

CSE 4004

Digital Forensics

**Lab
Session 9**

TOPIC: File signature analysis

Name: Makesh Srinivasan

Registration number: 19BCE1717

Slot: L49 + L50

Date: 7-October-2021-Monday

Faculty: Prof. Nagaraj

AIM: Exploring File signature analysis in Microsoft Windows 10 environment

INSTRUCTIONS:

File signatures are data used to identify or verify the content of a file. Such signatures are also known as magic numbers. Almost all file types contain a file signature at the beginning of a file and some contain particular data patterns at the end of the file. These patterns at the beginning of a file and the end of a file may be called as headers and footers respectively.

File signature analysis is done primarily to check files are what they claim to be. Changing the extension of a file does not change its contents. For example, suppose we have a genuine jpg file called file.jpg. Renaming it as file.txt will not change its contents. You may check this using a hex editor. So we can easily detect a jpg file impersonating as a txt file by doing file signature analysis.

A signature analysis will compare a file's header or signature to its file extension. A file header identifies the type of file and is located at the beginning of the file's data area. The Windows operating system uses a file's extension to associate the file with the proper application. UNIX and Linux operating systems also use a file's header information to associate file types to specific applications.

Download at least two files with each of the following extensions from the Internet and keep them in a folder: jpg, png, bmp, gif, pdf

Use a hexadecimal editor such as Winhex (see <https://www.x-ways.net/winhex/>) or some other hexadecimal editor (see https://en.wikipedia.org/wiki/Comparison_of_hex_editors) to look at the hexadecimal contents of the file in order to find headers and footers. Check whether headers and footers are the same for the same file type.

See the following sites for more information about how file signatures look like.

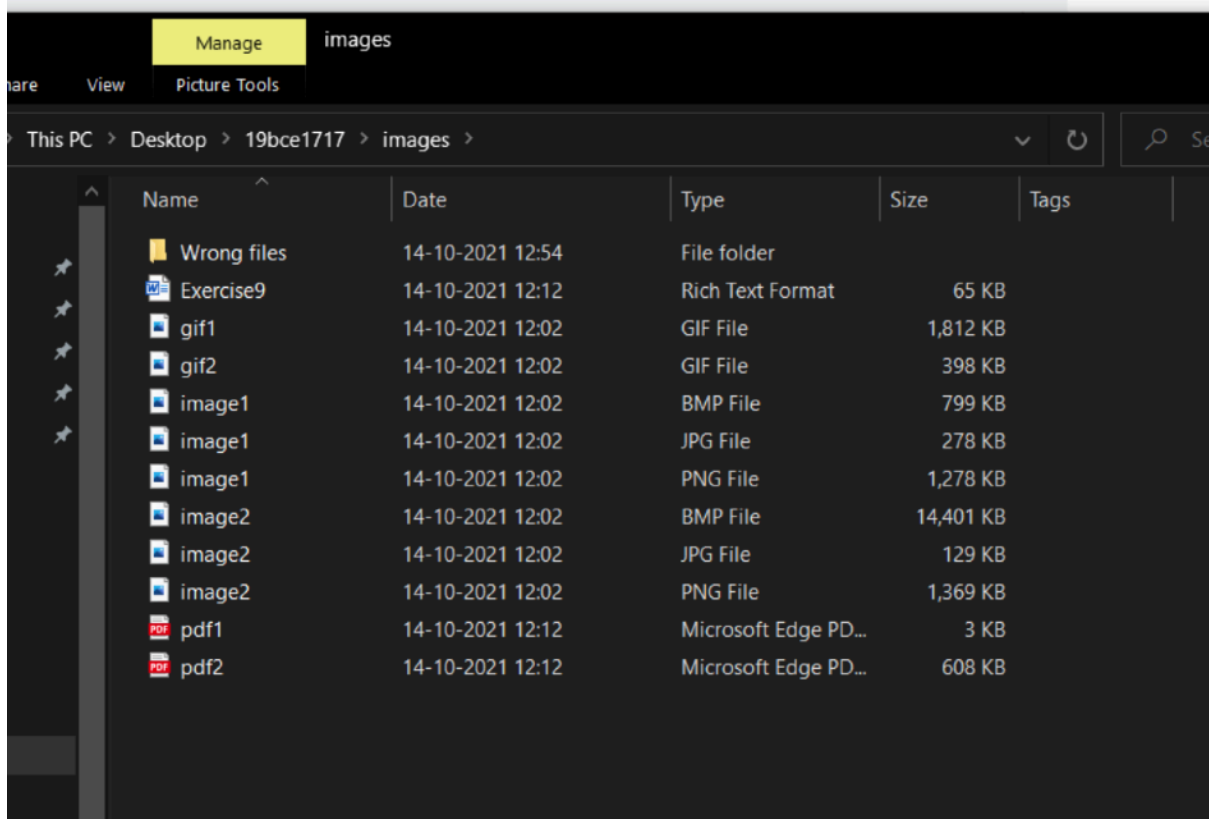
https://en.wikipedia.org/wiki/List_of_file_signatures

https://www.garykessler.net/library/file_sigs.html

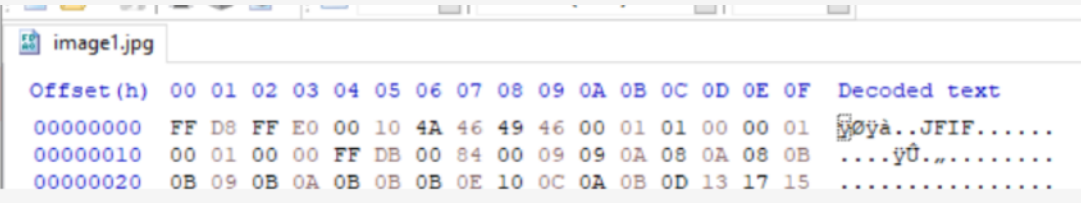
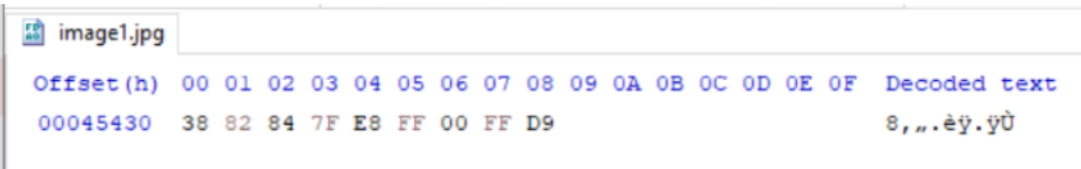
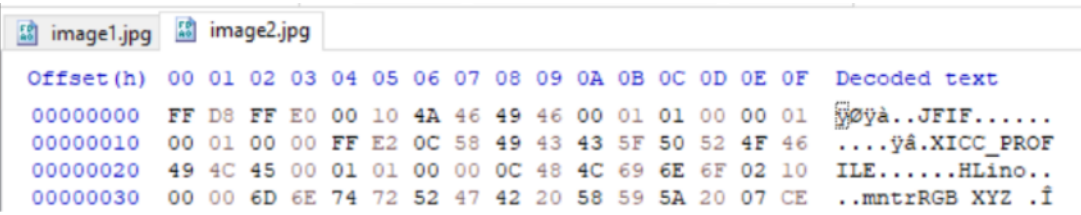
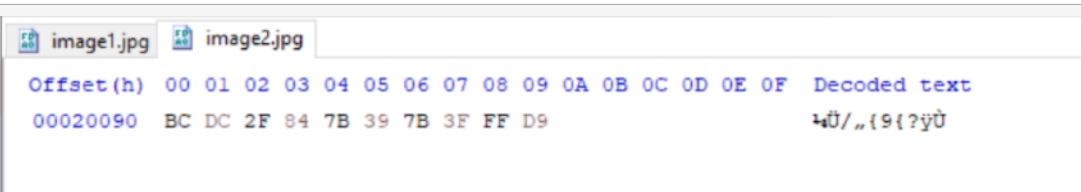
Include screenshots in your submission.

OBSERVATION:

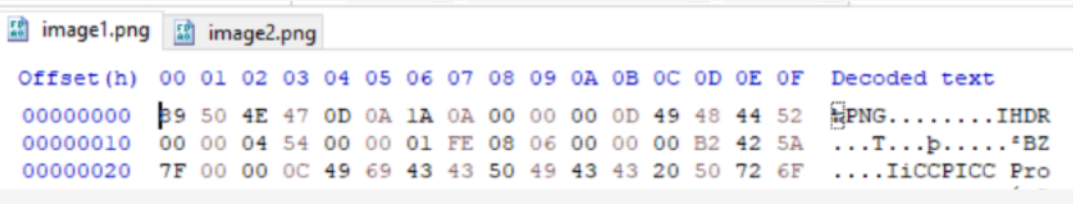
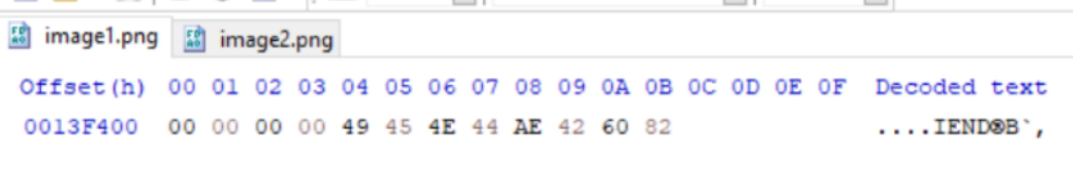
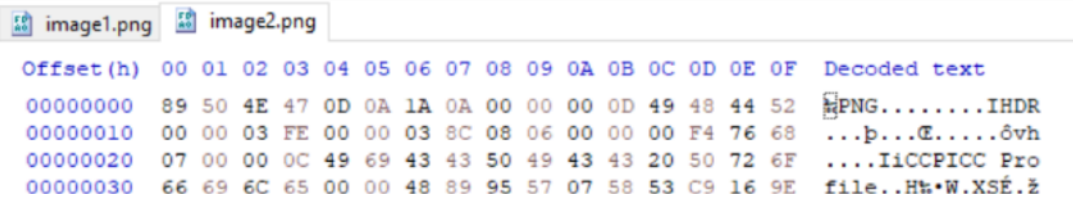
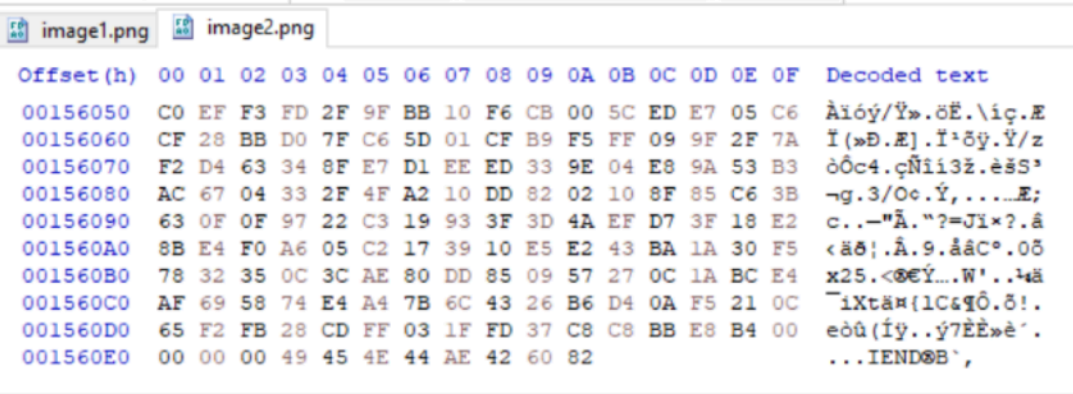
Working directory



File signature analysis

#	JPG
	File 1
Start	
End	
	File 2
Start	
End	

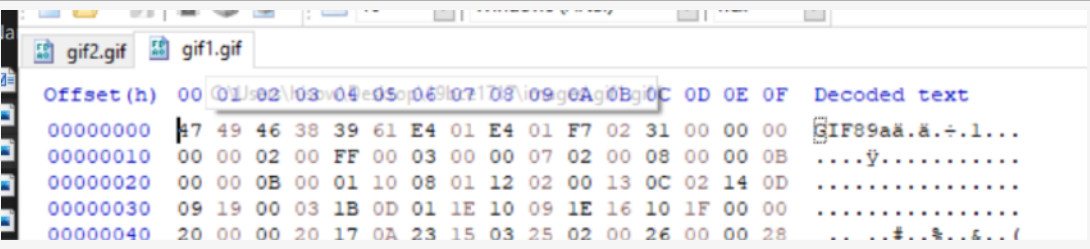
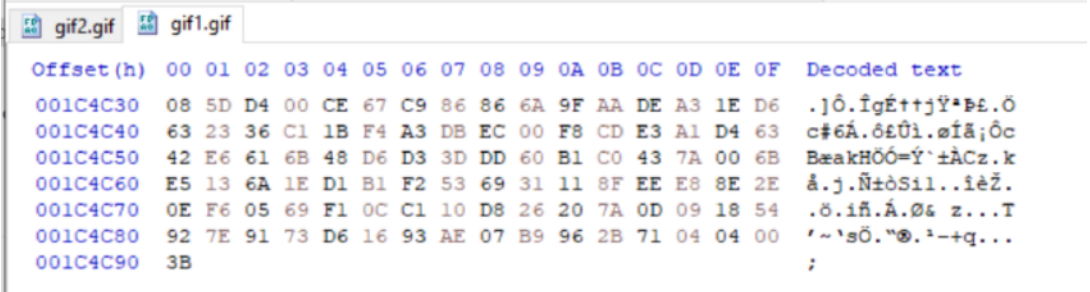
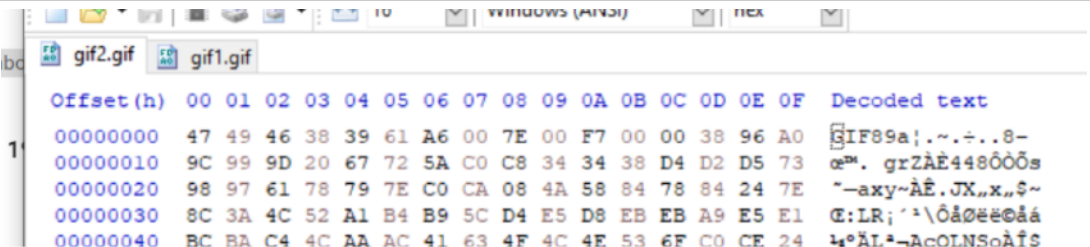
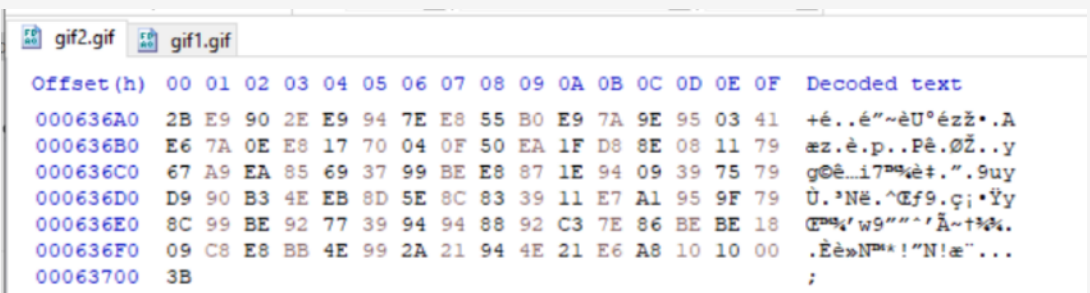
The files image1.jpg and image2.jpg have the same few starting bytes (FF D8 FF E0 10 ...) and this can be viewed in the decoded text section on the right. This is known as the header of the JPG file, this indicates that this is a JPG File. They also have the same bytes at the end, this is known as the footer (FF D9) of the JPG file. The header and footer allow the Operating System to recognise what kind of file this is even if the extension is altered or changed.

#	PNG
	File 1
Start	 <pre> Offset(h) 00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F Decoded text 00000000 89 50 4E 47 0D 0A 1A 0A 00 00 0D 49 48 44 52 PNG.....IHDR 00000010 00 00 04 54 00 00 01 FE 08 06 00 00 00 B2 42 5A ...T...p.....'BZ 00000020 7F 00 00 0C 49 69 43 43 50 49 43 43 20 50 72 6FIiCCPICC Pro </pre>
End	 <pre> Offset(h) 00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F Decoded text 0013F400 00 00 00 00 49 45 4E 44 AE 42 60 82IENDB`, </pre>
	File 2
Start	 <pre> Offset(h) 00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F Decoded text 00000000 89 50 4E 47 0D 0A 1A 0A 00 00 0D 49 48 44 52 PNG.....IHDR 00000010 00 00 03 FE 00 00 03 8C 08 06 00 00 00 F4 76 68 ...p...E.....δvh 00000020 07 00 00 0C 49 69 43 43 50 49 43 43 20 50 72 6FIiCCPICC Pro 00000030 66 69 6C 65 00 00 48 89 95 57 07 58 53 C9 16 9E file..Hh•W.XSÉ.ž </pre>
End	 <pre> Offset(h) 00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F Decoded text 00156050 C0 EF F3 FD 2F 9F BB 10 F6 CB 00 5C ED E7 05 C6 Àlóý/Ý».ðË.\iç.Æ 00156060 CF 28 BB D0 7F C6 5D 01 CF B9 F5 FF 09 9F 2F 7A Ĩ(»Ð.Æ).İ'ðÿ.Ý/z 00156070 F2 D4 63 34 8F E7 D1 EE ED 33 9E 04 E8 9A 53 B3 ôÔc4.çŇii3ž.èŠS³ 00156080 AC 67 04 33 2F 4F A2 10 DD 82 02 10 8F 85 C6 3B ~g.3/Oc.Ý,...Æ; 00156090 63 0F 0F 97 22 C3 19 93 3F 3D 4A EF D7 3F 18 E2 c..-"Ä."?=Ji×?.â 001560A0 8B E4 F0 A6 05 C2 17 39 10 E5 E2 43 BA 1A 30 F5 <âð .Ä.9.ââC°.0ð 001560B0 78 32 35 0C 3C AE 80 DD 85 09 57 27 0C 1A BC E4 x25.<ðËÝ...W'...4â 001560C0 AF 69 58 74 E4 A4 7B 6C 43 26 B6 D4 0A F5 21 0C ~iXtâ={lC&ŋÔ.ð!. 001560D0 65 F2 FB 28 CD FF 03 1F FD 37 C8 C8 BB E8 B4 00 eðû(İÿ..ý7ËË»è´. 001560E0 00 00 00 49 45 4E 44 AE 42 60 82IENDB`, </pre>

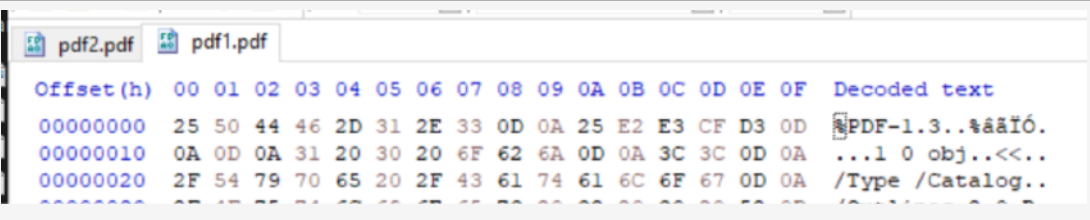
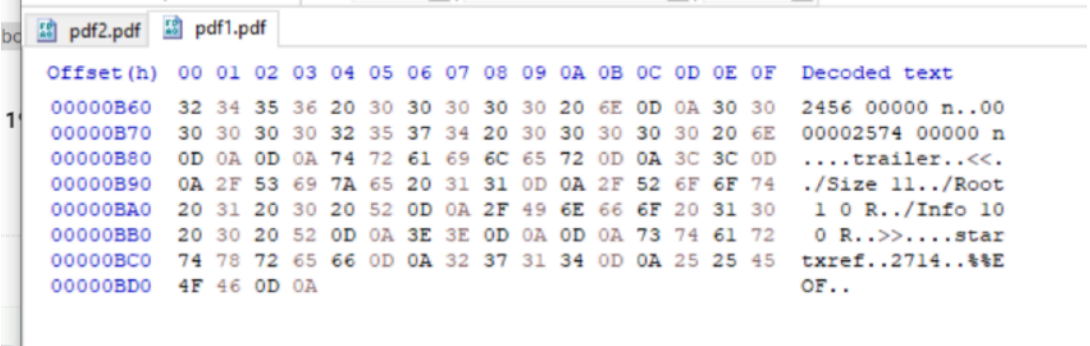
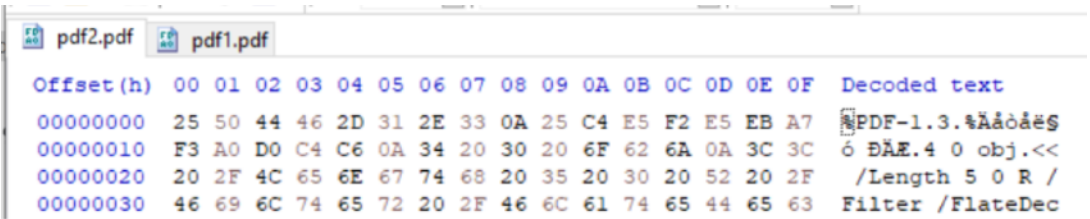
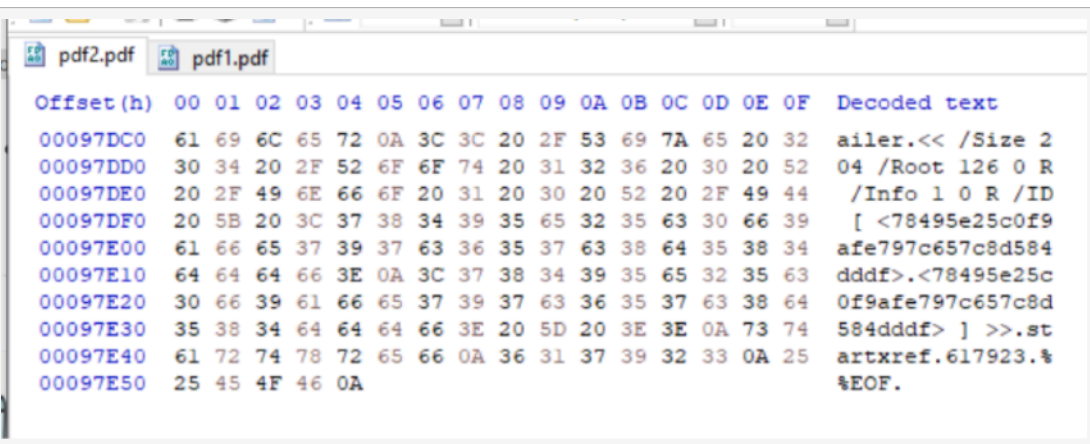
The files image1.png and image2.png have the same few starting bytes (89 50 4E 47 0D ...) and this can be viewed in the decoded text section on the right. This is known as the header of the PNG file, this indicates that this is a PNG File. They also have the same bytes at the end, this is known as the footer (42 60 82) of the PNG file. The header and footer allow the Operating System to recognise what kind of file this is even if the extension is altered or changed.

#	BMP
	File 1
Start	
End	
	File 2
Start	
End	

The files image1.bmp and image2.bmp have the same few starting bytes (42 4D 8A) and this can be viewed in the decoded text section on the right. This is known as the header of the BMP file, this indicates that this is a BMP File. They DO NOT have the same bytes at the end (footer of the file). The header allows the Operating System to recognise what kind of file this is even if the extension is altered or changed.

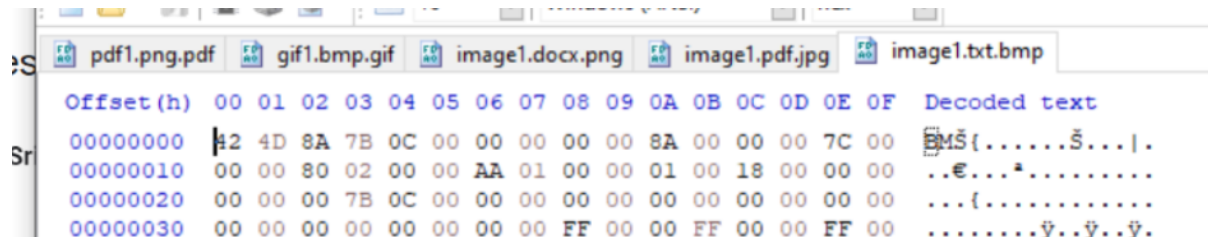
#	GIF
	File 1
Start	
End	
	File 2
Start	
End	

The files gif1.gif and gif2.gif have the same few starting bytes and this can be viewed in the decoded text section on the right. This is known as the header of the GIF file, this indicates that this is a GIF File. They also have the same bytes at the end, this is known as the footer of the GIF file. The header and footer allow the Operating System to recognise what kind of file this is even if the extension is altered or changed.

#	PDF
	File 1
Start	
End	
	File 2
Start	
End	

The files pdf1.pdf and pdf2.pdf have the same few starting bytes and this can be viewed in the decoded text section on the right. This is known as the header of the pdf file, this indicates that this is a PDF File. They also have the same bytes at the end, this is known as the footer of the pdf file. It is sometimes common to find “..” after EOF but the hexadecimal values 45 4F 46 are representing the EOF which is unique to PDF files

Example of when the file extortion is forcefully modified:



The file is a bmp type, but the extension txt when added does not change the signature of the bmp file. The first few bytes are those of the BMP file types as seen in the previous sections.

Conclusion

The file signatures (header and sometimes, footer) can help identify the type of file even if the extension is modified. The changes to the extension do not modify the contents of the file and this was explored using the HxD editor. The header and footer allow the Operating System to recognise what kind of file this is even if the extension is altered or changed.
