

Witness statement of Henk Sips and Johan Pouwelse

Background and expertise

Henk Sips and Johan Pouwelse are employed by Delft University of Technology. Professor Sips leads the Parallel and Distributed Systems (PDS) research group, has decades of experience, and co-authored numerous scientific papers and books in this field. Dr. Pouwelse works within the PDS group and both conducts and coordinates Peer-to-Peer file sharing research.

The PDS group has conducted numerous measurements of Peer-to-Peer (P2P) file sharing systems since 2001. In 2003 onwards, we started the worlds largest measurement of the Bittorrent file sharing network¹. As of June 2004, the Bittorrent network is the largest P2P file sharing network². During our file sharing measurement study we followed over 100,000 files, many of which contained copyrighted works. We followed one file in detail for several months and estimated that over 90,000 people downloaded it illegally. Dr. Pouwelse has appeared before the American Federal Trade Commission to testify as an expert on P2P file sharing.

On June 8th 2005 we were asked to create this statement about P2P file sharing in general, the Kazaa network, and the MediaSentry findings under standard university contract work rules.

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.

For example, the telephony software from Skype.com can be used to call ordinary telephones in the Argentine and Australia from The Netherlands at a rate of roughly one Euro per hour. This software only requires an Internet PC, microphone, and speakers. The Skype software exploits P2P technology to reduce telephony costs significantly. The paradigm shift of using The Internet and P2P technology for costs reduction is changing the entire telephony industry. Numerous companies are at risk of being put out of business³.

P2P file sharing is both controversial and popular. File sharing means connecting millions of computer hard disks together into a single network. Roughly two thirds of *all Internet traffic* consists of P2P file sharing traffic⁴. According to a study by Intomart GFK, more than half of the Dutch broadband users have used file sharing to download content⁵. 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 free⁶. On the other side, artists are also using P2P to bypass them. 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 sales⁷.

Measurements of file sharing networks

File sharing networks are difficult to measure. Only a few companies and

1http://www.isa.its.tudelft.nl/~pouwelse/Bittorrent_Measurements_6pages.pdf

2<http://www.cachelogic.com/research/slide9.php>

3http://www.theregister.co.uk/2005/01/10/skype_euro_telcos/

4<http://www.cachelogic.com/research/slide12.php>

5http://www.muzyiek-en-beeld.nl/Live_2005_juni/

6<http://money.cnn.com/2005/05/25/technology/piracy/>

7www.unc.edu/~cigar/papers/FileSharing_March2004.pdf

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 severely 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.

The Kazaa file sharing system

Only one detailed study has been conducted of the Kazaa file sharing network⁸. 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 the pollution in Kazaa⁹. Pollution refers to meaningless files and mismatches between filenames and their actual content. Kazaa was found to be severely 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 system¹⁰. 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 queries¹¹. 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 lying, high pollution levels, and multi-peer downloading it is nearly impossible to obtain solid evidence and detailed checks are therefore required.

8<http://cis.poly.edu/~ross/papers/KazaaOverlay.pdf>

9<http://cis.poly.edu/~ross/papers/pollution.pdf>

10<http://www.hpl.hp.com/news/2002/apr-jun/kazaa.html>

11<http://www.zeropaid.com/news/articles/auto/08262003a.php>

We believe that the following procedure takes the necessary precautions when trying to establish if a user is making copyrighted works available for download.

- Collect filenames by searching the network using keywords.
- Filter out polluted files by checking the actual content.
- 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.
- Investigate if the computer is possibly highjacked. Check if a computer is cracked, for instance, running an open proxy or a hacked Microsoft Internet connection sharing application. A thorough measurement would check if there is a significant difference in traceroute timings and Kazaa protocol response times.
- Track this computer for several days if it does not go offline for reliable IP-address translation by an ISP.

Approach of MediaSentry

The technical information provided by MediaSentry is limited and their measurement procedure is simplistic. MediaSentry did not conduct a thorough investigation such as outlined above to provide evidence of infringement.

The statement from Tom Mizzone hints in item 27 that they systematically searched the Kazaa network for certain keywords, by means of modified Kazaa software. How they resolved relevant technical problems such as superpeer hopping, NAT translation, and firewall relaying by Kazaa is unclear.

In item 28, it is stated that no actual complete file transfer took place; It was only initiated at this stage. Item 30 again hints that MediaSentry simply took filenames at face value and did not mention any correction for pollution on Kazaa. Pollution levels can be as high as 90% for some files.

Item 33 indicates that MediaSentry has no knowledge of the limitations of Kazaa in file searching. Not many of the 2,499,121 users online would be able to see the mentioned 736 files. Reliable global searching in P2P file sharing networks is still an unsolved problem. Only users connected to the same Kazaa Superpeer are *guaranteed* to see these files when Kazaa operates properly (roughly 100 to 150 users as measured by Prof. Keith Ross).

Item 36 states that no computer hygiene precautions were taken. The collected evidence of the spacemansam@KaZaA alias clearly contains multi-peer downloading contamination. Therefore, it is difficult to establish the contribution of the various IP-addresses. It is possible that some IP-addresses contributed 0 Bytes to an actual download, thus there was only involvement and no actual contribution.

Delft, 10 Juni 2005,

Prof. Dr. Ir. H.J. Sips

Dr. Ir. J.A. Pouwelse