Why Share on Peer-to-Peer?

An Interactive Qualifying Project
submitted to the Faculty
of the

WORCESTER POLYTECHNIC INSTITUTE
in partial fulfillment of the requirements for the
Degree of Bachelor of Science
by

Robert J. Camera Jr.

Date: December 18, 2003

Professor Oleg V. Pavlov, Advisor of Record

Professor James Kevin Doyle, Co-Advisor

Abstract

Mp3 music files have been available for free download on peer-to-peer networks for many years. Recently, the *Recording Industry Association of America* has started tracking down individuals who share large amounts of files on these networks. The study consisted of an internet survey of 46 WPI students. It was found that individuals see little reason to share these music files and that the added threat of a lawsuit has stopped many of these individuals from sharing their music files.

Acknowledgements

I would like to thank Professor Pavlov and Professor Doyle for their support on this project, and Professor Weaver for helping me arrange th e project in the first place. Thanks also go out to Irma S., Andrew M. and Elisa B. for keeping me sane while working on the project.

Table of Contents

Table of Tables

Introduction

The expanding popularity and ever - increasing speed of the Internet has allo wed information of all kinds to flow at speeds never before imagined. At no previous point in time could one read the latest news headlines while checking up on real -time market data all while shopping around for the best deal for a particular item. In a ddition to these potential 'serious' types of uses for the Internet, this network of networks provides a limitless amount of entertainment in the form of games, chat rooms, and file swapping among others.

This study tries to find out how many college st udents at a distinguished technical institute download music, share music, and why.

Background

Peer-to-Peer

The popularity and complexity of file swapping has grown significantly since the days of anonymous free FTP servers where anyone in the world coul d log onto a machine with a Dynamic Name Server (DNS) lookup or static Internet Protocol (IP) address. In fact, the growth of the internet is the very cause of why anonymous FTP is now nearly obsolete. As more and more individuals connected their persona l computers to the Internet, the IP protocol could no longer support the original goal of 'one machine, one IP'. Instead, internet providers would assign a dynamic IP address to users as they would connect to the service. There was no guarantee that the computer's IP would be the same the next time it connected to the service. These dynamic IP addresses made it impossible for DNS to work properly for these 'part -time' connected machines since it could take days for a single DNS update to propagate throug h the system. Since DNS and static IP addresses could no longer guarantee a lookup for every computer on the Internet, a new method of connecting to a known machine was needed. This hole

created in DNS by dial-up and other part-time connected devices led to the eventual creation of Peer-to-Peer (P2P) networks.[1]

The purpose of P2P networks is to bypass DNS and allow real-time updating of the IP address of a specific client. That is, some sort of unique identifier will follow either a particular computer or a particular user and update the IP address real-time. ICQ (I Seek You) was the first well-known P2P protocol, released in November of 1996. With this protocol, a user was able to 'sign on' to ICQ from any machine and chat with other individuals based on their ICQ name, not their DNS lookup or IP address. Under the hood of ICQ, the 'sign on' process specifies the IP address of the computer used to sign on. This IP address is updated almost immediately within ICQ's DNS replacement. When another user wishes to chat with the first user, he will tell ICQ the name of the user he wishes to chat with, ICQ will look up the IP address, and forward the message on to the original user's IP address. Signing on to ICQ as a particular user is seamless across IP address and host.

Other common P2P applications include the more recent AOL Instant Messenger (AIM), Scour, Napster, and Gnutella. These latter two P2P protocols as well as newer versions of AIM and ICQ allow individuals to download files directly from another user's machine, generally from a specified directory. The idea behind Napster and Gnutella is that an individual can search the shared directory of other computers on the same network for a particular file. When a matching file is found, the user has the capability to download it from one or more of the matching computers. This file sharing is done without any form of DNS lookup, allowing a user or host machine to change locations seamlessly.

File Sharing

As mentioned earlier, the concept of file sharing has been around for quite a long time, reaching back to the days of FTP. In order to share files, an individual would have to run an ftp server and set the root of the server to a directory containing the shared files (generally /pub from the FTP root directory). In order to download a particular file, an FTP client would have to type the entire path to the desired file. This can be fairly

difficult if the exact path of the file is unknown and the FTP server has large amounts of directories and files to browse. Software such as Globescape's CuteFTP and Microsoft's Internet Explorer offer graphical representations of FTP. This method is still used today, but it is not particularly popular due to security issues involved with this protocol and the availability of new, more versatile methods of file transfer.

With ICQ, an individual can choose to allow others to access a specified directory while signed on to the service. The process can be viewed almost the same as an FTP server except that the FTP server application is replaced by the ICQ client. Individuals would still need to browse a directory in search of a particular file. This one application runs as both a server and a client.

Similar to ICQ, Napster runs as both a server and a client. Once again, a single directory is shared. Unlike any other file sharing protocol, a host name is not needed to find a file. Instead of connecting to a host and looking for a file, Napster allows the user to search for a file without specifying a host. The Napster service would query other clients signed on to the service and return the results to the user. With these results, the user can simply download a file without necessarily knowing the host. This effectively allows a user to search a large number of other hosts without knowing each host. Furthermore, the user does not need to browse directories or know the full path to the desired file on the host.

With the collapse of Napster, there was a large market for a similar service that did not rely on central servers. This is the concept behind the Gnutella protocol. The second sentence of the Gnutella protocol states that "*Although the Gnutella protocol supports a traditional client/centralized server search paradigm, Gnutella's distinction is its peer-to-peer, decentralized model*" [2]. That is, it allows for the same type of query as Napster but removes the central server aspect. This fact would make it near impossible to shut down the Gnutella network. Other than the lack of a central server, Gnutella acts quite similar to Napster. Some Gnutella clients take an extra step forward and allow the user to download segments of the same exact file from multiple clients simultaneously and re-assemble the file once the downloads are complete. This assures

the user that the file will be downloaded successfully even if one or more file hosts disallow the file to be downloaded or disconnect from the network.

The ability to download all types of digital files has caused heated debates and even copyright infringement lawsuits. The original Napster service was court-ordered to shut down in 2001. The Recording Industry Association of America (RIAA) launched lawsuits against at least 216 individuals openly sharing at least 1000 copyrighted files in 2003. Many of these individuals settled out of court and were required to destroy all copies of copyright infringing material. In addition to suing music file-swappers, companies have made 'spoofed' files available in place of the actual files [3]. The idea behind this is that an individual will download a 'spoofed' file in place of the file they actually wanted. However, this just means that the individual has to download from somewhere else to obtain the desired file, and 'spoofing' becomes nothing more than a minor inconvenience.

Why Share Files?

Using these P2P file-sharing methods is a common thing to do in the US. As of April, 2001, it is estimated that nearly 30 million American adult Internet users have downloaded music files, totaling to around 29% of adult internet users. The percentage of children between the ages 12 and 17 to download music files is around 53% of children.[4] However, with all this sharing going on, there seems to be a very small percentage of individuals who actually share files on these systems. Using these services to download material while offering no material in return is known as freeloading. Previous studies have determined that freeloaders are the predominant form of P2P file-swappers. These studies show that 70% of Gnutella users are indeed freeloaders in one case [5], while other studies find up to 98% of Gnutella users freeloading [cited by 6]. Although these numbers vary by a large amount, both agree on the fact that most Gnutella P2P users choose not to share files.

With these facts in mind, this paper will discuss reasons individuals at a high-tech institute *do* choose to make files (music files in particular) available for download. Doing such provides no direct benefit to the individual and in a small number of cases

can even prove to be detrimental. Sharing files potentially wastes a user's bandwidth and other computer resources. In some cases, file sharing can even lead to legal actions by the RIAA.

Methods

In order to study why individuals share music files on P2P networks, I decided to create a simple internet survey. The survey asks questions regarding music file sharing habits, as well as questions regarding familiarity with P2P current events.

Target Population

In order to receive the best results in the survey, the survey group should have computers and preferably high-speed, always-on internet access. Although most households in the US have a computer, most still do not have high-speed internet access. Without always-on and high-speed internet access, one is less likely to share music files. A particular group that is almost guaranteed to own a computer and high-speed, always-on internet access can be found in college students. Worcester Polytechnic Institute (WPI) was chosen as the target school due to convenience to the author.

In order to sample the college properly, I must make sure not to sample from a particular major. A representative sample of college students can be found in a broad, introductory course. Examples of this type of course in an engineering school would be calculus or economics. Calculus is a requirement for most students to take at some point in their college careers, but is usually taken early on. Economics is not a required class at WPI and is not often a prerequisite for more advanced engineering classes. Based on this, an economics class was chosen as the survey sample population.

The class consists of 26% freshmen, 34% sophomores, 17% juniors, 17% seniors, 2% transfers, and 2% 5th year student. Juniors and seniors are slightly under-represented and sophomores are slightly over-represented in the class. There are approximately

three males to one female. Broken down by major, the class roughly fits the breakdown of the school by major. That is to say, where *Mechanical Engineering*, *Electrical and Computer Engineering*, and *Computer Science* are the three most pursued majors at the school, each of these majors is within the top four most common majors in the class. The class has a disproportionately high amount of *Management & Information Systems* majors (11.6% class vs. 5.5% of school), but this should not greatly effect the results of the survey. Overall, the chosen sample group represents the overall population of the school closely enough for this study.

Survey Type

In order to avoid disrupting class time, I chose to conduct an internet survey. This survey can be completed in about twenty minutes at the student's convenience from any computer with internet access. This type of survey also guarantees that the entire class can participate in the survey, not only students who show up at a specific lecture. An internet survey also benefits the surveyor because the results are stored in a pre-determined format on a computer, and therefore do not need to be entered manually by hand. Potential drawbacks of this type of survey are that the surveyor is not present to ask any immediate questions, and technical issues such as non-standard-conforming browsers can arise.

A simple program was written by the author to parse a small text file indicating the contents and format of the survey. With the information from this file, the HTML form can easily be generated. The HTML form contains standard form widgets – radio buttons, check-boxes, text-boxes, and buttons for reset/submit.

An additional custom application parses the return of the survey and stores all relevant data into a unique file. The debriefing indicates a unique identifier, which the individual must send to the professor of the target class to indicate their participation in the study. Individuals in the class who successfully complete the survey are given bonus points in the class. Participation in the survey is completely voluntary – there is no penalty to students who choose not to participate.

Once the survey is closed to the students, the unique survey result files are collected and compared to the unique identification numbers sent to the professor. Any survey not matching a student's unique ID is disregarded. Student IDs that do not match a survey are not given extra credit. This method assures us that each student completed a maximum of one survey, and it guarantees us that only students' survey results will be tallied. The valid survey files are merged into a single comma separated value file, a common file-type that allows our data to be transferred into many statistical analyzing software packages. Once the data has been brought into one of these packages, we are free to evaluate the final results of the study.

Survey Questions

The survey consists of questions regarding file swapping behavior, and is broken down into five distinct parts; *Downloading Behavior*, *Sharing Behavior*, *Sharing Knowledge*, *Personal Information*, and *Survey Feedback*. Each section answers different questions vital to the success of the survey. The full survey can be found below in Appendix B - The Survey.

Downloading Behavior

This section of the survey is intended to gain a feel for the individual's music-downloading habits. Not only does this section ask about how many music files the individual owns, but what file types they own as well. Since downloading from a P2P network is not the only way to acquire music files on the Internet, there is also a question regarding purchasing music files.

The questions '*how*', '*why*', and '*what*' about downloading are also asked. How these files are acquired (what application), why they are acquired, and what is done with them. Since some music files may be acquired as a sample of a particular genre of music or a particular band, these files are often deleted once they have been sampled. A question regarding the length of time files are retained is also asked.

Sharing Behavior

This section is the primary focus of this study. In this part of the survey, we ask the individual about their sharing behavior on P2P networks . Although the section is relatively short, these questions will be able to answer most, if not all, of our questions about file sharing . In addition to simply asking how many files an individual shares at the present , a separate question regarding histo rical sharing is asked. These questions in conjunction will potentially show us a file -sharing trend.

The survey questions why the individual shares music files, or why they choose not to share files . These questions must be open -ended because each pers on has their own unique answer to these questions.

The question regarding why the student shares music files was taken nearly directly from a previous study by Terrel Galloway and Douglas Kinnear [7].

Sharing Knowledge

The thi rd section of the survey asks users about their knowledge of P2P related issues including the campus' Acceptable Use Policy (AUP), copyright law, and lawsuits against file sharers. The survey asks if the individual personally knows a defendant in a case b rought by the RIAA. Approximately one in 600,000 file sharers has had charges brought against them.

Information regarding the campus AUP and US copyright law is made available to the individual in the briefing.

Personal Information

The personal information section asks questions that tell the surveyor who is taking the survey. Age, gender, and field of study are asked , as well as internet connection speed. This section will show us how close our sample relates to the population of the school. Additi onal questions regarding computer knowledge will potentially indicate correlationary results regarding file swapping.

Survey Feedback

This section is only displayed in the pilot study. It asks a few questions regarding the ease of the survey and includes an open-ended question to allow suggestions to improve the survey.

Results

Once the methods are finished, it is time to collect results. Data collection must be completed in an unbiased manner to avoid altering the numerical findings of the survey.

Pilot Results

A small pilot study consisting of approximately six individuals ranging from college students to professors pointed out a few grammatical and spelling errors. Additional questions were added to assist in the creation of a P2P economic model for Prof. Pavlov. These questions include questions 10 and 30. These results were collected by personal contact with the individuals reviewing the survey. Pilot data regarding the main topic of discussion was not analyzed.

Data Collection

Survey data was collected from December 5, 2003 through December 8, 2003, 09:00 EST. There were seventy-nine surveys successfully submitted. Of these seventy-nine, only forty-six were accounted for by members of the target class. There were a few incomplete surveys. A potential reason for there being so many unclaimed survey submissions could be misunderstanding of the directions to submit the unique identifier to the professor of the class. The valid responses include over half of the class.

Findings

Downloading Behavior

Findings show that 91.3% of the surveyed students have downloaded mp3 or similar music files, while only 13.0% of the same students have purchased these music files online. 88.9% of the students claim to have owned mp3 files at the time of the survey. 44.4% owned *Windows Media* Audio (wma), 15.6% owned *Ogg* Vorvis (ogg), and 13.3% owned other formats of music. These other formats consisted of wav, *Mpeg4* (aac), *Fully Lossless Audio Content* (flac), and *Monkey's Audio* (ape).

The average amount of CDs owned by the students is 65.25, with a standard deviation of 108.29. This number varies between zero to as high as 600 for one student. The number of mp3 or similar music files owned by the students can be found in Table 1 below, as well as the number of expected music file downloads for the next week.

# Files Owned	Percent
0	10.87%
1 – 20	4.35%
21 – 50	4.35%
51 – 100	6.52%
101 – 250	15.22%
251 – 500	10.87%
501 – 1000	10.87%
1001 – 1500	8.70%
1500 +	28.26%

Table 1.1 : Files Currently Owned

# Expected	Percent
0	17.78%
1 - 5	42.22%
6 - 10	15.56%
11 - 20	8.89%
21 +	4.44%
Don't Know	11.11%

Table 1.2: Expected File Download

Table 1: Files Currently Owned, Expected File Download

Mp3 or similar downloaded file format songs tend to be kept by the students for an extended period. 65.2%, nearly 2/3 of students, claim to hold on to these files for at least one year. An additional 26.1 % of the students keep files for between one month and one year, leaving less than 9 .0% of students who claim to delete files

less than a month after acquiring th em. Over 2/3 of students (67.4 %) burn downloaded songs to CD or DVD, so even if the files are deleted they are available on optic media. The availability of mp3 and similar format music files has decreased the number of CDs purchased by the students. Fewer than 18.0% of the students reported an increase of CD purchases while 42.2 % reported no change. 17.8% reported a slight decrease in their CD purchases and 22.2 % reported a large decrease in CD purchases.

The genre of music typically downloaded can be found bel ow in Table 2, as well as the application used for the se downloads. As would be expected, Rock / Pop is the most commonly downloaded genre. Surprisingly, blues came in second, followed by country. The most commonly used applicat ion for downloading was GnucleusLAN, followed closely by Kazaa. GnucleusLAN is the only service that works properly on the campus network.

Genre of Music	Percent
Rock / Pop	78.26%
Blues	76.09%
Country	69.57%
Rap / Hip-Hop	58.70%
Electronic	45.65%
Classical	43.48%
R&B / Soul	43.48%
Reggae	26.09%
Jazz	21.74%
New Age	21.74%
Folk/Traditional	17.39%
Latin	17.39%
Other	17.39%
Industrial	10.87%

Table 2.1: Genre of Music

Applications	Percent
GnucleusLAN	57.78%
Kazaa	55.56%
Other	31.11%
Napster	20.00%
Web Site	13.33%
Morpheus	8.89%
LimeWire	8.89%
Gnutella	4.44%
BearShare	2.22%

Table 2.2: Applications Used

Table 2: Genre of Music / Applications Used

'*Other*' genres of music included *A Cappella*, *Ambient*, *Comedy*, *Seasonal*, *Punk*, *Ska*, and *Techno*. Additional P2P applications included *Direct Connect* (DC), WinMX, and *iTunes Link Maker* (iTMS).

The list of reasons for downloading mp3 and similar music files yielded interesting results. Each of the listed reasons was seen as '*Very Important*'. In fact, for most of the listed reasons, the '*Very Important*' level was chosen at least 50.0% of the time, as shown by Table 3.

	Unimportant	Somewhat Unimportant	Neutral	Somewhat Important	Very Important
Instant access to songs	4.65%	4.65%	9.30%	20.93%	60.47%
Cheaper than CDs	8.70%	4.35%	17.39%	17.39%	52.17%
More convenient than buying CDs	6.52%	10.87%	13.04%	8.70%	60.87%
Not worth buying CDs for 1-2 songs	2.17%	8.70%	19.57%	17.39%	52.17%
To sample songs	6.52%	10.87%	15.22%	23.91%	43.48%

Table 3: 1-5 Importance of Download Reasons

The survey revealed that a majority of students were happy with the service provided by P2P networks. Students are generally able to find a particular song they are looking for, and are able to fully download the file without having the transfer cut off. These networks are not found to be slow. These results can be viewed in Table 4. Overall, it appears that the students are happy with the overall performance and service of P2P networks for music file downloading.

	Totally False	Somewhat False	Neutral	Somewhat True	Totally True
I can usually find a song I am looking for	2.22%	20.00%	22.22%	28.89%	26.67%
I am happy with the service available through P2P	6.82%	15.91%	34.09%	18.18%	25.00%
Often I cannot finish downloading a file because the connection to the serving node is lost	22.22%	26.67%	22.22%	17.78%	11.11%
P2P networks are usually slow	13.33%	24.44%	35.56%	11.11%	15.56%

Table 4: Satisfaction with P2P Networks

Sharing Behavior

According to the survey results, 80.4% of students in the survey have shared mp3 or similar music files at some point. 15.2% of surveyed students have never shared music files. The historical and current amount of music files shared can be found in Table 5. Note the relative closeness between the near 20% who claim never to have shared with the over 15% who claim never to have shared in a separate question.

Number of Files	Percent
0	15.22%
1 - 20	26.09%
21 - 50	6.52%
51 - 100	6.52%
101 - 250	8.70%
251 - 500	4.35%
501 - 1000	10.87%
1001 - 1500	8.70%
1500+	13.04%

Table 5.1: Historical Number of Files

Number of Files	Percent
0	63.04%
1 - 20	10.87%
21 - 50	4.35%
51 - 100	4.35%
101 - 250	4.35%
251 - 500	2.17%
501 - 1000	6.52%
1001 - 1500	2.17%
1500+	2.17%

Table 5.2: Current Number of Files

Table 5: Historical and Current Number of Files Shared

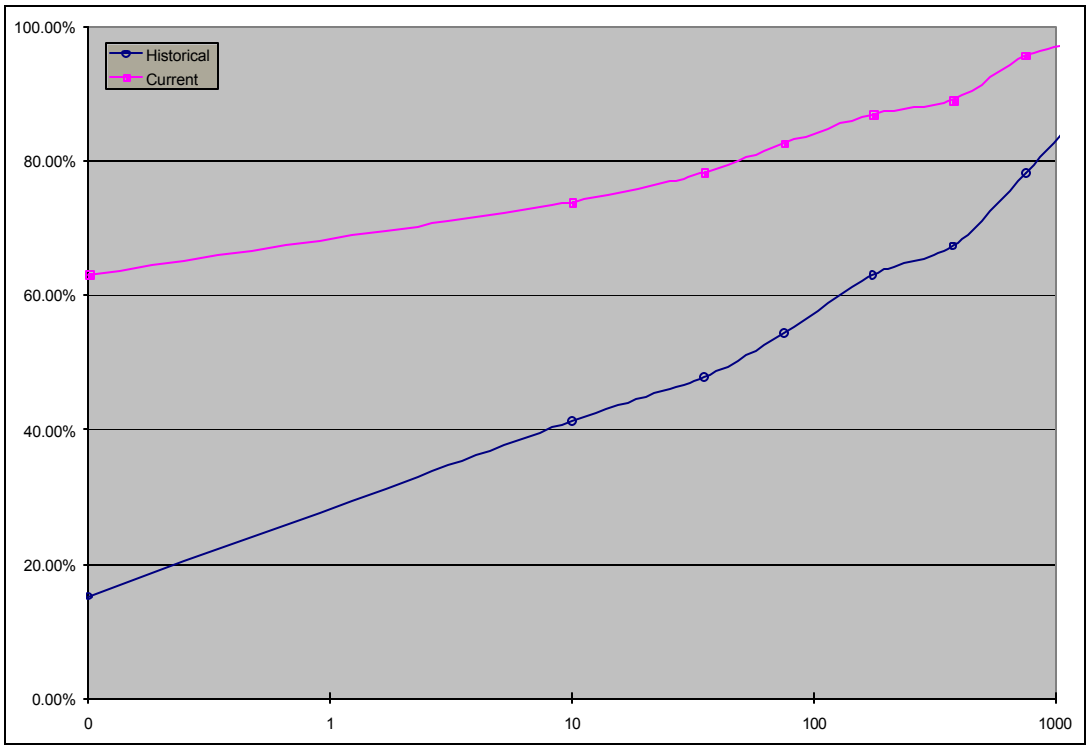

Chart 1: Historical vs. Current File -sharing

?

The amount of sharing has certainly decreased by a substantial amount, as seen in Chart 1. This is evident by comparing the percentage of students currently sharing no files with the percentage of students who historically shared no files.

Although many students are not currently sharing files, most have in the past. When asked why they shared files either now or in the past, students replied as follows in Table 6.

	Unimportant	Somewhat Unimportant	Neutral	Somewhat Important	Very Important
To return the favor to others	21.05%	13.16%	23.68%	18.42%	23.68%
Because pre-recorded CDs are overpriced	23.68%	18.42%	15.79%	15.79%	26.32%
To share music that you enjoy	15.79%	15.79%	28.95%	26.32%	13.16%
Software used to get files makes you share files as well	28.95%	13.16%	26.32%	13.16%	18.42%
To move society toward a new way of doing business	39.47%	18.42%	21.05%	15.79%	5.26%
To defy the wished of music companies	57.89%	10.53%	13.16%	13.16%	5.26%
To defy laws you don't agree with	65.79%	7.89%	13.16%	5.26%	7.89%
To be part of an internet community	52.63%	21.05%	15.79%	5.26%	5.26%

Table 6: Reasons for Sharing Music Files

This table indicates that the most important reasons why students share are '*to return the favor to* others' and because '*pre-recorded CDs are* overpriced'. The least important reasons why the students share their music files are that they wish to defy the wishes of music companies, they wish to defy laws they do not agree with, and to be part of an internet community. When the students are explicitly asked why they choose to share or why they choose not to share, there were a few interesting answers. There were no reasons as to why a user chooses to share. Many of the reasons why users do not share include fear of the RIAA (*four answers*), fear of others accessing files on their computer, and use of hardware/network bandwidth resources. When asked for any additional thoughts on file sharing, many students mentioned that struggling bands are helped by file sharing because of publicity and sampling of music.

Sharing Knowledge

Nearly 85.0% of surveyed students feel that music file swapping should be allowed by law. However, only 8.7% of those surveyed claim to know most of the relevant copyright law. 67. 4% claim to know most of the main points, leaving 23.9% who say they do not know anything about copyright law. In contrast, 23.9% of the students know most of the campus rules regarding music swapping. 52. 2% know some of the rules, and 23.0 % claim not to know the campus rules on music swapping. 13.0% of the students have had their network connection shut down by their ISP and 6.7% have been reprimanded by the school or others for file swapping.

Despite the high publicity of the recent file swapping law suits brought by the RIAA, 15.6% of the students s urveyed have not heard about them. Two thirds of the students have heard of the lawsuits but do not follow closely, and 17.8% of the students follow closely. One student, or 2.2% of the surveyed students , personally claims to know a defendant in one of t hese cases. Knowing about these lawsuits, only 2.2% of student s claim to share more often than before the lawsuits. 60.0 % of the students share the same amount, and 37.8 % share less.

When asked for why sharing should or should not be legal, students str ongly supported legalization of sharing. The students that defended not legalizing audio file sharing all mentioned copyrights. The arguments for legalizing audio file sharing were much more varied. Students argued that it is impossible to shut down the P2P networks, so fighting P2P is a losing battle. Others argued that most people buy a CD containing the songs they download. Others argued that money is hard to come buy and CDs are overpriced, so they are forced to download music files. The answers range from '*No one should profit from music*' to '*They are only data files*' to '*It is like listening to the radio*'. Only one student gave a good argument regarding the original intent of copyrighting, which is to prevent individuals from profiting off of someone else's work, and that there is no profit made by P2P music file transfers.

Personal Information

Everyone participating in the survey owns a computer. Only one individual uses a dial-up service. 47.8% connect via cable or DSL, 43. 5% connect with a T1 or greater (campus or fraternity house), and 6.5% are unsure (also probably campus or fraternity house).

A majority of the studen ts surveyed (63.0 %) consider themselves to be somewhat '*computer savvy*'. 26.1% consider themselves to be very ' *computer savvy*', leaving 10.9% to be not at all ' *computer savvy* '. Common Internet uses can be found below, in Table 7. Other Internet uses not included in the questionnaire were banking, business, public forums, and writing.

Internet Uses	Percent of Users
Research	97.83%
Instant Messaging	93.48%
Shopping (not including music)	73.91%
Games	71.74%
Programming	36.96%
CD shopping	26.09%
Maintaining a personal server	23.91%
Digital music shopping	10.87%
Other	10.87%
Dating	0.00%

Table 7: Typical Internet Uses

The students consisted of 28. 3% freshmen, 3 7.0% sophomores, 15.2% juniors, and 19.6% seniors. The male to female ratio nearly matched that of the school at 75. 6% male. The average age of the sampled students was 19.7 years old, with a standard deviation of 1.6 years . Overall, this is a fair sample of the population of the entire undergraduate body.

Analysis

The results yield some interesting information regarding who tends to upload and download music files in mp3 or similar format.

Who Downloads Music Files

There seems to be a strong positive correlation (*0.42*) between discretionary spending the number of CDs owned , and the number of CDs owned has a strong positive correlation (*0.39*) with the amount or mp3 or similar music files owned . There also is a strong correlation (*0.48*) between discretionary spending and number of mp3 or similar music files owned. Knowledge of RIAA lawsuits seems to have no serious effect on the amount of mp3s owned – in fact, those aware of RIAA lawsuits tend to own more mp3 or similar music files than those not aware of these cases (correlation of *0.47*). This makes sense because the more a particular student downloads, the more inclined they would be to follow news related to this particular activity. Those who do not share would have less personal interest in such news.

Who Shares Music Files

Historically, there was a strong correlation (*0.48*) between the number of files owned and the number of files sh ared. This has dropped to a correlation of *0.08* – considerably less than before. Historically, those who thought that music file -sharing should be legal were directly correlated with the amount of files shared. Now there is an negative correlation betwe en the same factors. Across the board, all correlations regarding sharing have dropped from their historical values as the amount of sharing has decreased.

Discussion

The results of the survey support many suspected concepts. This support ranges from t he fact that mp3 availability is correlated with less spending on legitimate music, to the fact that individuals can find few good reasons to share files .

Effects of Sharing

It would seem as if the RIAA has been somewhat successful in their attempt to thw art file sharing using P2P services. Many students answered in open-ended questions that they do not share due to fear of the RIA A. Most of the students were aware of RIAA lawsuits, and the amount of music files shared is significantly less than it has b een in the past (t-test: 0.000 16). The students seem to see little reason to share , and they perceive more risk than they deem worthwhile to share m3p or similar music files. In fact, the agreed on reason for sharing was that '*Software used to get files makes you share files as well*', which indicates that more people share because they think they have to as opposed to because they wish to. The very fact that fewer individuals are sharing could begin a downward spiral of availability of particular genres and songs on these P2P networks. The less files are available, the less draw the system has to individuals, which, in turn, causes others to d iscontinue use of the service.

Although the students seem comfortable and willing to shop online, only one quarter of the students purchase CDs online. A mere eleven percent purchase digital music on the internet. The students enjoy the service of P2P which allows them to download audio files at no cost . This is most likely a case of ' *Why pay for what can be rece ived for free?*'

Peer-to-Peer Misperceptions

There is much confusion within the world of P2P audio file sharing. Results from the survey show that the students believe that CD sales are increasing due to mp3 and other music files being available. For s ome reason, there is a popular belief that most of these downloads are individuals seeking new types of music and new groups to listen to. If an individual enjoys the new music, they will buy the CD. A few students even directly stated this belief in the open-ended questions. However, the answers from the survey show just the opposite. Most students download rock and pop tunes, most of which they have probably heard before. Instead of sampling these songs, most of the time the files are kept for a year or longer, and are often burned to permanent optical media. In addition, the amount of CDs purchased by the students has decreased since the availability of digital audio downloading. The evidence found in this survey points

to the scenario where individuals download songs they have heard and liked, and keep the file instead of buying a legitimate copy of the song.

Conclusion

The results of the study show that the RIAA lawsuits have had an effect on the file-sharing community. However, it is not known from the results of this survey if there are other major reasons for individuals to share music files. After this study was completed, there was a study conducted by the *Pew Internet and American Life Project* that found similar results [8]. According to the study, the number of internet users downloading music files has halved.

In addition to the *Pew Internet and American Life Project* study, the US Court of Appeals for the District of Columbia overturned a previous ruling which stated that Internet Service Providers (ISPs) must release personal information of customers suspected of file sharing to the RIAA in accordance with the Digital Millennium Copyright Act (DMCA) [9]. Now an individual must first be proven to be sharing music files in order to be subpoenaed. This overruling could mislead individuals into feeling that the RIAA no longer has any way to track down file sharers, which is an incorrect assumption.

Going forward, it would be interesting to see how this latest ruling affects the world of P2P sharing and whether this downward trend in music sharing continues. Another interesting study would be to run a similar experiment at a different, perhaps non-technical, school and see if the results vary from the findings of this study. It will be interesting to see what happens to P2P networks as they gain more recognition from computer users and corporations alike.

References

1. Shirky, Clay. (2000) "What is P2P… And What Isn't". *http://www.openp2p.com/pub/a/p2p/2000/11/24/shirky1-whatisp2p.html*, Accessed 17 December 2003.
2. The Gnutella Protocol Specification v0.4 (Document Revision 1.2) Dec. 2003, http://www9.limewire.com/developer/gnutella_protocol_0.4.pdf, Accessed 17 December 2003.
3. Kushner, D. (2003). "Digital Decoys", *IEEE Spectrum* 5(40)
4. M. Graziano and L. Rainie,(2001). "The Music Downloading Deluge". Pew Internet & American Life Project, at *http://www.pewinternet.org/reports/toc.asp?Report=33*, Accessed 17 December 2003.
5. Adar, Eytan & Berna rdo A. Huberman. (2000) "Free Riding on Gnutella". *First Monday*, 5(10) *http://www.firstmonday.dk/issues/issue5_10/adar/*, Accessed 17 December 2003.
6. Oram, Andy. (2000) "Peer-to-Peer Makes the Internet Interesting Again". *http://linux.oreillynet.com/pub/a/linux/2000/09/22/p2psummit.html?page=1*, Accessed 17 December 2003.
7. Galloway, Terrel & Douglas Kinnear. (2001) "Unchained melody: A price discrimination-based policy proposal for addressing the MP3 revolution". *Journal of Economic Issues* 2(35)
8. L. Rainie & M. Madden (2004). "Sharp decline in music file swapping: Data memo from PIP and comScore Media Metrix". Pew Internet & American Life Project, at *http://www.pewinternet.org/reports/toc.asp?Report=109*, Accessed 4 December 2004.
9. McGuire, David. (2003). "Court Re jects Music Industry Subpoenas". *http://www.washingtonpost.com/wp-dyn/articles/A14836-2003Dec19.html*, Accessed 8 December 2004.

Appendix A – The Consent Form

Thank you for your interest in this survey. Your answers will help me to determine how frequently and why college students share music files. At the end of the survey, you will be given a unique number to send to your professor as verification that you completed the survey. Please note that the survey must be *fully* completed in order to receive credit.

If you wish to fill out the survey, click on the link below. Your results will be kept confidential. If you choose not to complete the survey, close the browser at any time.

Take the survey

If you have any technical problems with the survey, please inform (*support address*).

Appendix B - The Survey

This survey is being conducted in order to determine file -sharing habits over peer-to-peer networks. Your answers will be compiled anonymously with resul ts of the other survey participants.

Section 1: Downloading Behavior

The following questions ask about your music -file <u>downloading</u> behaviors.

1. Have you ever downloaded mp3 or similar music files from the internet using a peer-to-peer network? Example app lications would be *Napster*, *Gnutella*, *Morpheus*, *LimeWire*, *BearShare*, *Kazaa*, *GnucleusLAN*, *etc.*
    a. Yes
    b. No

2. Have you ever purchased mp3 or similar music files online?
    a. Yes
    b. No

3. What types of music files do you currently own, if any?
    a. Mp3
    b. Wma
    c. Ogg
    d. Other (Please specify) _____

4. Approximately how many songs do you own in mp3 or similar format?
    a. 0
    b. 1 – 20
    c. 21 – 50
    d. 51 – 100
    e. 101 – 250
    f. 251 – 500
    g. 501 – 1000
    h. 1001 – 1500
    i. 1500+

5. How many CDs, tapes, records do you own? _____

6. Approximately how many mp3 or similar files will you dow nload in the next week?
    a. 0
    b. 1 – 5

c. 6 – 10
d. 11 – 20
e. 21+
f. Don't Know

7. What types of music do you download in mp3 or similar format?
   a. Blues
   b. Classical
   c. Country
   d. Electronic
   e. Folk/Traditional
   f. Industrial
   g. Jazz
   h. Latin
   i. New Age
   j. R&B/Soul
   k. Rap/Hip-Hop
   l. Reggae
   m. Rock/Pop
   n. Other (Plea se specify) _____

8. What application(s) do you use to download mp3 or similar music files?
   a. Napster
   b. Gnutella
   c. Morpheus
   d. Limewire
   e. BearShare
   f. Kazaa
   g. GnucleusLAN
   h. Web Site
   i. Other (Please specify) _____

9. Why do you download music files in mp3 or similar format? (please ra nk importance where 1 is not important, 5 is very important)
   a. Cheaper than CDs
   b. More convenient than buying CDs
   c. Not worth buying CD for 1-2 songs
   d. To sample songs
   e. Instant access to songs
   f. Other (Please specify)

10. How true are each of the following statements? (please rank the severity of the truth where 1 is false and 5 is true)
    a. I am happy with the service available through P2P
    b. I can usually find a song I am looking for
    c. P2P networks are usually slow

d. Often I cannot finish downloading a file because the connectio n to the serving node is lost

11. How long on average do you keep your mp3 or similar formats songs on your computer?
    a. One day or less
    b. Less than one week, but greater than one day
    c. Less than one month, but greater than one week
    d. Less than one year, but greater than one month
    e. One year or greater

12. Do you burn downloaded music files to CD/DVD/etc?
    a. Yes
    b. No

13. If you have downloaded songs in mp3 or similar formats, how do you feel the ability to download these songs has affected your CD purchasing habits?
    a. Large increa se in CD purchases
    b. Small increase in CD purchases
    c. No change in CD purchasing habits
    d. Small decrease in CD purchases
    e. Large decrease in CD purchases

Section 2: Sharing Behavior

The following questions ask about your music -file <u>sharing</u> behaviors.

14. Have you e ver shared mp3 or similar music files on the internet using a peer -to-peer service?
    a. Yes
    b. No

15. What is the largest number of music files you have shared at any point <u>ever</u>?
    a. 0
    b. 1 – 20
    c. 21 – 50
    d. 51 – 100
    e. 101 – 250
    f. 251 – 500
    g. 501 – 1000
    h. 1001 – 1500
    i. 1500+

16. How many mp3 or similar music files are you currently sharing on a peer-to-peer network?
    a. 0
    b. 1 – 20
    c. 21 – 50
    d. 51 – 100
    e. 101 – 250
    f. 251 – 500
    g. 501 – 1000
    h. 1001 – 1500
    i. 1500+

17. Why do you share songs in mp3 or similar format? (please rank importance where 1 is not important, 5 is very important)
    a. To return the favor to others
    b. Because pre-recorded CDs are overpriced
    c. To be part of an internet community
    d. To share music that you enjoy
    e. Software used to get files makes you share files as well
    f. To defy the wishes of music companies
    g. To defy laws you don't agree with
    h. To move society toward a new way of doing business
    i. Other (Please specify) _____

18. If you do not share music files, please explain why.

19. Please indicate any additional thoughts on file-sharing.

Section 3: Sharing Knowledge

The following questions will ask you about your knowledge of file-swapping.

20. Do you feel that you should be allowed by law to swap mp3 and similar music files?
    a. Yes
    b. No

21. Why or why not?

22. How familiar are you with current copyright laws regarding sharing of mus ic files?
    a. Not at all
    b. Know the main points
    c. Know most/all relevant legislation

23. How familiar are you with current campus rules regarding swapping of music files on the campus LAN?
    a. Do not know campus rules regarding music swapping
    b. Know some campus rules rega rding music swapping
    c. Know most/all campus rules regarding music swapping

24. Have you ever had your network connection shut down by your ISP (NetOps, Verizon, etc)?
    a. Yes
    b. No

25. Have you ever been reprimanded for file -sharing?
    a. Yes
    b. No

26. Are you aware of lawsuits brought by the *Recording Industry Association of America* against individual file -swappers?
    a. No, have not heard of any lawsuits
    b. Yes, have heard about lawsuits but do not follow them
    c. Yes, have heard about lawsuits and follow them

27. If you are aware of lawsuit s against individual music file swappers, do you personally know a defendant involved in one of these cases?
    a. Yes
    b. No

28. If you are aware of lawsuits against individual music file swappers, how has this knowledge effected your sharing behavior?
    a. Considerably more file swapping
    b. A little more file swapping
    c. No change in file swapping
    d. A little less file swapping
    e. Considerably less file swapping

Section 4: Personal Information

The following questions ask about your personal information. This information will only be used for the survey.

29. Do you own a computer?
    a. Yes
    b. No

30. What internet connection type do you use most often for file swapping?
    a. Dial-up
    b. Cable
    c. DSL
    d. T1+
    e. Don't know

31. Which of the following services do you use the internet for?
    a. Research
    b. Games
    c. Dating
    d. Shopping (not including music)
    e. CD shopping
    f. Digital music shopping
    g. Instant messaging
    h. Programming
    i. Maintaining a personal server (or servers)
    j. Other (Please specify) _____

32. Do you consider yourself to be *computer savvy*?
    a. Not at all computer savvy
    b. Somewhat computer savvy
    c. Very computer savvy

33. What year are you in school?
    a. Freshman
    b. Sophomore
    c. Junior
    d. Senior
    e. Other (Please specify) _____

34. What is your age? _____

35. What is your intended major? _____

36. What is your gender?
    a. Male
    b. Female

37. What is your average monthly discretionary spending? (d iscretionary spending covers dining, entertainment, etc)
    a. $50 or less
    b. $51 - $100

c. $101 - $150
d. $151 - $200
e. $201 - $300
f. $301 - $500
g. $501+

38. What are your total average monthly expenses, including rent, car, groceries, discretionary, etc, excluding tuition/scho ol fees?
    a. $200 or less
    b. $201 - $400
    c. $401 - $600
    d. $601 - $800
    e. $801 - $1000
    f. $1001 - $1500
    g. $1500+

Appendix C – Debriefing Form

Your unique ID is XXXXXXX. Please report this number to your professor.

Thank you for participating in this study. If you are not aware of campus sharing policy or American copyright laws, the following links may prove interesting.

WPI's AUP

Title 17 - Copyrights

If you have any questions regarding this survey, please email (*support address*).

Thanks for your time!