Blossom—Hands-on exercises for computer forensics and security

File Content Analysis

BLOSSOM
Manchester Metropolitan University
(Funded by Higher Education Academy)
l.han@mmu.ac.uk

1. Learning Objectives

This lab aims to conduct the forensic analysis of file content.

2. Preparation

1) Under Linux environment

2) Some files that you will need from /home/user/BlossomFiles/FileContentAnalysis:

   - 'FileContentAnalysisLabFiles.ZIP'

3) Some documents that you may need to refer to:

   - 'Virtual-MachineGuide.pdf'
   - 'Linux-Guide.pdf'
   - 'BLOSSOM-UserGuide.pdf'

3. Tasks

Setup & Installation:

- Start a single virtual machine as you have done with previous exercises (see Virtual Machine Guide)

  ```
  # kvm -cdrom /var/tmp/BlossomFiles/blossom-0.98.iso -m 512 -net nic,macaddr=52:54:00:12:34:57 -net vde -name node-one
  ```

- Unzip the 'FileContentAnalysisLabFiles.ZIP' file:

  ```
  # unzip FileContentAnalysisLabFiles.ZIP
  ```

Task 1 File Content Analysis – Content Identification

1.1 Vital information for any forensic investigation can be discovered within the files that a user stores on their computer; however, it is important to take in to account how easily someone can edit the file signatures of certain types of files as a way of obfuscating the true contents of the file. As an example, someone could change the file extension of every AVI file to have the extension .BIN, and then by associating .BIN files with a video player, the information contained within the files would be hidden from normal view.

1.2 The content of files can be identified based on their content by using the command 'file', which makes use of key structures inherent to certain types of files in order to determine a given file's type. Try using the file command on the file '997495.pdf', then change the extension to that of another file format and view it again.

Task 2 File Content Analysis – Hex

2.1 Although commands such as file can be used to verify the content of a file, it should not be relied upon in all situations – more importantly, a forensic examiner should be able to confirm correct or incorrect operation of such commands. For this, we must be able to verify the file content manually using a hexadecimal view of the file.

2.2 Simply use hexedit to view the hexadecimal format of the file '997495.pdf'. This can reveal information relating to the format of the file, as well as other useful information such as the file type version.

Question: What version of PDF format version is '997495.pdf'

Task 3 File Content Analysis – Extract / Hachoir

3.1 The extraction of metadata is another important part of forensic investigations, and for this we will look at two pieces of sotware, Extract and Hachoir-Metadata.

View the help documentation for Extract:

```
#extract -h
```

As can be seen, there are various options that can be used when extracting metadata from a file, some of which are critical to forensic analysis, such as computing hashes of the file in question, or removing duplicate files in order to reduce the width of the file search.

Use extract to view the metadata of the files ‘997423.gz’ and ‘997431.jpg’.

Question: What metadata is discovered from these files that could prove useful when performing a forensic investigation?

3.2 Hachoir-Metadata supports fewer types of files that Extract, but provides other metadata that Extract fails to provide depending on the file format. Use Hachoir-Metadata to view the metadata of the same files we viewed with Extract and take the note of the difference in metadata returned.

Question: What metadata is returned by Hachoir-Metadata, but not by Extract?

Task 4 File Content Analysis – Images

4.1 Images are simple enough; they contain data that is then rendered as a graphic, which can be of multiple different file formats. The potential for metadata extraction from images can range from simple text comments, all the way to location information from where the image was created. To analyse the metadata of an image file, we will be using the ‘identify’ package from ‘imagemagick’ as follows:

```
#identify -verbose 276861.jpg
```

The –verbose option provides extremely detailed information about the image. The result should be an extraordinary amount of information relating to this small image file, including information such as date created and date modified.

4.2 As mentioned earlier, metadata can also reveal information such as GPS Locations. We will use exiftool to view extended metadata from the file ‘808913.jpg’:

```
#exiftool 808913.jpg
```

As a result of this, we will see an extensive set of information relating to such details as the camera make / model, location of the photo, and even the settings and encoding of the image.

Question: Using the GPS co-ordinate metadata located within the file, perform a Google map search in order to locate the rough location of which the photo was taken, whereabouts was the photo taken?

4.3 GIF images are an image format that are primarily used for icons and simple graphics and will usually be created by image editing programs, and it is for this reason that built-in metadata hasn’t really been a requirement meaning that the metadata contained within a GIF file will be very sparse; however, it is still important to know what kind of metadata will be available within a GIF on the off-chance that a GIF file analysis is required.

View the file ‘135900.gif’ within both exiftool and hachoir-metadata and analyse the metadata that is returned.

The other two main types of image formats are PNG and TIFF. PNG is similar to GIF in the sense that metadata available is extremely sparse; however, TIFF images can contain quite a lot of information as it does support internal metadata tags.

Task 5 File Content Analysis - Archives

5.1 Archive files are container files which hold various other files. The metadata contained within these files can provide information relating to the user who supplied the file, as well as system time stamps for when files were added to the container, and information pertaining to the system of origin.

We will use the unzip command to retrieve information from a ZIP archive:

```
#unzip -v BiffView.zip
```

This shows us the dates and times when each file was added to the ZIP archive.

Question: At what dates was the ZIP file modified?

5.2 We can also use other simple tools such as 7zip to view archive information. Use 7zip to view the file contents of 'PuttyHijackV1.0.rar':

```
#7z l PuttyHijackV1.0.rar
```

This will show dates and times of when the files were added in a similar fashion to unzip, which shows that the files were all added from the dates June 2008 to July 2008. This information could be useful if, for example, the RAR archive was found on a compromised system and the extracted EXE file had a modification date of 2007. This would show explicit evidence of file time stamp alterations.

5.3 TAR files are the standard method of archiving and compressing data on a Linux system. There are multiple different tools involved in the process of compressing and archiving a file in this manner. The 'tar' command concatenates multiple files into a single archive, and the tar archive is then compressed using tools such as GZIP or BZIP2. It is because of this, that the files will have two extensions, such as 'Turtle.tar.gz', which means the tar file was compressed using GZIP.

We can use gunzip to view the compression layer:

```
#gunzip --list --verbose Turtle.tar.gz
```

This will show the last modified date, but that is the only useful piece of information available, so we can also drill down a layer and access the tar archive itself:

```
#tar --list --verbose --gunzip --file Turtle.tar.gz
```

Question: What is the name of the owner of the archive, and at what dates did he add files to the archive?

Task 6 File Content Analysis - Documents

6.1 Documents can come in a variety of different formats, such as DOC, DOCX, PDF to name a few, and the potential for forensically valuable information is extremely high. As an example, if a word document did contain suspicious information or had some relation to a suspicious individual, we would need to use metadata to connect it an individual. This can be done using tools such as wvSummary:

```
#wvSummary darknet5.doc
```

This will display metadata information about the file ‘darknet5.doc’.

Question: Who the author of the file, when was it created, and what is the email address of the author?

6.2 PDF Files can store a significant amount of metadata relating to the author of the document, the tool that was used to create it, modification and creation dates, as well as identification for the document itself. These are all extremely critical for forensic investigating and can be viewed in a number of different tools:

```
#exiftool 997495.pdf
```

From this, we can see that the file was created in February 2006, by someone called ‘derwin’, amongst other snippets of information.

6.3 We can also use a program called ‘pdfresurrect’ to display historical information about PDF files:

```
#pdfresurrect -q 025835.pdf
```

This will show us how many previous versions of the PDF file existed, and we can now search this file for information relating to the modification dates:

```
#pdfresurrect -i 025835.pdf | grep ModDate
```

This information could then be used to further delve in to the history of the PDF file, by performing such tasks as recovering the previous versions so that comparisons could be made between each version.

Question: How many times was the file modified, and at what dates? What use could the historical version information be used for?