Forensic Studies in BitTorrent

Optimising file downloads is a dream for some; for others it raises challenging piracy issues.

BY JAMIE ACORN AND JOHN AUSTIN

Royal Holloway
University of London

Produced by the Information Security Group at Royal Holloway, University of London in conjunction with TechTarget.

INTRODUCTION

Given the volumes of recorded and transmitted data in today's computerised environment, the collection and processing of digital evidence is an even more delicate and complex business than in years past. Bram Cohen's creation of BitTorrent in 2001 enables and optimises the downloading of files, any size and any type, to remote computers. Now available for all common operating systems, its popularity is soaring and is available in many browsers including Mozilla Firefox. For some it is a dream; for others, including those trying to combat the unauthorised downloading or 'pirating' of copyright material, it poses yet another significant problem.

Jamie Acorn

Information Security Group, Royal Holloway, University of London, Egham, Surrey, U.K.

John Austin

Information Security Group, Royal Holloway, University of London, Egham, Surrey, U.K

Forensic Studies in BitTorrent

Given the ever-growing use of BitTorrent as a means of file sharing, and the associated costs to media-based industries and legal issues, the need for a forensic understanding of this system of file sharing is becoming increasingly important. To date, there are no (known) published studies investigating the forensic aspects of BitTorrent. This study is a preliminary investigation into the forensic artefacts created by BitTorrent use.

'BitTorrent' is a peer-to-peer application that uses metadata files known as torrents. The metadata provides instructions to a BitTorrent client, facilitating the connection to remote computers and the downloading of files (of any size and type).

The diagrams shown *(next page)* are taken from and depict how the BitTorrent protocol works. Firstly, an individual creates a torrent using either a BitTorrent client or torrent-making application, and publishes it on a website or forum. This individual is known as the 'initial seeder'. Figure 1 shows the 'initial seeder' distributing fragments of a file to different machines connected using a BitTorrent client. It is usual for the shared file to be virtually split into many smaller chunks

This article was prepared by students and staff involved with the award-winning M.Sc. in Information Security offered by the Information Security Group at Royal Holloway, University of London. The student was judged to have produced an outstanding M.Sc. thesis on a business-related topic. The full thesis is available as a technical report on the Royal Holloway website http://www.ma.rhul.ac.uk/tech.

For more information about the Information Security Group at Royal Holloway or on the M.Sc. in Information Security, please visit http://www.isg.rhul.ac.uk.

Figures 1 – 4: The BitTorrent file sharing process

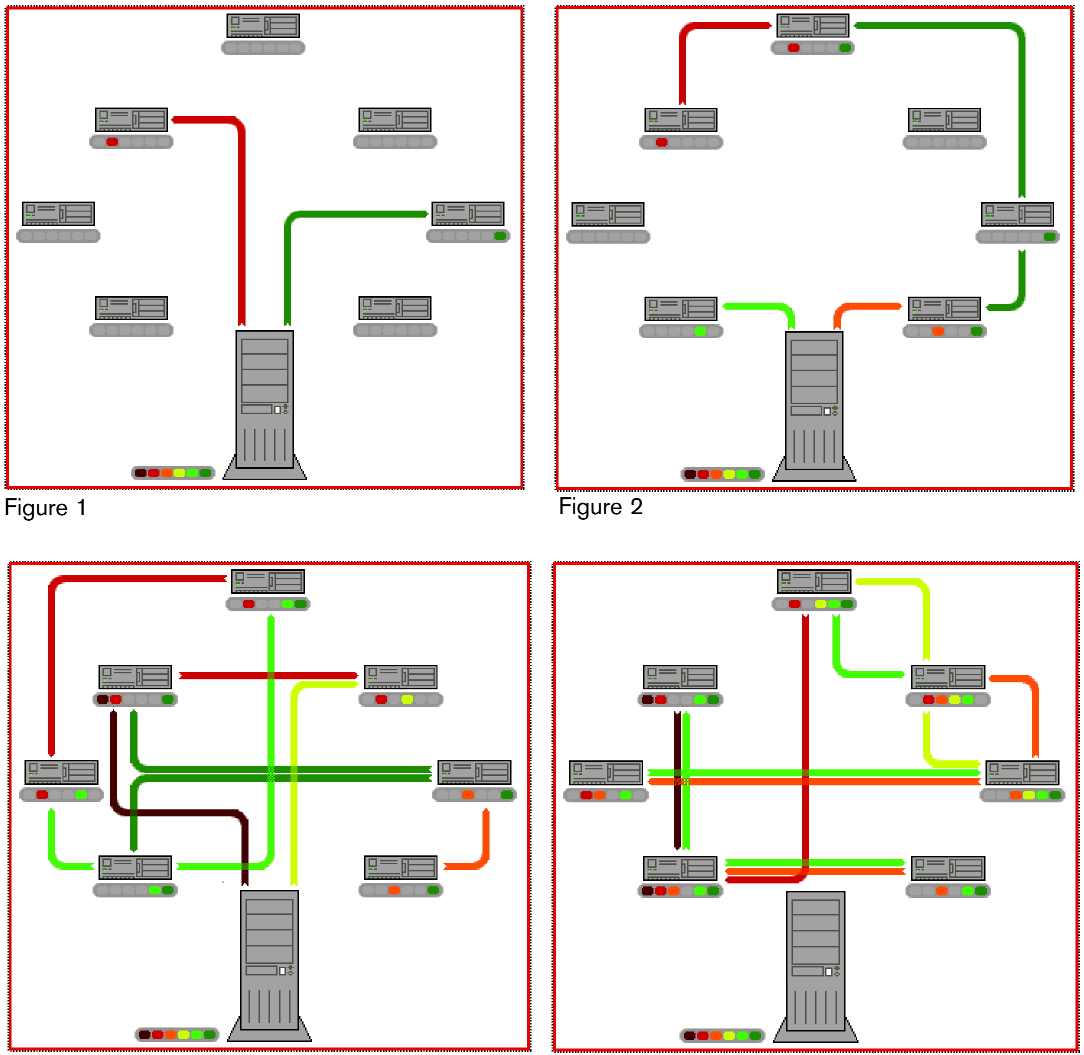

Figure 1

Figure 2

Figure 3

Figure 4

SOURCE: Wikimedia Foundation, Inc. (September, 2007): BitTorrent, http://en.wikipedia.org/wiki/BitTorrent

of data of equal size to aid file transfer (it is not always possible to divide the file equally and therefore the last chunk may be truncated). The connected individuals are collectively known as a 'swarm'.

Figures 2 and 3 show the swarm sharing data chunks between each other as well as the 'initial seeder'. Figure 4 shows the point where the 'initial seeder' has shared all the small data chunks and now no longer needs to seed these files. The individuals that form the swarm now possess the sum of all parts of the file being shared. The swarm will continue to share data with each other and any newly connected individuals. The beauty of this protocol is that an 'initial seeder' only needs to share each data chunk once in order for the file to be shared with many individuals and this means the initial seeder's bandwidth is not being constantly depleted by the people they are sharing files with.

The instruction information contained in torrent files are essentially: the name of the file to be shared, the size of each piece and the number of pieces that make up the file, and the Uniform Resource Locater (URL) of a tracker. A 'tracker' is a dedicated server that links all the peers (remotely connected computers) associated with a particular torrent file. Remote individuals download the torrent from the website. When the .torrent file is opened within a BitTorrent client, it points the client to the 'initial seeder' using the tracker URL.

The effectiveness of the protocol relies on every individual sharing pieces of the file they are downloading; hence, while an individual is downloading pieces of the file, they are also uploading or seeding the pieces of the file they already have. It is possible to prevent seeding by changing preferences within the BitTorrent client but trackers and individuals will ban these users or limit their download speed. Thus, it is the general rule that, by using BitTorrent to download files, the user is also sharing files.

The specific aims of the study were as follows:

1. To identify forensic artefacts produced by BitTorrent file sharing, and to establish whether the artefacts lead to identification of the downloaded or shared files.
2. To identify any settings that within client configuration files which may be useful to aid forensic examination.
3. To identify any artefacts that determine IP addresses of remote computers from which data was downloaded, or shared, during the test phase.
4. To identify whether any of the torrents had been created and seeded by the user.

Five BitTorrent clients (ABC, Azureus, Bitcomet, BitTornado, and uTorrent) were selected for testing as these were determined to be the most 'popular' at the time of this study. A number of torrents were selected for download and then different scenarios created to emulate normal usage, such as stopping a torrent during the download, removing a torrent from the client during download, completing a full download and letting the torrent seed). Torrent files were also created by each client (except BitTornado), linked to a public tracker and left to seed until the files seeded were completely uploaded. Each client was then analysed using forensic software on generated image files and also in situ.

All the clients tested have a 'settings or preferences' function where the user can tweak operation configurations. Analysis of the clients showed that they vary in complexity and operability and thus varied in the amount of useful forensic information stored in the settings. However, investigating the settings of a client is key to understanding

It is possible to prevent seeding by changing preferences within the BitTorrent client but trackers and individuals will ban these users or limit their download speed.

how it has been used, as the settings can determine information such as:

- where downloads will be stored,
- if default settings have been altered,
- where torrents will be stored, deletion settings,
- if logging is enabled,
- if a password has been set,
- the version of the client used,
- the ports used,
- the last time the client was used,
- the seeding settings etc.

Thus analysis of the various settings can be used to form a profile of a user and to distinguish a zealous user from a recreational user.

TORRENT FILES

Torrent files are a fundamental component of the BitTorrent file sharing procedure. They are, in effect, pointers to the target files that are to be shared–meaning that there is no difference between a torrent file that is used to share or download a file. The only way to determine what a torrent file has been used for (i.e. to download or share a file) is to investigate artefacts produced by the BitTorrent client used. There are different ways that torrent files may arrive on a computer. A user can create a torrent, and save it anywhere on the system. A user can open a torrent from a website; this causes the torrent file to be saved in the 'Temporary Internet File' folder (if Internet Explorer is used as a web browser). It is also possible a user might save a torrent from a website, email, IRC, external storage device, to any location on the system. With the exception of BitTornado, all clients analysed create a backup of torrent files (these are stored in application specific directories) when they are opened. The backups are direct copies of the torrent files opened. Azureus, uTorrent and ABC store all backup torrent files within their designated directories where they remain stored even after torrents are removed from the GUI. The backup torrent files for these clients can be deleted by the user using the GUI, but not by using the main removal tab. The BitComet client does not continue storing the torrent backups once they are removed from the GUI; hence only backup torrent files currently loaded in the GUI are stored.

Torrent files contain information such as the names and sizes of files that are downloaded or shared. This information can be

Torrent files contain information such as the names and sizes of files that are downloaded or shared.

used as a guide to determine which files may have been downloaded or shared but the presence of a torrent file alone is not evidence of file downloading/sharing; further evidence would have to be gathered showing that the torrent has been opened within a client. As previously discussed, backup torrents are created (within specific application data directories) when torrents are opened in the clients (except BitTornado) and this is evidence of intention to download or share files.

CACHE FILES

All the clients generated files containing data regarding the 'state' of downloads. These files are used as a recovery system so that torrents can resume from the same point if the program is interrupted (either by being stopped or when the client is closed). These 'cache' files can provide informative data such as:

- the directory where downloads are saved
- the amount of data downloaded or uploaded for specific torrent files
- the time and date torrents were started or stopped
- the status of the torrent (i.e. complete, downloaded/seeding, or stopped)

These 'cache' files, thus, provide evidence for the downloading or seeding of specific files.

The ABC client stores a file containing information on every torrent that has ever been opened within the client; thus a complete picture of the downloading history of ABC can be obtained. Similarly, the BitComet client can provide a complete downloading history by analysing data entries within an application data file, but this data is only stored for every torrent file opened and 'shared'. Users have to select a tab during the torrent opening process in order for the torrent files to be 'shared'. The data for each torrent remains within the system file as long as the corresponding torrent files are not deleted from BitComet using 'Delete Task and Downloaded Files' option. No other BitTorrent client produced artefacts that reveal a complete history of torrents downloaded and opened.

Torrents that have been created and seeded produced identifiable artefacts in the ABC, BitComet, and uTorrent clients. The analysis of files generated by these clients identified data that defined the torrents that

Torrents that have been created and seeded produced identifiable artefacts in the ABC, BitComet, and uTorrent clients.

are used as 'initial seeders' to share file data. Such evidence can be used to substantiate file sharing. It should be noted that this evidential data identified in the BitComet and uTorrent artefacts only relates to torrents currently loaded in their respective clients.

Analysis identified that it was possible to acquire the IP addresses of connected peers for each torrent currently loaded in the client, but only for uTorrent, Azureus, and BitComet. The type of peer information that was stored differed between these clients. The IP addresses, the amount of data exchanged, and the direction of exchange can be acquired from Bitcomet, but only for the point in time at which the torrent was last stopped/closed.

Azureus and uTorrent, on the other hand, store a list of connected peers that are recorded throughout the duration of download or upload. It was ascertained that IP addresses stored in uTorrent and Azureus were not just stored from the moment of torrent closure, as IP addresses were retrieved from relative files corresponding to torrents that had finished downloading and were no longer connected to peers.

Further tests would need to be conducted in order to determine whether every IP address connected becomes stored in the areas identified during analysis. No data artefacts relating to 'peer' information were discovered when analysing the ABC and BitTornado clients. Concretely identifying peers that have downloaded, or shared, illegal or confidential material is valuable intelligence. Internet service providers can be sanctioned to release the details of individuals that pertain to the IP addresses identified. Further investigations can then transpire.

EMULATION

'Emulation' is a useful tool in forensics as it provides a quick approach to understanding how applications work, and it can also be used to visually display information on the computer as seen by the user. The uTorrent application is a self-contained application and does not install program files to the computer. This property makes it easy to emulate. Files stored in the directory 'C:\Documents and Settings\<user>\Application Data\utorrent\' can be extracted and loaded into the same directory on another computer (with no internet connection).

Running the uTorrent application will begin

Concretely identifying peers that have downloaded, or shared, illegal or confidential material is valuable intelligence.

the emulation and display the 'state' of file sharing as it was when uTorrent was last closed. Analysis of the emulated uTorrent provided information regarding the number of times uTorrent had been opened, the total amount of data uploaded and downloaded, the number of added torrents and the total time that the application ran for. This information had not previously been observed via the 'normal' method of analysis. Emulating other clients may be possible using the same technique and could result in data not previously found by the usual analysis techniques.

Analysis of registry files of BitTorrent activity for each client produced many artefacts, although these artefacts are of little evidential value. They basically identify that a client has been run on the machine. Artefacts created within application files also show this information and much more, for example, time and date data references are provided. The most significant data discovered in the registry was identified in the Bitcomet sub key 'HKEY_CURRENT_USER\Software\BitComet\BitComet'. This key contains a record of the website URL from which the last torrent was downloaded and opened, and the 'title' of the web page (as seen in the source code of the website). This is valuable as web page information cannot be found from analysing BitComet specific files, and details of the last torrent opened within BitComet are not explicitly stored in any file artefact. Two other useful registry keys provide information regarding torrent files. Creations of these keys are a result of actions pertaining to torrent files rather then the BitTorrent client. The registry keys:

'HKEY_LOCAL_MACHINE\Software\Microsoft\Windows\CurrentVersion\Explorer\ComDlg32\OpenSaveMRU\torrent' and HKEY_CURRENT_USER\Software\Microsoft\Windows\CurrentVersion\Explorer\RecentDocs\.torrent' contain information on the names and paths of the files most recently saved or copied, or opened torrents. This information could be of value in circumstances where backup torrent files have been removed from the client.

The development of BitTorrent client applications is still continuing. Newer versions of UTorrent, BitComet and Azureus have been made available for download since the time testing began. The sheer number of available clients means their popularity may change over time, and indeed new clients may be developed thus, further forensic tests will need to be carried out.*

Ron Condon

UK bureau chief
searchsecurity.co.UK

Ron Condon has been writing about developments in the IT industry for more than 30 years. In that time, he has charted the evolution from big mainframes, to minicomputers and PCs in the 1980s, and the rise of the Internet over the last decade or so. In recent years he has specialized in information security. He has edited daily, weekly and monthly publications, and has written for national and regional newspapers, in Europe and the U.S.