available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/diin

Digital Investigation

Identifying an existing file via KaZaA artefacts☆

Paul Sanderson

Sanderson Forensics Ltd.

ARTICLE INFO

Article history:
Accepted 9 August 2006

Keywords:
KaZaA
P2P
FastTrack
MD5
Hash
Peer to peer

ABSTRACT

This child exploitation case study details two newly developed techniques for investigating illegal activity on peer-to-peer networks (names have been changed as the case is ongoing). The defendant in this case is accused of taking three indecent digital video clips of a minor, and sharing these on the KaZaA peer-to-peer file sharing network. The first search method focuses on post-mortem analysis of a forensic duplicate of the defendant's hard drive. The second method employs techniques for live searching of the peer-to-peer network. Tools used to automate these searches are presented. Although the tools and techniques developed for this investigation did not locate additional evidence, they are valid and may be useful in future investigations.

1. Case background

In August 2004, Mr. Jones contacted police to report that his son was sexually assaulted by a family friend, Mr. Smith. It is alleged that in July 2004, whilst on holiday that Mr. Smith made a number of sexual advances on Richard. Richard claims that Mr. Smith took photographs of him whilst in bed and had conversations relating to 'gay' sex and that Mr. Smith performed solo sexual activities in Richard's presence. Richard's sisters confirm the allegations of photograph taking.

During interviews in August 2005, Mr. Smith confirmed that he took a picture of Richard, but claims that he deleted them straight away. He also admits to the solo sexual activities but claims that at the time of the activities he believed that Richard was asleep.

The initial police investigation of Smith's computers revealed a number of indecent images of children. Specifically, they found references to three digital videos with titles "12yo boy 1of3.mpg", "12yo boy 2of3.mpg" and "12yo boy 3of3.mpg" in the KaZaA databases. In the additional description saved in the KaZaA database there is a reference to the boy's name 'Richard', and the dates associated with these three files are shortly before the initial interview and arrest of the defendant.

Further information was found in chat logs that indicated that Smith had discussed taking the movies of Richard, and once the movies had been taken had informed the third party via chat that Smith would upload them to KaZaA where the other user would subsequently be able to download them.

A search was made by the police computer crime unit to recover one or more of the MPG files from the computers but the search was unsuccessful; however, it is a function of the KaZaA application that if there is an entry in the KaZaA database there must have been an associated file on the relevant media at some time. The computer crime unit of the police force investigating this case approached the author for advice on whether these files could be found by any other method. The approaches and tools developed for this investigation are described in the following sections.

2. Overview of KaZaA

To understand the following it is necessary to first have an understanding of the KaZaA system specifically in relation to the KaZaA hash and the methods used by KaZaA to search for a file by hash and by keywords.

☆ The following is based on a project submitted by the author as part of an MSc in Forensic Computing and Cyber Crime Investigation with the School of Computer Science and Informatics at University College Dublin.

E-mail address: paul@sandersonforensics.com

1742-2876/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.diin.2006.08.011

KaZaA is a peer-to-peer file sharing program used by many millions of people worldwide to share files, or more correctly KaZaA is the name of one of the programs used on this particular machine to connect to a peer-to-peer network known as FastTrack. The FastTrack network is the heart of a vast peer-to-peer system (originally written by the company FastTrack) but now owned by Sharman Networks. Sharman Networks licence the FastTrack peer-to-peer technology to various companies so that they can re-brand it as their own. There are a number of applications using this technology currently KaZaA (Sharman Networks also own the KaZaA application), Grokster and iMesh are the most commonly available. KaZaA Lite is also a very common KaZaA derivative. I will use the term KaZaA, as this was the variant used by the defendant, and FastTrack interchangeably to refer to all of these applications unless otherwise specified, however, the techniques detailed here apply to all of the variants.

Peer-to-peer file sharing systems allow a user of the system to search for files that are on offer by other users of the system and to selectively download files that are found as a result of this search.

2.1. *DBB files*

FastTrack clients maintain a list of files that are currently shared, and files that were previously shared in a collection of database files with a .dbb extension. These files are typically named "data256.dbb," "data1024.dbb," "data2048.dbb," "data4096.dbb", etc., i.e. the numerical portion of the file name is a binary power. The only difference between the different dbb files is that a "data1024.dbb" file can record up to 1024 bytes of data about a particular shared file and likewise a "data4096.dbb" file can record up to 4096 bytes of data, and so on.

Files stored on a computers' hard disk are normally referenced by file name only, the owner of the file would normally be aware of the content of the file from the file name although files that are incorrectly named are not unknown. However, FastTrack allows users to add additional descriptions and lists of keywords to the file (known as meta-data) thus helping other FastTrack users to more easily find content that they are interested in. Most of the additional "meta-data" stored with a file is displayed across the main KaZaA display and is usually much more than can easily be displayed on one screen. This information is also recorded in the dbb files.

The format of a DBB file is reasonably well understood (see http://members.home.nl/frejon/55/ft/KazaaFileFormats.htm#dbbformat). This has permitted the coding of certain applications to view this data. Along with the file name, file hash, and file size the dbb files record all of the meta-data associated with a particular entry, including title, keywords, artist, and description. The dbb files can be decode using KaZAlyser (www.sandersonforensics.com) or Encase (www.encase.com). KaZAlyser was used by the police as part of their initial investigation.

2.2. *Download*.dat files*

FastTrack also supports interrupted downloads. If a "LocalUser" is downloading a file from one or more "RemoteUsers" of the FastTrack network and for some reason the downloading LocalUser decides to shut down his/her computer, FastTrack will allow the download to continue next time the LocalUser starts up their computer and connects to the internet. To support this feature when a LocalUser selects a file to download the first thing that FastTrack based programs do is to create a temporary download file (usually named "downloadxxxxxxxxxxxx.dat" where the x's represent a unique number) that records, amongst other information, where the file is being downloads from (i.e. a list of RemoteUsers), what the final file name of the file will be, the keywords associated with the file and most importantly a digital signature (hash) that uniquely identifies the particular file. As the file download progresses the actual data for the file is added to the front of the temporary file and the details of where the file is being downloaded from (the IP addresses), how much has been downloaded etc. is continually updated at the end of the temporary file. When the download is completed the file is renamed to its proper descriptive file name.

Another major feature of KaZaA is that of swarming downloads. Swarming downloads are downloads were the file that is being download is obtained from different source/remote peer computers. Using this technique allows a KaZaA user with a large bandwidth to obtain a file from multiple different users with lesser bandwidth without noticing any significant performance degradation. To facilitate this function KaZaA records the IP and port number (and some additional information) of each user that it knows has a copy of the file that is being downloaded in the download.dat file. It is not uncommon to see 20 sets of IP and port numbers stored in the trailer information at the end of a file that is being downloaded. The format of the download.dat file is detailed at http://members.home.nl/frejon/55/ft/KazaaFileFormats.html#datformat.

2.3. *The FastTrack hash*

To identify multiple copies of the same file with potentially different files names on the KaZaA network, a hash is calculated for each file.[1] The hash is a combination of the MD5 of the first 307,200 bytes of the file and a CRC calculated across regular portions of the entire file. Specifically, the MD5 portion of the KaZaA hash is calculated across the first 307,200 bytes of the file (or the entire file if less than this). The CRC portion of the KaZaA hash is calculated across the first 307,200 bytes of each binary Mega Byte (1,048,576 bytes), i.e. from 1 MB for 307,200 bytes, 2 MB for 307,200 bytes, 4 MB for 307,200 bytes, 8 MB for 307,200 bytes, etc. and the final 307,200 bytes. Graphically this can be represented, for an example 8.7 MB file as shown in Fig. 1.

The dark 307,200 byte block in Fig. 1 is the portion of the file upon which the MD5 hash is calculated, and the hatched 307,200 byte blocks are the portions of the file on which a cumulative 32 bit CRC are calculated. The 128 bit MD5 and the 32 bit CRC are concatenated, with the MD5 first, to create a 160 bit, 20 byte, KaZaA hash.

[1] Th KaZaA hash is a digital signature that is usually effectively unique to a given file.