

# P2P Marshal™: Automatic Extraction of Peer-to-Peer Data

Dr. Frank Adelstein, Dr. Robert A. Joyce, Mr. Judson Powers



ATC-NY

33 Thornwood Drive, Suite 500  
Ithaca, NY 14850

## Abstract:

Generally, digital forensic investigators find peer-to-peer, or file sharing, software present on the computers, or the images of the disks, that they examine. Investigators must first determine what P2P software is present and where the associated information is stored, retrieve the information from the appropriate directories, and *then* analyze the results. P2P Marshal™ is a tool that automatically detects and analyzes peer-to-peer client use on a disk. The tool automates what is currently a manual and labor-intensive process. It determines what clients are currently installed (or have been installed) on a machine, and then extracts per-user usage information for each client, including lists of shared or downloaded files and peer-servers contacted. P2P Marshal was designed to perform its actions in a forensically sound way, including maintaining a detailed audit trail of all actions performed. This paper describes the general design and features of P2P Marshal.

**Keywords:** peer-to-peer, P2P, forensics, LimeWire

## 1 Introduction

The FBI defines *cybercrimes* as actions by “people who use the Internet and computers to illegally penetrate business and government computer systems, including stealing trade secrets and intellectual property, trafficking in child pornography, enticing children from the safety of their homes, and attacking critical infrastructure such as computer networks and power grids” [2]. In their 2006 Annual Report, the FBI classifies crimes into 8 types: terrorism, counterintelligence, cybercrimes, public corruption, civil rights, organized crime, white collar crime, and major theft/violent crime. Cybercrime is the number one category of crime investigated at 11 of the 13 FBI Regional Computer Forensic Labs (RCFLs), and is number two at the remaining two labs. Further, specifically, child pornography or exploitation comprises 35% of the cases of the Philadelphia RCFL, 38% of the cases for the Rocky Mountain RCFL, 51% of the cases for the Intermountain West RCFL, and 65% of the cases for the Western New York RCFL [2].

Often, peer-to-peer (P2P) file sharing networks are widely used in these crimes, and represent a significant source of evidence on computers that are under investigation. Of particular interest to

investigators are the configuration parameters (user name, password, peers/servers used), times of use, time of install, log files of any transactions, and the downloaded (or shared) files themselves. Currently, an investigator must gather, categorize, and analyze all of this information by hand. This typically requires the investigator to research the specific P2P software to determine the location on the disk where the software stores files, the names of configuration files, and their content. In addition, the investigator may need to obtain some secondary software (beyond the investigator's normal tools) that translates a log or cache file into a human-readable format. Clearly, this is a time consuming process, can yield inconsistencies, and can result in problems with the forensic integrity of the examination. Investigators need tools that automate this process and bundle together all of the information relating to each P2P network.

Currently there are a few dozen networks and several dozen P2P programs in general use on the Internet. While a small handful of programs comprise the majority of P2P usage, each program is slightly different. Having a customizable tool that processes evidence from many different P2P clients would be of great benefit to investigators. Few tools exist for examining P2P systems. KaZAlyser [1], probably the best known, analyzes FastTrack-based systems, such as KaZaA, iMesh, and Grokster. However, it is not extensible to non-FastTrack systems, and analyzes the database files generated by the P2P client *after* they have been found. It does not determine what clients have been used. General purpose forensic tools like EnCase can be extended through scripting, but are not designed to analyze P2P evidence and cannot easily parse P2P systems' configuration files or database formats. In general, most P2P analysis is done by hand.

The problem is how this automated extraction and analysis can be done in a cost-effective, yet extensible way. It is essential that the analysis tool must allow knowledge of new P2P software tools to be added through the use of "plug-ins" or configuration files. In addition, the process must be done in a forensically valid way. Specifically, it must give consistent and accurate results for every run. The process cannot be a "black box"; it must be well documented.

## 2 P2P Marshal

*P2P Marshal* is a digital forensic tool for the extraction and analysis of data from peer-to-peer software on client machines. P2P Marshal automates the tedious and time-consuming process of looking for evidence of peer-to-peer usage. P2P Marshal performs these tasks in a forensically valid way and presents them in an easily readable form on-screen and in a format that can easily be incorporated into a report. P2P Marshal's modular, extensible design makes it possible to add extensions for new types of peer-to-peer clients and networks.

P2P Marshal quickly determines what P2P clients were present on a disk image and presents per-user information on those clients, including shared files, downloaded files, and peer servers.

In this section, we describe the overall operation and capabilities of P2P Marshal. The first subsection describes the three modes of operation. The next subsection describes logging and report generation, followed by a description of search capabilities. The registry library is described next, followed by the P2P Marshal user interface and the back-end configuration file.

## 2.1 Phases of Operation

P2P Marshal operates on a mounted disk image. An investigator invokes P2P Marshal, creates an *inquiry*, and starts the analysis. There are three phases to the investigation: discovery, acquisition, and analysis, plus report generation at the end. Figure 1 shows the phases and the information each phase passes to the next.

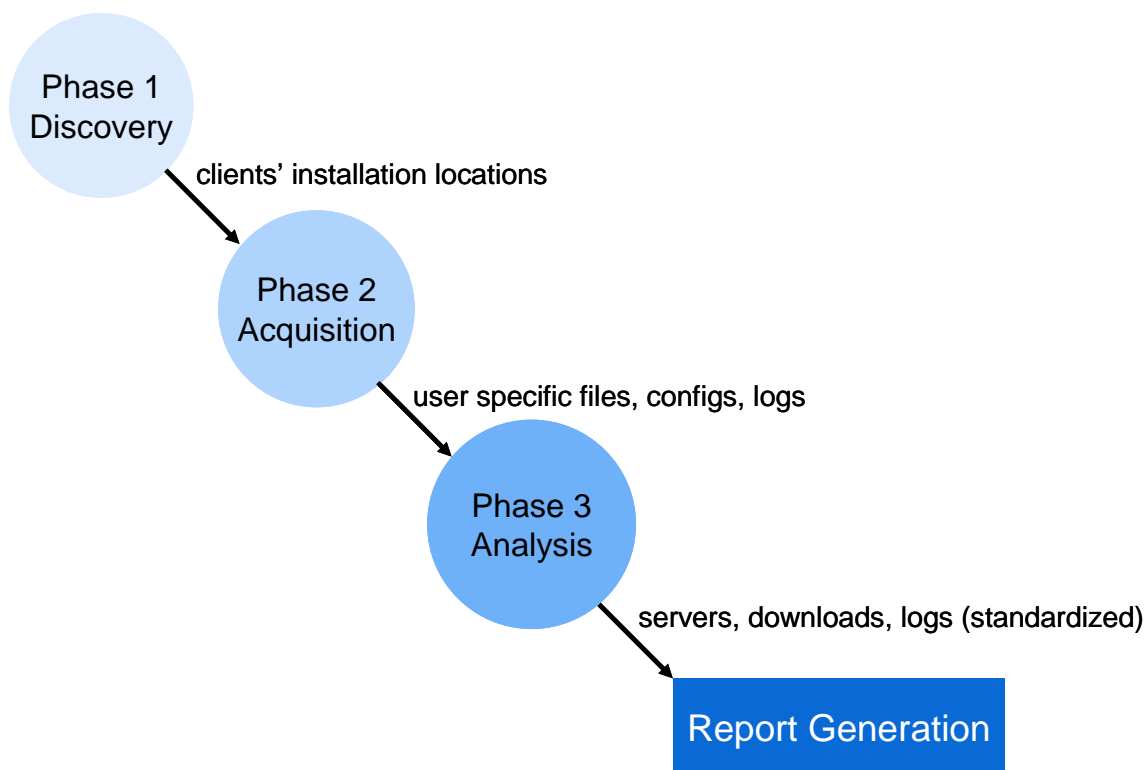


Figure 1. The P2P Marshal investigation process

In the *discovery* phase, P2P Marshal examines the target disk image and determines what peer-to-peer clients are currently, or were previously, installed. To perform this check, P2P Marshal looks for the presence of files, directories, and registry keys and values. Configuration files specify the artifacts that indicate if a particular client was installed. In some cases the programs may have been deleted, but the data directory remains. Registry keys for user preferences may also persist after the user uninstalls the P2P client, or reside in backup versions of the registry generated when the operating system creates a system restore (checkpoint). Files are specified by a pathname. In addition, they can be specified by a hash (currently MD5, but others can be supported). Registry entries can include the (sub)keys, values, and their data.

In the *acquisition* phase, P2P Marshal gathers user-specific usage information for specific P2P clients. For each user, P2P Marshal gathers configuration and log information, including peer or bootstrap servers contacted, files downloaded and shared, and other forensically-relevant data maintained by the specific P2P client. Again, the specific files are defined in the configuration

file. The configuration file lists the Java modules (classes) to be used for parsing; new parsers can be created as needed using a straightforward API.

In the *analysis* phase, P2P Marshal displays the information gathered and allows an investigator to view details (such as the contents of files) and sort data by various fields (IP address, date last contacted, etc.). Investigators can view downloaded files by launching an appropriate viewer (e.g., Acrobat for PDF, Firefox for HTML, Photoshop for an image).

## **2.2 Logging and Report Generation**

P2P Marshal logs all operations it performs. The log file provides very detailed, low-level information on what actions were performed, thus maintaining the forensic integrity of the investigation. The log file provides details on how the back-end tool was invoked, as well as any return or error codes. The audit log is not intended to be easily readable by humans, but rather it allows investigators to verify exactly what actions were taken (and by the same token, what was *not* done) during an investigation, and would be appropriate to be included as an appendix in a report.

P2P Marshal generates a summary report of the findings in a format that can be included in an investigator's report. Supported formats include HTML, PDF, and RTF, so that a P2P Marshal report can be easily inserted into a larger forensics report.

## **2.3 Search Function**

P2P Marshal allows the investigator to search for various usage-specific items (see Figure 2). This includes IP addresses and DNS names of peer servers, names of files, and file hashes (MD5, SHA-1, etc.). For instance, if an investigator wants to trace all contacts with a particular sever, the search function would return all contacts regardless of the P2P client or clients used.

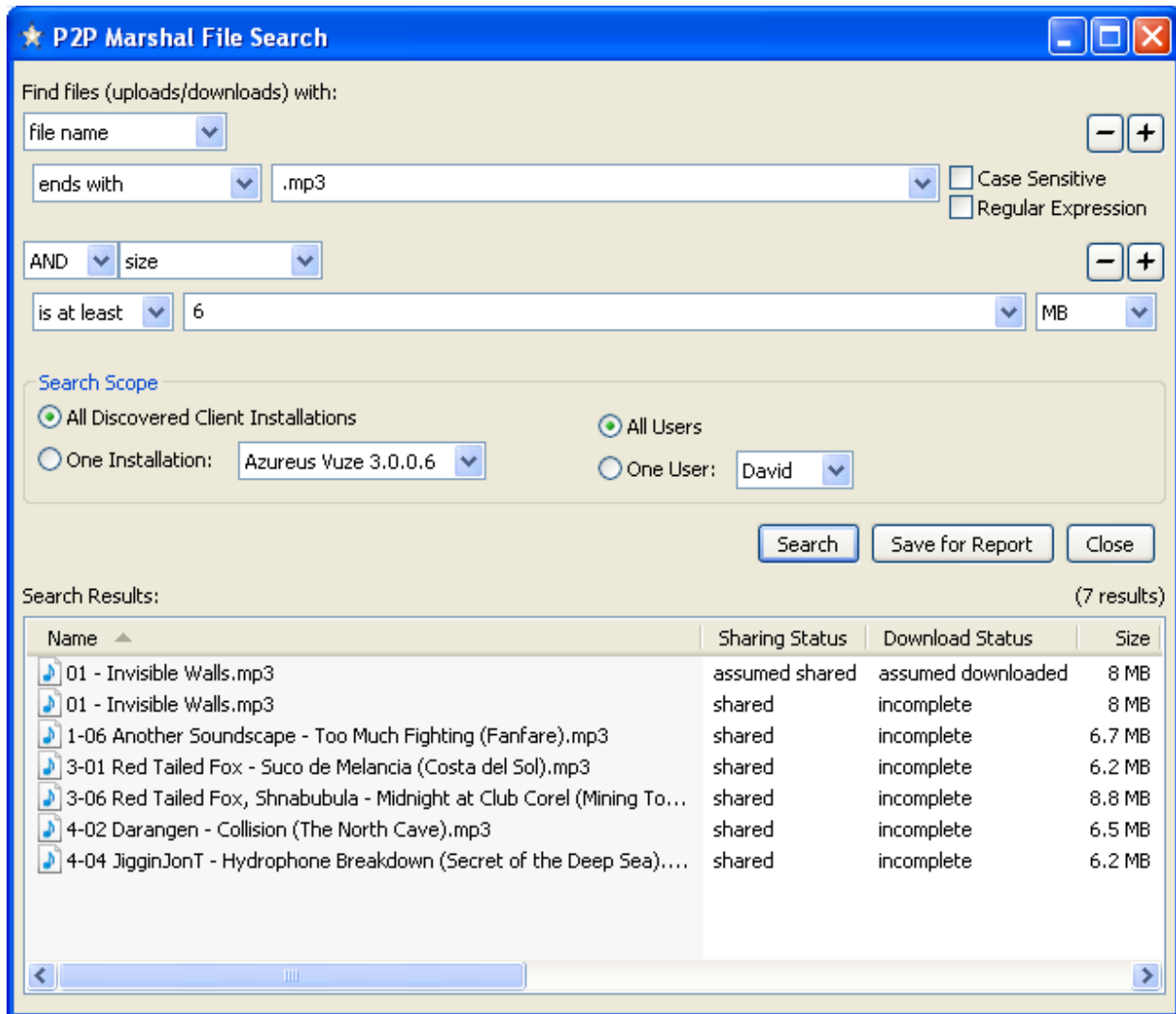


Figure 2. The P2P Marshal search interface

## 2.4 Registry Library

In the discovery phase, P2P Marshal looks for artifacts indicating that a P2P program has been installed or used. One artifact it examines is the registry. However, because P2P Marshal performs an offline analysis of static registry files, there is little support for retrieving keys and values from a file (as opposed to the registry of the running system). Therefore, using documentation of the registry file format found on the Internet, we created our own library to support parsing and interpreting registry files<sup>1</sup>. The library supports enumerating keys (and subkeys), their values, the type associated with each value, and the data content of each value. In addition, it supports directly looking up a key by name. It also supports retrieving the mtime (modification time) and the access control list (ACL) associated with a key.

<sup>1</sup> Limited information on the format of the registry file exists. We used the information from the URL <http://home.eunet.no/pnordahl/ntpasswd/WinReg.txt> as a source and verified the results using regedit.

## 2.5 User Interface

The P2P Marshal user interface, shown in Figure 3, presents information about each P2P client it detects<sup>2</sup>. Within each tab (one tab per client), it presents information specific to each user account in the disk image that has evidence relating to using that client. In the example, six client tabs are shown (Azureus Vuze, LimeWire, Google Hello, Ares, uTorrent, and BitTorrent), with the Azureus Vuze tab selected.

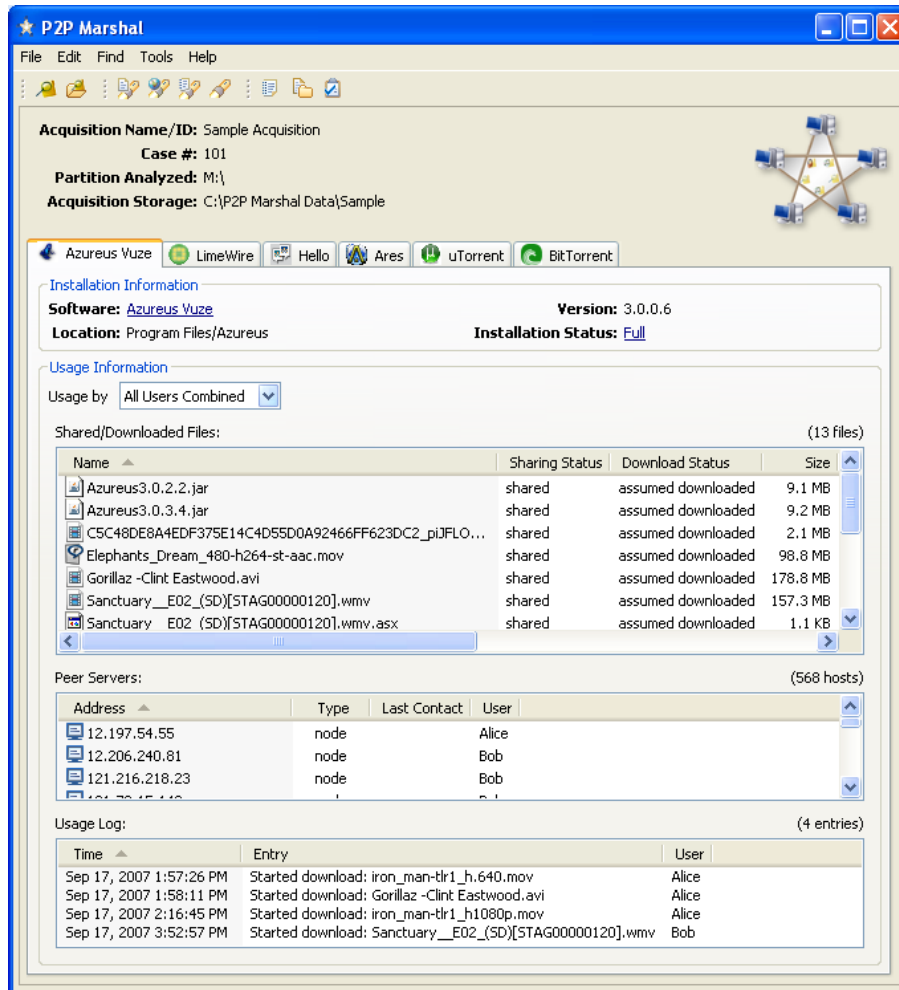


Figure 3. The P2P Marshal user interface

The installation information provides details about where the client was installed, what version, and whether it is a full or partial installation. Partial installation indicates that a P2P client has been on the system but has been (at least) partially removed. In addition, a web page link provides more information about the client when clicked.

<sup>2</sup> Note: The figure shows a test case in which a number of P2P clients were used to download legal content from public sites.

The usage section describes how that client was used by specific users. A pull-down menu allows the investigator to select individual users or “All users combined” to view all P2P activity on the disk image. At the bottom of the window, three tables provide summary information on peer servers, shared files, and log entries.

### 3 Current Status

Version 1 of P2P Marshal was released in early 2008 and is currently available at no charge from [www.p2pmarshal.com](http://www.p2pmarshal.com). A one-day training class on peer-to-peer forensics and how to use P2P Marshal is also available (<http://p2pforensicstraining.com>).

#### Acknowledgement

The authors appreciate the feedback of Mr. James P. Thompson S.C.E.R.S., Assistant Director, Broome County Government Security Division Computer Analysis & Technical Services Unit, who served as our law enforcement partner during the development of P2P Marshal.

This project was supported by Award No. 2006-DN-BX-K013 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication/program/exhibition are those of the author(s) and do not necessarily reflect the views of the Department of Justice.

### 4 Bibliography

[1] KaZAlyser, Sanderson Forensics, available at:  
<http://www.sandersonforensics.com/kazalyser.htm>.

[2] “RCFL Program Annual Report for Fiscal Year 2006,” US Department of Justice, Federal Bureau of Investigation, downloaded on May 16, 2007, from:  
[http://www.rcfl.gov/Downloads/Documents/RCFL\\_Nat\\_Annual06.pdf](http://www.rcfl.gov/Downloads/Documents/RCFL_Nat_Annual06.pdf).

#### Author Biographies:

Dr. Frank Adelstein is the Technical Director of Computer Security at ATC-NY, and provides oversight and guidance to projects at ATC-NY relating to computer security. His areas of expertise include digital forensics, intrusion detection, networking, and wireless systems. He has co-authored a book on mobile and pervasive computing. He received his GIAC Certified Forensic Analyst certification in 2004. A recent research focus is in the area of live forensics. He was the Principal Investigator on a project that resulted in the OnLine Digital Forensic Suite™, a live forensics tool. Dr. Adelstein is the vice-chair of the Digital Forensics Research Workshop.

Dr. Robert A. Joyce is the Technical Director of Information Management at ATC-NY. His research interests include distributed information storage and transformation, computer forensics, image and video processing, network and media security, visualization and design, and human-computer interaction. Since joining ATC-NY in 2002, he has led several research and development efforts in the area of information management and has made significant

contributions to many other projects within the organization. Dr. Joyce was a substantial contributor to the development of the OnLine Digital Forensic Suite™.

Mr. Judson Powers, Computer Scientist at ATC-NY, served as lead developer of P2P Marshal™.