ANALYSIS OF INTERNET MUSIC CONTENT DISTRIBUTION

Sasha Slijepcevic
UCLA CS Dept.
Los Angeles, CA 90095
sascha@cs.ucla.edu

Richard Muntz
UCLA CS Dept.
Los Angeles, CA 90095
muntz@cs.ucla.edu

Miodrag Potkonjak
UCLA CS Dept.
Los Angeles, CA 90095
miodrag@cs.ucla.edu

Vincent Busam
UCLA CS Dept.
Los Angeles, CA 90095
vince@ucla.edu

ABSTRACT

The Internet allowed previously unattainable ease of information exchange. However, the Internet also enabled a fast and widespread unauthorized distribution of copyrighted content. A solution for that problem that will retain a free flow of information, but protect the rights of the content owners is still sought.

In this project, we analyze the availability of the copyrighted music content on the Internet. Our goal is to acquire technological, geographical, and sociological information about the users of the music content on the Internet. We believe that such information may significantly help to define directions for future of the content distribution on the Internet.

1. INTRODUCTION

The goal of this project is to give a statistical analysis of behavior of Internet users toward pirated audio material. We believe that insights into technological, geographical, and sociological characteristics of users of the Internet can significantly help when evaluating different proposals to protect from privacy the musical content on the Internet. In this paper, we analyze the data we gathered observing behavior of the users on the Internet. We also describe the tools we developed for data gathering and analysis.

The ease with which all kinds of data are exchanged on the Internet enabled an overall improvement in communication, but it also allowed unlimited exchange of the copyrighted content. Although illegal copies of software products and movies are available online, the exchange of music files attracted the most attention. The introduction of a new compression standard for music, MPEG 1 Layer 3 [1], allowed users to compress hundreds of megabytes of music into 3 or 4 megabytes. However, only when several Web sites offered listings of available music content, among them the most popular was Napster [2], content protection attracted a wide attention in music industry and media. Although Napster is not currently active, there are many other ways and technologies that ensure that the problem is still open. Since the existence of a central server made Web sites vulnerable to legal charges, most of the current music content distribution networks are based on peer-to-peer communication [3], but many FTP servers offering copyrighted music content are also available.

Our goal was twofold. The first goal was to supply statistical data to the discussion that would let us more clearly develop possible strategies to diminish effects of widespread piracy. We wanted to detect possible patterns in spreading of music content. Also, we wanted to confirm widely stated beliefs of university networks as the main source of illegally copied materials.

2. DATA GATHERING TECHNOLOGY

In order to achieve the stated goal, we had to develop software tools that will allow us to keep track of the trends in music content distribution on the Internet. The shift in popular technologies and a wide range of available sources of music content required a parallel development of various software tools. We directed our efforts along two directions. On one side, we developed software that tracks FTP servers. Although FTP servers are rarely used anymore, they offered a good oportunity to have a complete overview of all files that Internet users may have on their machines. FTP servers allow a client to get a complete list of all available files, while that is not the case on Napster and other server-based and peer-to-peer networks. For FTP servers, we acquired lists of popular servers from the sites that offered such lists [4][5][6]. On the other side we developed tools for tracking Napster users. A simple way to get information about the users that offered music content on Napster was to develop a client that connects to one or more Napster servers and issues requests.

In both cases, for FTP and for Napster, we wanted to acquire IP addresses of the users. IP addresses allow us to find certain information about a particular user. In the case of the FTP servers, the lists from the given Web sites include only DNS names, and not IP addresses. We did not connect to each of them to acquire their FTP addresses, but we performed DNS

lookups. We acquired IP addresses of Napster users by issuing a large number of requests for the most popular songs. Our Napster client would then record IP addresses, returned by a Napster server, of the users that offered a particular song.

Once when we acquire an IP address, using *whois* service, we can find the owner of the network from which a user connects to the Internet. From the owner of the network, in many cases we can find the geographical location of the user, the type of connection that the user uses, and the type of the institution from which the user connects.

3. MOTIVATION

In this section, we describe our motivation for gathering and analyzing specific type of data. We do not propose specifically for which purposes the data that we offer should be used. However, the most discussed question concerning music content distribution is related to the possibility to stop unauthorized file sharing. Therefore, we use our results to estimate potential viability of proposed solutions for music content protection.

Various solutions have been offered in order to protect music content on the Internet. They range from legal prosecution of the owners of Web servers and the users who serve as sources of the illegal music content to devices that would be able to play only registered content, which on the other side could not be played on existing audio devices.

One of the first reactions to the music distribution revolution initiated by Napster was to disallow Napster connections from university dorms. We give statistics here that show that although important, the traffic from universities represents less than 25% of the users of Napsters. More generally, the legal approach has more chance of success if the distribution of users is such that small number of users offers most of the available musical content. We found that the users on broadband connections are significantly more likely to own and distribute illegal musical content. Combined with the fact that the number of providers offering broadband Internet access is still relatively small, it may seem that for the time being, success in limiting offer of illegal music content on broadband connections may significantly decrease music content piracy. However, the broadband connections are inevitably more and more widespread, so even if currently the given approach may cut piracy, in the long run that solution does not solve the given problem.

Finally, the legal battle against music content piracy might be successfully led in the United States. However, there are many countries in the world where prosecution of illegal content distribution may not be easily achieved. Therefore, we believe that geographical distribution of the Internet users that offer illegal music content is important when the music content protection solutions are discussed.

4. RESULTS

The number of users that use broadband connections to connect to the Internet grows steadily. The percentage of broadband users climbed from 7% in 2000 up to 16% in 2002 [7].

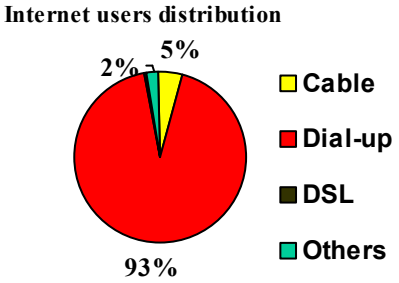

Fig 1. Distribution of the Internet users by the type of connection.

In Fig. 1 we show the data for the year 2000, when we performed the analysis. Even if the number of users that had a broadband connection at that time was 7%, they made more than quarter of Napster users, as shown in Fig. 2. On Napster there is no definite division between the content servers and the clients. However, in reality, as shown in [8], the majority of users (70%), on such networks as Napster and Gnutella [9], do not share any files, while on the other side 1% of users offer 37% of all available files. The results for FTP servers, given in Fig. 3, present even higher discrepancy between the number of the broadband users among the general Internet population and among the owners of the FTP servers with the copyrighted music content. The clients that connect to FTP servers cannot be tracked so easily. However, the FTP servers function using exchange mechanism, therefore the network of the FTP servers behave rather as a peer-to-peer network, and not as a standard client-server type of network.

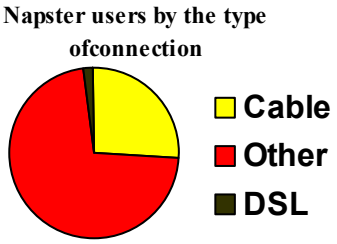

Fig 2. Distribution of Napster users by the type of connection.

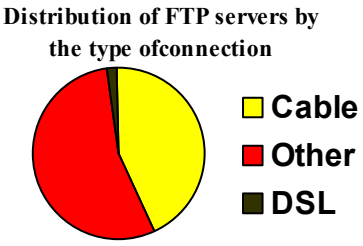

Fig 3. Distribution of FTP servers with copyrighted music content.

The number of files available at the FTP servers at that time was around 230,000. On 58 FTP servers, the number of files was more than 1,000. One FTP server in Canada offered 8,500 different MP3 files. The data that we gathered, together with the data from [8], show that legal prosecution of relatively small group of users might significantly decrease the amount of copyrighted material available on the Internet.

From April 2000 to June 2000, Napster experienced its biggest expansion. At the same time, the number of available FTP servers with music content decreased rapidly. We have found 3800 different FTP servers available in April, while only 390 available in June. Besides Napster, the reason for such a rapid decrease was in the policy of the cable companies not to allow incoming traffic to the well-known TCP ports.

The next set of the acquired data deals with geographical locations of FTP servers and Napster users. The importance of that data mainly lies with the fact that majority of the users are located in North America and Western Europe, where legal protection of copyrighted content can be achieved. In case of FTP servers, the geographical location of the server can be easily different from the location of the owner of the server. However, even in July 2000, when Napster ceased to operate, there were no inclinations that FTP servers are moved to geographically distant locations.

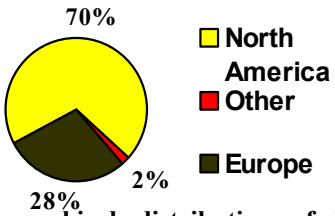

Fig 4. Geographical distribution of FTP servers.

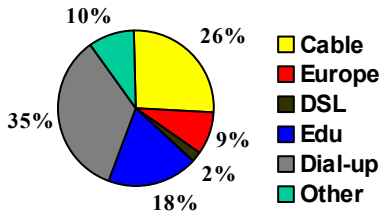

Fig 5. Domain and Geographical distribution of Napster users.

Napster users in April 2000 were overwhelmingly located in North America. In Fig. 5 we present together geographical distribution of Napster users and the distribution of the users between high-level domains.

The observations that can be made from Fig. 5 is that the users from the educational domain are represented in smaller percentage than it was perceived in the media at that time. According to this observation, proliferation of unauthorized music content could not be achieved by blocking the access to Napster from university networks that include dorms.

5. RELATED WORK

The distribution of music content on the Internet has attracted a lot of attention for several reasons. Significant scientific and technological improvement, as Mpeg 1 Layer 3 [1] and peer-to-peer networks [3], were brought to wider attention or even inspired by the music content distribution. On the other side, unauthorized distribution of the music content is considered a serious threat to the music industry [10]. In [11], the results of interviews with 2,555 college students are given. Only 500 reported that they used Napster to download one or more songs. 40% out of 500 students stated that Napster impacted their music purchases, so that they are buying fewer CDs than before. On the other side, 20% of 500 students stated that they use Napster to help them decide what to buy or to make better selection. The research in [11] was focused on college students, assuming that the college students make majority of the owners of unauthorized music content on the Internet. However, in [12] and also in our analyses we can see that the users of the music content are distributed across the general population.

Behavior of users on popular peer-to-peer networks is the topic of several research projects. Peer-to-peer networks are based on assumption that the users voluntarily contribute their resources to establish a network. However, as shown in [8] and [12], both Napster and Gnutella networks exhibit the characteristics of content-centralized systems where small number of users, 7% on Gnutella as reported in [12], are significant sources of music content, while

the rest of the users do not share or share a small number of files.

6. CONCLUSION

The unauthorized distribution of the music content remains one of the most interesting and the most important problems on the Internet. Proposed solutions include different legal and technological mechanisms that may have long-lasting consequences on the information exchange on the Internet.

The goal of our project is to evaluate possible solutions in light of the observed characteristics and behavior of the users of the music content. We showed that in spite of technological framework of the peer-to-peer networks, where each user can be a server or a client, offering or downloading the content, in reality there is a small number of users on Napster that offer majority of the content. However, looking at the number of the users with broadband connections that take part in the exchange of the music content, we can expect that with increase of the broadband connections, the distribution of the users who offer the content will be more evenly distributed. Geographical location of the users is important information, if any of the legal mechanisms are to be used to protect music content. The majority of the users are in the US, Canada, and Western Europe, where the legal protection of the copyrighted content can be achieved.

Finally, we show that the college students in university dorms are a significant fraction of the users who download MP3 files, but at 18% limitation of the content distribution on university networks alone, cannot solve the problem of unauthorized distribution.

REFERENCES

[1] Leonardo Chiariglione: MPEG: A Technological Basis for Multimedia Applications. IEEE MultiMedia 2(1): 85-89 (1995).

[2] http://www.napster.com

[3] Ian Clarke, Oskar Sandberg, Brandon Wiley, and Theodore W. Hong, "Freenet: A Distributed Anonymous Information Storage and Retrieval System," in Designing Privacy Enhancing Technologies: International Workshop on Design Issues in Anonymity and Unobservability, LNCS 2009, edited by Hannes Federrath. Springer: Berlin (2001).

[4] http://music.lycos.com

[5] http://en.sonico.com

[6] http://www.oth.net

[7] http://cyberatlas.internet.com/markets/broadband/article/0,,10099_1016281,00.html

[8] E. Adar, B. Huberman, Free riding on Gnutella, First Monday, Vol. 5-10, October 2, 2000.

[9] http://www.gnutella.com.

[10] SoundScan study, http://www.riaa.org/PDF/fine.pdf.

[11] http://www.riaa.org/PDF/jay.pdf.

[12] http://www.pewinternet.org/reports/toc.asp?Report=33