Introduction to useful digital forensics concepts for solving UniWA-CTF

Michael Kokkos, Alexander Soloviev

Aigaleo - Athens, Greece

Abstract

Nowadays, more than ever digital forensics related skills can be useful for conducting investigations for various different fields, ranging from small cyber security enterprises to law-enforcement and military agencies. An understanding of digital forensics is crucial for any individual that wants to pursuit a career in the cyber security field. This paper introduces individuals whom participate in UniWA-CTF into the field of digital forensics by discussing the topics of file signatures, EXIF metadata and network forensics.

Keywords: Investigation, file signatures, EXIF metadata, packet analysis

1. Introduction

Digital forensics is a branch of forensic science encompassing the recovery and investigation of material found in digital devices, often in relation to computer crime. [1] [2] Such investigations can be conducted by military or law enforcement agencies or by private enterprises and individuals.

The reason that an organization or an individual may want to start a digital forensics investigation can vary. Examples of applications of digital forensics may include cases like a law enforcement agency that needs to find the IP address of a malicious user by analysing log files, a private enterprise that has been infected by a computer virus may want to find if the virus infected the network from inside, or an individual that wants to recover from a hard drive a file that he or she has deleted.

There is a number of different tools and techniques that can be used for conducting a digital forensics investigation. This paper provides the reader with introductory-level knowledge into topics that are involved in the "forensics" category of the challenges of UniWA-CTF. These topics are also heavily used in various beginner level CTFs, as scenarios, that involve these skills, exist in digital forensics investigations.

2. The Magic Bytes

Figure 1: QR code that leads to the "Magic Bytes & Security: When file categorisation goes wrong" video, from the channel "247CTF"

The magic bytes, also called magic numbers or file signatures, are numbers embedded at or near the beginning of a file and they indicate its file format [3]. The magic bytes method is one of the three methods of file type detection, the other two are the extension-based and the content-based method [4].

Users can examine or edit the magic bytes of a file by opening the file with a hex editor. In the figure below are presented the magic bytes of an example JPEG file opened with "hexedit" in a BASH terminal.

```
FF D8 FF E0 00 10 4A 46 49 46 00 01
```

Figure 2: Magic bytes of a JPEG file

It is worth to been said that this magic bytes signature is not the only one that is used for JPEG files, when a user wants to identify the file type, he or she is advised to read an updated list with the various signatures. Numerous lists exist that match file types with their magic bytes, like the one on [5], although these lists are not exhaustive, they usually contain the most common file formats.

3. Metadata

The word "Metadata", is defined as "data that provides information about data" [6]. According to the National Information Standards Organization (NISO), three main types of metadata exist [7]:

- Descriptive metadata: Describe a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords.
- Structural metadata: Indicate how compound objects are put together, for example, how pages are ordered to form chapters or how the text is aligned.
- Administrative metadata: Provide various information that help to manage a resource. These information can include the time and the date that it was created, the file type and the access rights, along with other technical information.

  There are also several subsets of administrative data that exist, like Rights management metadata and Preservation metadata.

Figure 3: QR code that leads to the "Obtain Valuable Data from Images During Recon Using EXIF Extractors" video, from the channel "Null Byte"

Among other uses, metadata can be written into a digital photograph. These metadata can contain various information such as the time and the date that the picture was taken, the manufacturer and the model of the device, GPS coordinates, etc [8].

There are many different standards that are related with digital photography metadata. One of these standards is Exif, which is, briefly, explained below.

3.1. Exchangeable Image File Format

Exchangeable image file format (officially Exif) is a standard that was initially developed by the Japan Electronic Industries Development Association. This standard specifies images and sound formats, along with tags that are used in digital cameras (including these which are embedded in smartphones), and other systems that handle image and sound files recorded by digital still cameras [8].

The process of viewing the Exif metadata is a relatively easy process, regardless of the operating system that is being used. Figure 4 shows how a GNU/Linux user could access the information provided within the Exif metadata, by simply using "exiftool" within the operating system's terminal. In Figure 4 the reader can see the various metadata that exist within a digital picture. These tags are not restrictive, different pictures can have different tags. However the tags that are shown in Figure 4 are usually included in the majority of digital photographs, if these Exif metadata are not erased by the user or the server for privacy reasons or the server that hosts the picture for the reasons of saving network bandwidth and storage.

```
user@linux_pc:~\$ exiftool picture.png
ExifTool Version Number         : 11.88
File Name                       : picture.png
Directory                       : .
File Size                       : 110 kB
File Modification Date/Time     : 2020:03:05 01:12:08+02:00
File Access Date/Time           : 2020:12:29 16:32:49+02:00
File Inode Change Date/Time     : 2020:12:29 16:33:43+02:00
File Permissions                : rw-rw-r--
File Type                       : PNG
File Type Extension             : png
MIME Type                       : image/png
Image Width                     : 1584
Image Height                    : 717
Bit Depth                       : 8
Color Type                      : RGB with Alpha
Compression                     : Deflate/Inflate
Filter                          : Adaptive
Interlace                       : Noninterlaced
Image Size                      : 1584x717
Megapixels                      : 1.1
```

Figure 4: Output of "exiftool" in a GNU/Linux terminal

Figure 4 shows an example of usage of "exiftool", which is a popular command-line tool that extracts the Exif metadata and then prints them as output in the terminal. It is compatible with GNU/Linux, MS Windows and macOS operating systems. This particular file does not contain a lot of metadata but, even in that case, the reader can see the amount of information that can be easily accessed, just by running the particular tool with a digital photograph file, that its extension is supported, as an input.

The same process could be easily done in an MS Windows operating system environment. If the analyst doesn't want to use exiftool, he or she can just right click on the file, then select "Properties", and by clicking on "Details" can extract the same information.

4. Network Forensics

Figure 5: QR code that leads to the "What is Packet Capture?" video, from the channel "NetFort"

Network forensics is a sub-branch of digital forensics relating to the monitoring and analysis of computer network traffic for the purposes of information gathering, legal evidence, or intrusion detection [9].

Different types of network-based evidence exist. All of which have pros and cons with respect to forensic analysis [9]. UniWA-CTF introduces its participants to the topic of Full content data forensic analysis, with a challenge where they are given a "packet capture" (.pcapng) file, which contains every single piece of information that passes across a network.

4.1. Packet Analysis

Packet analysis is a primary traceback technique in network forensics, which, providing that the packet details captured are sufficiently detailed, can play back even the entire network traffic for a particular point in time [10].

Packet analysis can be conducted by using specialized software (e.g. kismet, tcpdump, WireShark etc.) or hardware (Packet capture appliances) that has been developed for this specific purpose. These pieces of software and hardware can provide the user with a user friendly Graphical User Interface (GUI) or a Command Line Interface (CLI) for the more advanced users, and allow the analysis of the data streams as, among their other capabilities, they are capable of sorting data by time, protocols, IP, etc. Within their uses are also real-time capturing of the network traffic to assist, for example, a Security Operations Center.

The user can analyze the packets one by one, and extract a big number of data, like these of IP and MAC addresses, files which were moved on the network, the protocols that were used, etc. These data are crucial in a network forensics investigation.

The presence of a Graphical User Interface is considered an advantage on many cases, however, there are a lot of scenarios were the analyst may need to process manually, with the use of a script. This can be done easier by using command line tools like Wireshark's utility "tshark", tcpdump, etc.

5. Conclusions

Digital forensics is definitely a vast and interesting field, which requires knowledge in many different fields. CTF challenges is a great and entertaining way that can help individuals to acquire those skills and motivate them to deepen their understanding in digital forensics and how computers and computer networks work in general.

References

[1] M. Reith, C. Carr, G. Gunsch, An examination of digital forensic models, Int. J. Digit. EVid. 1 (2002).

[2] B. D. Carrier, Defining digital forensic examination and analysis tool using abstraction layers, Int. J. Digit. EVid. 1 (2003).

[3] The Linux Information Project, 2006, Magic number definition, URL: http://www.linfo.org/magic_number.html.

[4] Meghanathan, N., Boumerdassi, S., Chaki, N., Nagamalai, D. (Eds.), Recent Trends in Network Security and Applications (International Conferences, NeCoM 2010, WiMoN 2010, WeST 2010,Chennai, India, July 23-25, 2010. Proceedings), volume 90 of *Communications in Computer and Information Science*, Springer-Verlag Berlin Heidelberg, 2010. doi:10.1007/978-3-642-14493-6.

[5] G. Kessler, 2020, GCK'S file signatures table, URL: https://www.garykessler.net/library/file_sigs.html.

[6] Merriam-Webster, 2020, Metadata, URL: https://www.merriam-webster.com/dictionary/metadata.

[7] J. Riley, Understanding metadata, National Information Standards Organization (NISO), 2017.

[8] JEITA CP-3451, Exchangeable image file format for digital still cameras: Exif Version 2.2, Standard, Japan Electronics and Information Technology Industries Association, 2002.

[9] E. U. A. F. C. (ENISA), Introduction to Network Forensics, European Union Agency For Cybersecurity (ENISA), 2019. doi:10.2824/995110.

[10] L. Sikos, Packet analysis for network forensics: A comprehensive survey, Digital Investigation 32C (2020) 1–12. doi:10.1016/j.fsidi.2019.200892.