UMG Recording Inc., et al. v. Lindor
ED – NY Case Number: 05-cv-1095
Expert witness report by Dr. J.A. Pouwelse

General statements on Peer-to-Peer

The topic of Peer-to-Peer (P2P) is attracting wide spread attention. This new technology enables people to distribute information and communicate at only marginal cost.

P2P file sharing is both controversial and popular. File sharing means connecting millions of computer hard disks together into a single network. Roughly 74% of all Internet traffic consists of P2P file sharing traffic[1]. Content creators are under pressure from two sides. On one side, their customers are using P2P file sharing to download movies, music, and songs for free[2]. On the other side, bands such as Radiohead using the Internet to bypass them[3]. With P2P artists themselves can reach a worldwide audience of millions at only marginal cost. Within KaZaA, users can use "micropayments" to pay artists directly and download legally. The economic impact of file sharing is still poorly understood. For instance, a leading study by Harvard researchers was unable to find a relation between illegal downloading and decreases in Audio CD sales[4].

Measurements of file sharing networks

File sharing networks are difficult to measure. Only a few companies and universities in the world have the required expertise to conduct measurements of file sharing networks. It is very difficult to directly establish that a certain computer contains copyrighted works and makes them available to others through a file sharing application.

The first problem is that we need to have an understanding of the file sharing application itself. This is difficult due to the complexity of such applications and lack of detailed documentation about their inner workings. The second problem is that we often do not have physical access to the computer under investigation. When we can only observe this computer through The Internet, we are severly limited in our observational power. The third problem is that The Internet and P2P are dark places where people commit fraud and abuse. All obtained information must be treated with suspicion. Users use fraudulent means to obtain a higher download speed from their broadband ADSL connection, install abusive software to obtain higher downloads on a file sharing network (at the cost of other people), and like to fool other people with fake content on file sharing networks.

1 http://www.ipoque.com/media/internet_studies/internet_study_2007
2 http://money.cnn.com/2005/05/25/technology/piracy/
3 http://entertainment.timesonline.co.uk/tol/arts_and_entertainment/music/article2602597.ece
4 www.unc.edu/~cigar/papers/FileSharing_March2004.pdf

The KaZaA file sharing system

Only one detailed study has been conducted of the KaZaA file sharing network[5]. This study is conducted by the research group of Professor Keith Ross from Brooklyn Polytechnic University. They investigated how KaZaA operates and measured it extensively.

This research group focused on the pollution in KaZaA[6]. Pollution refers to meaningless files and mismatches between filenames and their actual content. KaZaA was found to be severly polluted. For many recent pop songs, more than 50% of the copies were polluted. Our research group at Delft University has found similar pollution levels in KaZaA for all types of content.

There are three causes of pollution. First is the unintentional pollution by average users when they insert files such as "credit_card_statements.doc" into the system[7]. Second is the intentional pollution by users for fun. For example, a file named "hot big blond women playing around.mpeg" that contains a movie of a laughing clown. Third is the active pollution by companies in an attempt to reduce piracy. Several companies exploit weaknesses in KaZaA in order to pollute the search results of popular queries[8]. Their aim is to reduce the usability of KaZaA in searches for popular copyrighted works.

The KaZaA-lite software is also described in the measurements of Keith Ross's team. This popular, modified version of the official KaZaA client provides improved performance. However, this performance gain comes at the cost of others and KaZaA-lite lies to KaZaA users to obtain more performance. This phenomenon indicates that information from the KaZaA network must be treated with suspicion.

The KaZaA software communicates with numerous other computers on The Internet during its operation. Communication can consist of transmission of advertisement data, instant messages, actual file transfers, and control traffic for maintaining the file sharing network. KaZaA has a special feature to increase file downloads, called multi-peer downloading. When the same file is present on several computers it is possible to download pieces of this file in parallel from multiple computers.

Accurate file sharing measurements

Due to the complexity of file sharing applications, limited observation powers, rampant deception, high pollution levels, and multi-peer downloading it is nearly impossible to obtain solid evidence and detailed checks are therefore required.

I believe that the following 6-step test takes the necessary precautions when trying to establish if a computer is making copyrighted works available for download.

---

5 http://cis.poly.edu/~ross/papers/KaZaAOverlay.pdf

6 http://cis.poly.edu/~ross/papers/pollution.pdf

7 http://www.hpl.hp.com/news/2002/apr-jun/kazaa.html

8 http://www.zeropaid.com/news/articles/auto/08262003a.php

1. Collect filenames by searching the network using keywords.
2. Filter out polluted files by checking the actual content.
3. Establish that a specific file can be downloaded from a certain computer. File sharing applications often talk to numerous other computers at once. Sufficient hygiene precautions should be taken by blocking traffic from all possible other computers.
4. Investigate if the computer is possibly highjacked or the Internet connection is shared with others. Check if a computer is cracked, for instance, running an open proxy or a hacked Microsoft Internet connection sharing application. A measurement is needed to establish if there is no significant difference in traceroute timings, SYN responses, and KaZaA protocol rendezvous times.
5. Track this computer for several days if it does not go offline for reliable IP-address translation by an ISP.
6. Establish that no IP address spoofing, BGP hijacking, or other tampering with IP addresses has taken place.

Review of case material

After reviewing the material listed below I conclude the following:

A) two reports by Dr. Jacobson where based in total on roughly an hour of work

Plaintiffs witness Dr. Jacobson deposition transcript at page 53 states:

> *"Q. And how much time did you spend on the April 2006 report in this case?*
> *A. Without seeing the billing records, I can only guess but I think it was 45 minutes."*

and on page 54 states:

> *"Q. And how much time did you spend on the December 19th declaration?*
> *A. Maybe 15 minutes."*

In my opinion this limit amount of effort spend investigating matters supports a notion that there has been a lack off both in-depth analysis and proper scientific scrutiny. It is impossible to go through all the exhibits in one hour. For instance, examination of exhibit 11 (a 139 page document) and discovery of anomalies and forensic clues such as "desktop.ini' and "kmd251_en.exe" requires a few hours.

B) the April 2006 report includes in my opinion factually erroneous and misleading statements

The first witness report of Dr. Jacobson dated April 7, 2006 marked as exhibit 16 shows in statement marked 12 on "The Internet and Addressing":

> *"The Internet is a collection of interconnected computers or network devices. In order to be able to deliver traffic from one computer or network device to another, each computer or network device must have a unique address within the Internet. The unique address is called the Internet Protocol (IP) address. This is analogous to the postal system where each mail drop has a unique address."*

The above statement is factually erroneous as networks of networks can have many duplicate IP addresses. Many computers can be connected to the Internet with identical IP addresses as long as

they remain behind control points such as routers, firewalls, proxy servers, or similar technologies. Furthermore, the comparison of IP addresses to mail drop points in the postal system is misleading as this suggests a degree of accuracy, simplicity, reliability, certainty, and robustness to fraud. The same deposition shows in statement marked 13 on "Peer-to-Peer networks":

> *"The users of the peer-to-peer network often think they are anonymous when they distribute files. In reality, they can be identified using the IP address. The IP address of the computer offering the files for distribution can be captured by a user during a search or file transfer.*

The above statement is factually erroneous as an IP address captured from a peer-to-peer network during search or file transfer cannot identify a **user** (see the "Accurate file sharing measurements" section above on **computer** identification). This statement suggests precision where precision does not exist. Numerous technical measures exist and are in use to make such identification impossible. For instance, computers can share an external IP address, computer on the same subnetwork can steal IP addresses, a computer can be cracked and used by others as a proxy, or one can seize control of a large block of adjacent IP addresses with a method know as "BGP hijacking".

C) there is lack of knowledge on MediaSentry procedures, methods, and failure rate

The first report of Dr. Jacobson dated April 7, 2006 marked as exhibit 16 shows in statement marked 15 on "conclusions":

> *"I will testify to the procedures used and results obtained by MediaSentry coupled with the information supplied by defendants ISP, to demonstrate the defendant's Internet account and computer were used to download and upload Copyrighted music from the Internet using the KaZaA peer-to-peer network."*

This report indicates that Dr. Jacobson has knowledge of "procedures used" by MediaSentry. However, plaintiffs witness Jacobson deposition transcript at page 32 states:

> "Q. *Do you know what processes and procedures MediaSentry employed?*
> A. *I do not know the inner works of MediaSentry processes and procedures.*
> Q. *Do you know what software they used?*
> A. *No."*

The latter indicates that Dr. Jacobson is not competent to judge the accuracy of information supplied by MediaSentry and his analysis can in my opinion be regarded as hearsay information from third party MediaSentry.

Evidence exists that information supplied by MediaSentry was flawed in other cases. Numerous institutions have received false MediaSentry claims regarding peer-to-peer activity on their computer network. MediaSentry supplied information often involved non-existant or inactive IP addresses. Erroneous MediaSentry claims have been reported by: Yale University, Princeton University, University of California Los Angeles, University of California Santa Barbara, UNC Chapel Hill, University of Northern Iowa, Virginia Tech, College of William & Mary, Georgetown University, Glasgow University Computing Service, Metropolitan State College of Denver,

Western Michigan University, Cleveland State University[9].

It is important to note that in the above cases the claims made by MediaSentry where checked for their validity by full-time network administrators that employ numerous complex technical tools which have direct access to detailed network accounting data. Such full-time administrators, tools, and data are not available in the case of Ms. Lindor.

Finally, to my understanding no independent review of MediaSentry procedures and methods has ever taken place. Their operation, accuracy, and error rate is unknown. From the presented evidence in this case I believe their procedures and methods are simplistic and fail the 6-step "Accurate file sharing measurements" test, as described previously.

C) there is lack of knowledge on Verizon procedures, methods, and failure rate

Plaintiffs witness Dr. Jacobson deposition transcript at page 128 states:

> *"Q. Do you know what procedures Verizon employed to link Ms. Lindor's name and address to the alleged IP address?*
> *A. No."*

The witness therefore has no knowledge that provide insight into Verizon procedures and methods for linking names to IP addresses. Exhibit 19 shows evidence of faulty MediaSentry information and/or faulty Verizon information with regard to linking IP addresses. Page 1 of exhibit 19 shows that:

> *"With regard to an additional eight (8) IP addresses, after diligent searching, Verizon has not located any information in its possession, custody, or control that is responsive to the above-referenced subpoena. No session information exists for the timestamp provided (see Exhibit B)."*

The Verizon response in exhibit 19 is similar to the reports listed above concerning erroneous MediaSentry claims. It is also possible that Verizon procedures and methods are the cause for this misalignment. For instance, an IP spoofing attack, a BGP hijack, or a simple clock skew of a DHCP server could account for the problem of the missing information on eight IP addresses. Such a clock skew would mean all Verizon supplied information is faulty, including the information on IP address 141.155.57.198. One can only speculate on what exactly has happened without further information from both Verizon and MediaSentry. The missing IP addresses on Exhibit 19 prove that the subpoena which allegedly puts blame on Ms. Lindor is flawed.

D) the exhibits contradict the conclusion of copyright violations

The exhibits contradict the conclusion that Mr. Lindor used KaZaA on her computer to distribute copyrighted works. The exhibits show no link between MediaSentry information and wrong doing by Ms. Lindor. The computer of Ms. Lindor is investigated by plaintiffs witness Dr. Jacobson. This investigation found "*no evidence of the KaZaA program*", as stated on the most recent December 2007 document titled "supplemental declaration and expert report" on Page 3 item 17:

> *"I will testify based on the forensics examination of the hard drive that was copied from the computer owned by the defendant that the computer had no evidence of the KaZaA program nor was there any evidence of the KaZaA program ever being installed on the*

9 Reports from the "UNIversity Security Operations Group" (UNISOG) at https://lists.sans.org/mailman/listinfo/unisog

computer, although the MediaSentry data showed the computer connected to the defendant's Internet account was running the KaZaA program."

As described in the section on "measurements of file sharing networks" it is very difficult to establish links. The lack of KaZaA hard disk evidence means the claim of copyright violations by Ms. Lindor is unfounded.

E) the investigative process has been unprofessional

In my opinion the three reports and deposition by witness Dr. Jacobson indicate that the investigative process had the following characteristics:

a) alternative explanations where not investigated,

b) no checks where conducted to check the accuracy of finding (potential rate of error),

c) no standards or controls exist,

d) the used methods are self-developed and unpublished,

e) the methods are not peer reviewed and not accepted by the scientific community.

This opinion is based on both the contents of the reports and the following deposition statements.

Plaintiffs witness Dr. Jacobson deposition transcript at page 46 states:

"Q. I'm sorry, I misspoke. Do any of your three reports discuss the possibility of any alternate explanations other than KaZaA appearing on a computer owned by Marie Lindor?

A. No.

Q. Are you, as we sit here, capable of thinking of some alternate explanations?

A. Yes."

and at page 38 it is stated:

"Q. How did you learn your method of determining from the MediaSentry materials whether a particular computer has been used for uploading or downloading copyrighted works?

A. It was a process that I developed.

Q. You developed it on your own?

A. Yes."

page 41 and beyond state:

"Q. Has your method of determining from the MediaSentry materials whether a particular computer has been used for uploading or downloading copyrighted works been tested by any testing body?

A. Not that I have submitted.

Q. Do you know anyone else that is using your method, other than you?

A. Not that I'm aware of.

Q. Has your method of determining through the MediaSentry materials whether a particular computer has been used for uploading or downloading copyrighted works been subjected to any form of peer review?

A. Not that I'm aware of.

Q. Has your method of determining from the MediaSentry materials whether a

computer has been used for uploading or downloading copyrighted works been published?

A. No.

Q. Is there a known rate of error for your method?

A. No.

Q. Is there a potential rate of error?

MR. GABRIEL: Object to the form.

A. I guess there is always a potential of an error.

Q. Do you know of a rate of error?

A. To my process, no.

Q. Are there any standards and controls over what you have done?

A. No.

Q. Have your methods been generally accepted in the scientific community?

A. The process has not been vetted through the scientific community."

Due to the above listed characteristics the investigative process can be regarded as unprofessional.

Materials considered

I have reviewed four written statements of expert witness Dr. Douglas W. Jacobson (April 2006, October 2006, December 2006, and December 2007), the deposition transcript, and exhibits 1 through 19.

Conclusions

The material considered and the review of case material described above shows borderline incompetence of plaintiffs witness Dr. Douglas W. Jacobson and the allegations of copyright violations are not proven.

My qualifications

1. As of Jan 1st 2008 I am the technical & scientific director of a 19 Million Euro research project investigating the next-generation of Peer-to-Peer technology, called P2P-Next. The P2P-Next research is sponsored by a research grant from the European Union under the 7th framework program. P2P-Next includes 22 pan-European partners, such as the Finish national research organization (VTT), the public broadcaster of the UK (BBC), the research unit of the German public broadcasters (IRT), the European Broadcasting Union (EBU), and several large companies.
2. I am an active scientist in the area of Peer-to-Peer technology and regularly present recent advances in this field at scientific conferences and workshops in this area. My scientific publications in the area of Peer-to-Peer technology and resource management have been cited over 600 times[10].
3. As an assistant professor at Delft University of Technology I'm coordinating a group of currently 18 researchers conducting experimental Peer-to-Peer file sharing research. At the time of this writing this team is the worlds largest non-profit group in this area. My complete CV is available

10 http://scholar.google.com/scholar?q=pouwelse

online[11].

Full disclosure

I have been asked by the defending counsel for my opinions on the accuracy of the statements made by Dr. Jacobson. This declaration is made for the standard university fee of 220 Euro per hour plus (travel) expenses.

I Dr. Janis Adriaan Pouwelse, Assistant Professor at Delft University Technology in The Netherlands declare under penalty of perjury that the foregoing is true and correct.

Date: 13 Feb 2008

Signature:

11 http://www.ilrweb.com/viewILRPDF.asp?filename=umg_lindor_070514PouwelseCVREDACTED