Newtwork Usage



To illustrate the problem, we spent a couple hours searching the Gnutella network for sensitive personal documents; the resulting files we found should be disconcerting to users of P2P networks:

▸ *Birth Certificate* – 45 Results
▸ *Passport* – 42 Results
▸ *Tax Return* – 208 Results
▸ *FAFSA* – 114 Results

The Free Application for Federal Student Aid (FAFSA) and the U.S. Government's "EFILE" program both encourage individuals to complete forms online. When these forms are complete and full of potentially harmful information, applicants are asked to save a copy for their records. Similarly, those who are worried about credit scores often visit sites such as freecreditreport.com

and annualcreditreport.com which, after asking several questions, return the customer a pdf file with their credit history. These types of files leak out onto the P2P networks because of their inherent digital nature.

We downloaded a selection of these files and verified that they were indeed real. We observed one particular individual who was sharing a scanned copy of his passport. However, he did not only scan his passport, he also decided to scan his driver's license at the same time and include both in the same file. This information made him an easy target for anyone looking to commit identity theft. The passport and driver's license gave us two recent photographs of him, as well as his full name, address, date and place of birth, height, eye color, driver's license number, passport number, and two signatures. Furthermore, we were able to obtain his phone number and aerial photos of his house by using the gathered information in Google and Google Earth. Thieves are likely to download many more files from the individual's computer after finding such a document knowing that they have found much of the needed information to commit fraud.

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: user education, proper controls on corporate information, site blocking, periodic tests, and P2P network monitoring. We believe that the vast majority of information leaks are the result of 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.

Growing usage and network heterogeneity means more leaks. 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.

Set and forget increases losses. 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.[3] 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 significant 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.

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. As we will show, files certainly migrate globally and threats can come from any corner of the globe.

Digital wind spreads files. A firm that has the unfortunate circumstance of sharing a name with a popular performer or song will experience far more activity. Users looking for a media target may upload unrelated files with similar names thus spreading a file. 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 be found and spread, increasing the likelihood that they will be found by others. We call this "digital wind." Many factors can drive the spread of files including the file naming conventions. Moreover, 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 or the company the software is registered to). For example searching for a live performance from the Wachovia Center in Philadelphia 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.

Malware. While the overwhelming majority of traffic on P2P networks is entertainment content (games, movies, music, etc.), also lurking on P2P

Figure 2.

```
To: "Sara FranklíX" <sarakitten1X@hotmail.com>
From: "Joe FranklíX" <joefranklin197Xyahoo.com>
Subject: Grandma sent you stuff

Sara,

     Grandma sent you a $25 prepaid visa card and a telephone calling
card in the mail for Christmas. She didn't have your address there at
school. She said you better call her or else because now you don't have
any excuses. Here's the info from the cards:

Visa:

4436-9811-8709-XXXX expiration date is 03/07. That three digit number
that some places require is 636.

Phone card:

Here's what the back of the calling card says on how to use it:

TO PLACE A CALL FROM WITHIN THE US:

1. DIAL 1-800-471-1805
2. PRESS 1 FOR ENGLISH
3. ENTER PIN
4. PRESS 1 TO CALL WITHIN THE US PRESS 2 TO CALL ANY OTHER COUNTRY

The pin number is 557-696-XXXX. It has 210 minutes on it.

I know that you'll probably buy something online at bodyworks but if
not let me know and I'll drop them in the mail.

XOXOXOXO

Dad
```

networks are files that pose severe security risks.[7, 13] Viruses 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.[5, 14]

Experimental Results Illustrating Threat

With a clear understanding of the vulnerability, what about the threat? To illustrate the threat, we ran a set of experiments in conjunction with Tiversa, Inc. In our first experiment, we posted the text of an email message (Figure 2) containing an active VISA (debit) number and AT&T phone card in a music directory that was shared via Limewire. 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 (Figure 3). 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 online 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 U.S. to two U.S. 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 (Table 1).

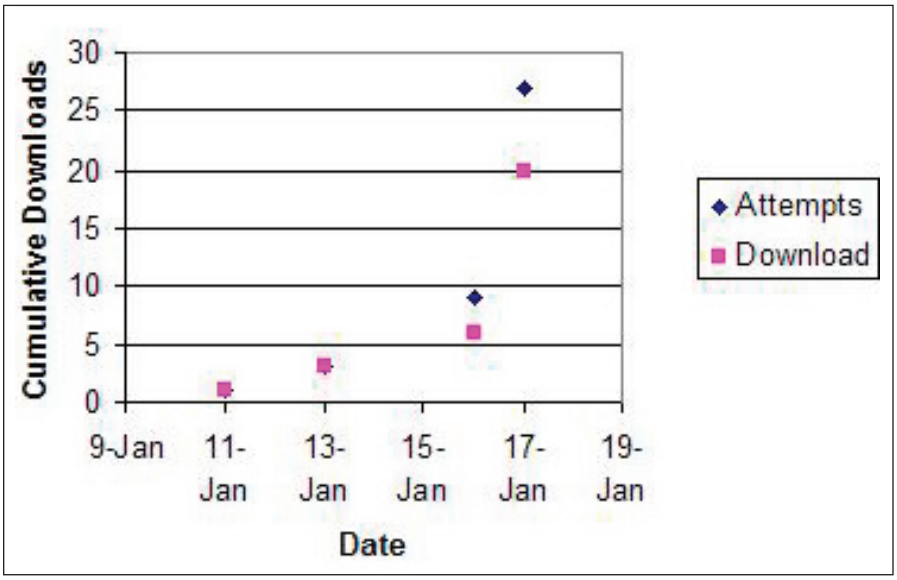
Table 1.

Hanover, NH	1/10	
Little Rock, CA	1/11	
Schenectady, NY	1/13	
Lincoln, NE	1/16	
Portland, ME	1/16	
Lancaster, CA	1/17	Stopped
Portland, ME	1/18	Sharing
Mexico	1/19	
Windsor, Canada	1/19	
UK	1/20	
Burbank, CA	1/21	
Little Rock, CA	1/21	
Singapore	1/22	
Sterling, VA	1/24	
Bakersfield, CA	1/25	
Germany	1/30	
Montreal, Canada	1/31	
Chattanooga, TN	2/1	
New Orleans, LA	2/2	

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. Such a document represents strategic business information that could be valuable in many ways, including the possibility of exploiting the information in stock trades. The second was simply as a (publicly available) press release from a major bank announcing the completion of a merger. It would again represent business information that the takers might think 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 12 times – eight for the major bank 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 document after taking it, saving it into a directory that was not shared. Others continued to

Figure 3.



share the documents leading to another six secondary disclosures.

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.

Conclusion

The popularity of peer-to-peer (P2P) file sharing has created many new security risks for individuals and organizations. In this article, 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 likely 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 evolving security threats mean that the best security advice for many users is to avoid these networks altogether. In most cases, firms are well advised to block P2P activity on their networks and devices.

However, P2P sharing can be a very effective mechanism for distributing files and collaborating with other users. We see several approaches to reduce security risk including:

User interface design. As discussed by Good and Krekelberg, interface design has a significant impact on security. Client developers should incorporate features that clearly show users what files are being shared and uploaded along with reducing the ease of sharing (or even blocking) nonmedia files. Visualizing system activity and integrating the client configuration into routine user action as suggested by de Paula et al.[2] would certainly improve security. However, as we noted earlier, given the business models of many P2P client developers it is not clear they currently have the incentives to improve the security of their interfaces. Thus users must beware and select appropriate clients. Likewise, firms should consider steps to improve user visibility of security gaps.

User education. Understanding the risks is a key step in reducing exposure. Firms should ensure employees, contractors, suppliers, and customers understand the risks.

File naming and organization. Firms and users should also introduce file naming conventions and policies to reduce the "footprint" of their documents. These types of initiatives reduce the threat of documents being found and spread. Folder organization to segregate files types is also important. For many firms, steps to block P2P participation on firm equipment along with policies for storing data on home machines are often warranted.

In related work, we are examining the implications for financial services[6] and health care 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. C

This work was supported under award number 2000-DT-CX-K001 from the Office of 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.

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 article were conducted in collaboration with Tiversa who has developed a patent-pending technology that monitors global P2P file-sharing networks in real-time.

References

1. Christin, N., Weigend , A. S., Chuang, J. Content availability, pollution and poisoning in file sharing peer-to-peer networks. *Proceedings of the 6th ACM Conference on Electronic commerce*, Vancouver, BC, June 05-08, 2005, 68-77.
2. de Paula, R., Ding, X., Dourish, P., Nies, K., Pillet, B., Redmiles, D. F., Ren, J., Rode, J. A., Filho, R. S. In the eye of the beholder: a visualization-based approach to information system security. *International Journal of Human-Computer Studies*, 63, July 2005, 5-24.
3. Gerber, A., Houle, J., Nguyen, H., Roughan, M., and Sen, S. P2P The Gorilla in the Cable. *National Cable & Telecommunications Association (NCTA) 2003 National Show*, Chicago, IL, June 8-11, 2003.
4. Good, N. S., and Krekelberg, A. 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.)
5. Ingram, M. 66,000 Names and Personal Details Leak on P2P. April 29,2006; http://www.slyck.com/news.php?story=1169
6. Johnson M.E. Information risk of inadvertent disclosure: An analysis of file-sharing risk in the financial supply chain. *J. MIS 25*, 2 (2008), 97-123.
7. Kalafut, A., Acharya, A., Gupta, M. A Study of Malware in Peer-to-peer Networks. *Proceedings of the Internet Measurement Conference*, ACM 2006.
8. Karagiannis, T., Broido, A., Brownlee, N., Claffy, K., Faloutsos, M. File sharing in the Internet: A characterization of P2P traffic in the backbone. Technical Report, UC Riverside, 2003.
9. Karagiannis, T., Broido, A., Faloutsos, M., and Claffy, K. 2004.Transport Layer Identification of P2P Traffic. *Proceedings of the 4th ACM SIGCOMM conference on Internet Measurement*. (Taorima, Siciliy, Italy), 121-134.
10. Mennecke, T. Slyck News – P2P Population Continues Climb. (June 14, 2006.) http://www.slyck.com/news.php?story=1220
11. Pew Internet Activities and Trends Report. Survey Question: Ever Share files from your own computer such as music, video, or picture files, or computer games with others online? June 2005.
12. Pew Internet Project Data Memo. July 2003; http://www.pewinternet.org/pdfs/PIP_Copyright_Memo.pdf
13. Shin, S., Jung, J., Balakrishnan, H. Malware prevalence in the KaZaA file-sharing network. *Proceedings of the Internet Measurement Conference*, ACM 2006.
14. W32.Antinny.Q – Symantec.com; http://www.symantec.com/security_response/writeup.jsp?docid=2004-053016-5101-99&tabid=2
15. Zhao, S., Stutzbach, D., and Rejaie, R. Characterizing files in the modern gnutella network: A measurement study. Proc of In *Multimedia Computing and Networking*, (S.Chandra and C Griwodz, Eds), 2006.

M. Eric Johnson (M.Eric.Johnson@darmouth.edu) is Director of Tuck's Glassmeyer/McNamee Center for Digital Strategies and Professor of Operations Management at the Tuck School of Business, Dartmouth College.

Daniel McGuire was a Tuck MBA student at Dartmouth College during the time of this study. He is now working as an Associate in Booz Allen Hamilton's Operations practice.

Nicholas D. Willey is a Master of Engineering Management student at Thayer School of Engineering at Dartmouth College.