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

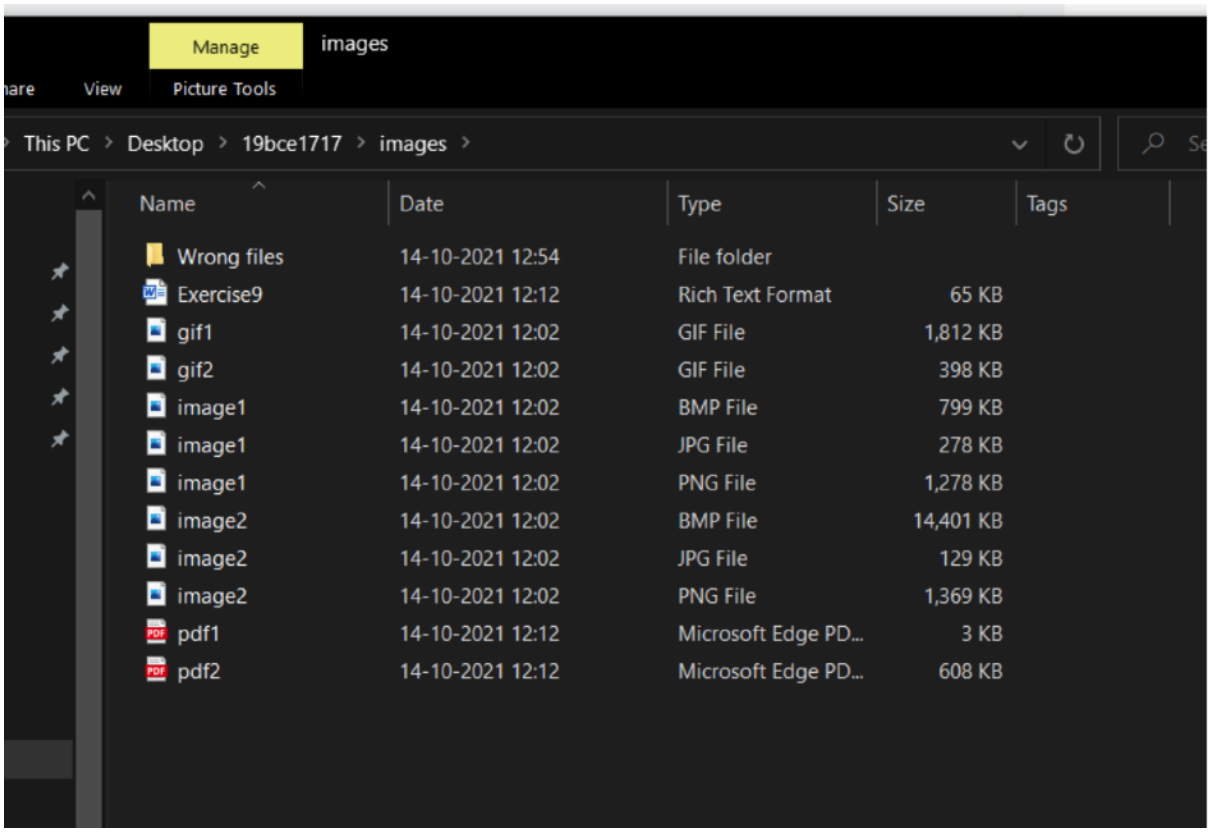
Use a hexadecimal editor such as Winhex (see https://www.x-ways.net/winhex/) or some other hexadecimal editor (see https://en.wikipedia.org/wiki/Comparison_of_hex_editors) to look at the hexadecimal contents of the file in order to find headers and footers. Check whether headers and footers are the same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

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

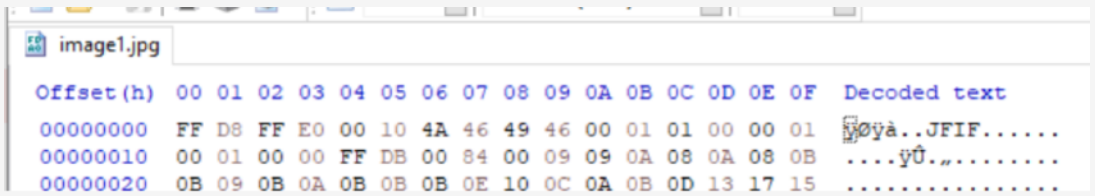
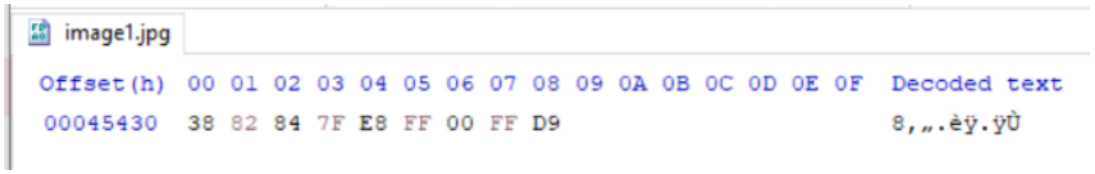
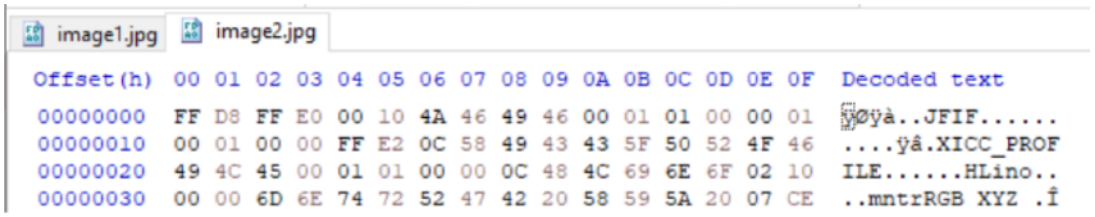
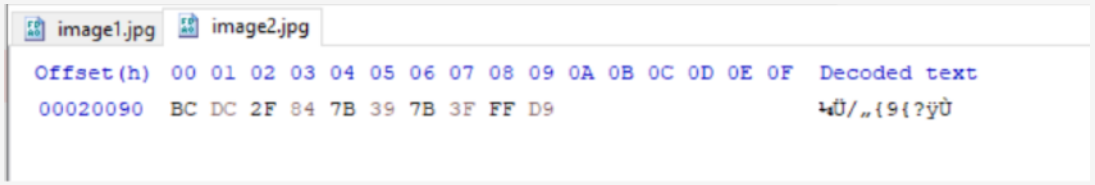
https://en.wikipedia.org/wiki/List_of_file_signatures
https://www.garykessler.net/library/file_sigs.html
Include screenshots in your submission.

OBSERVATION:

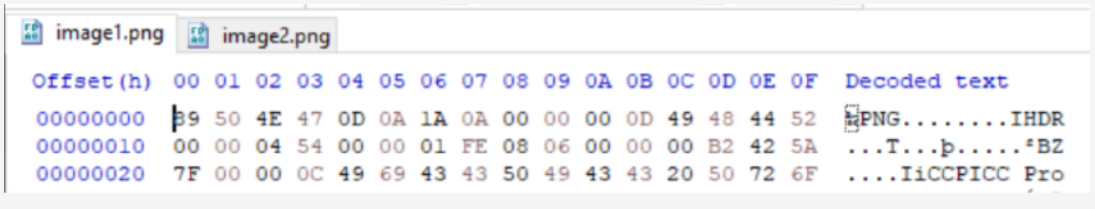
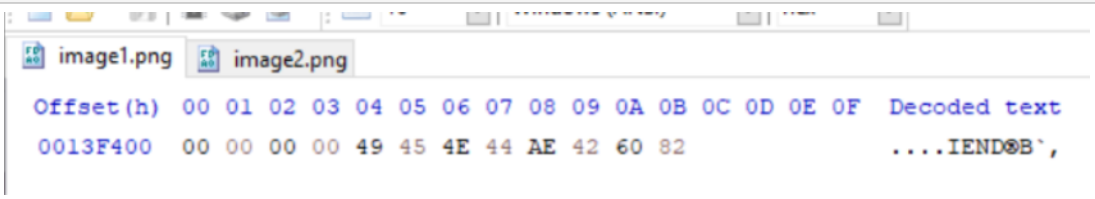
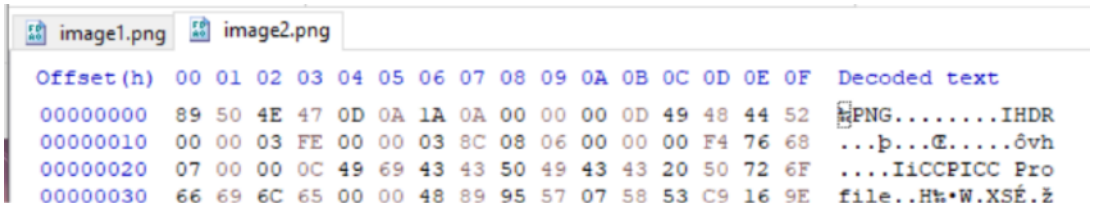
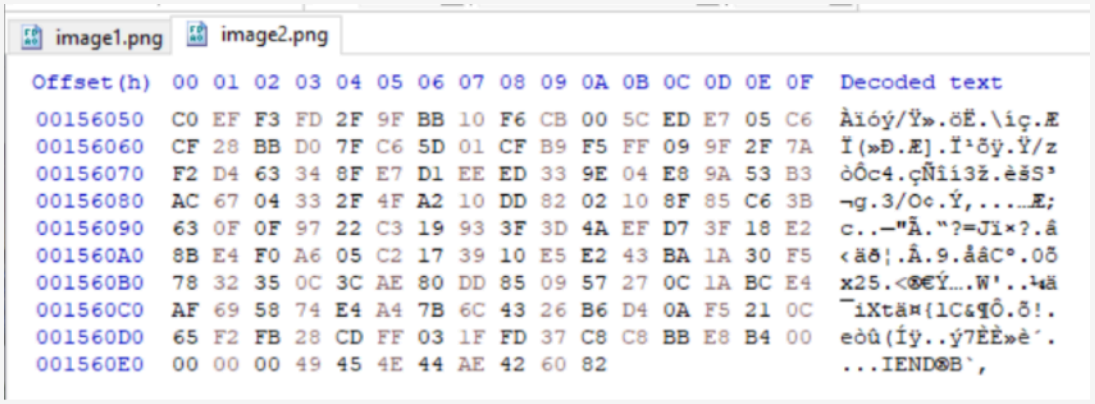
Working directory



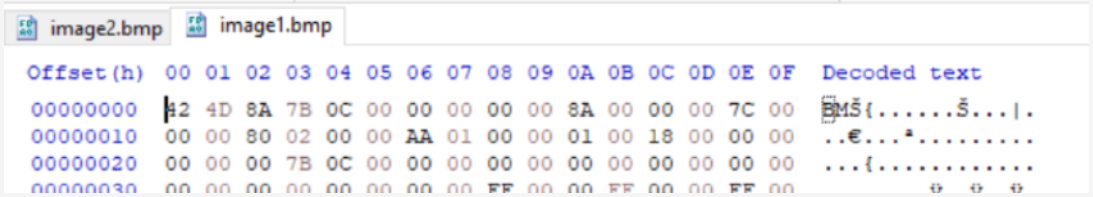
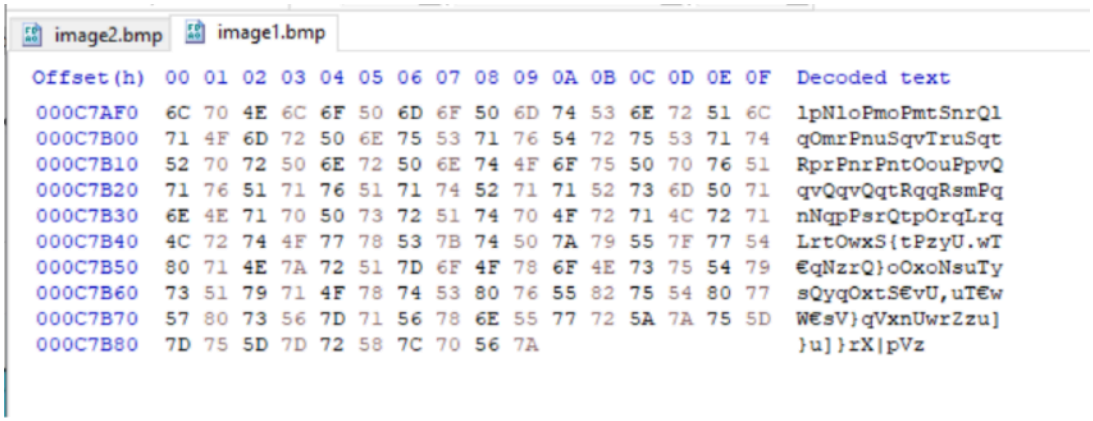
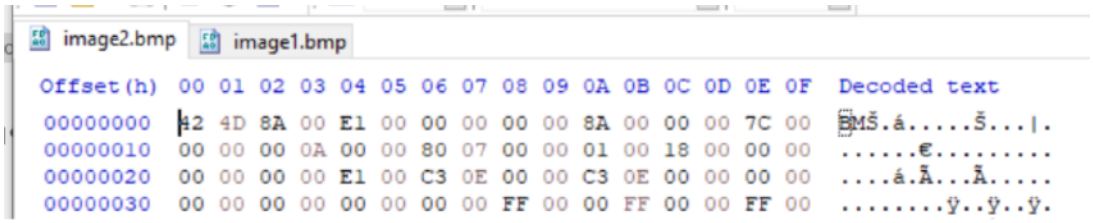
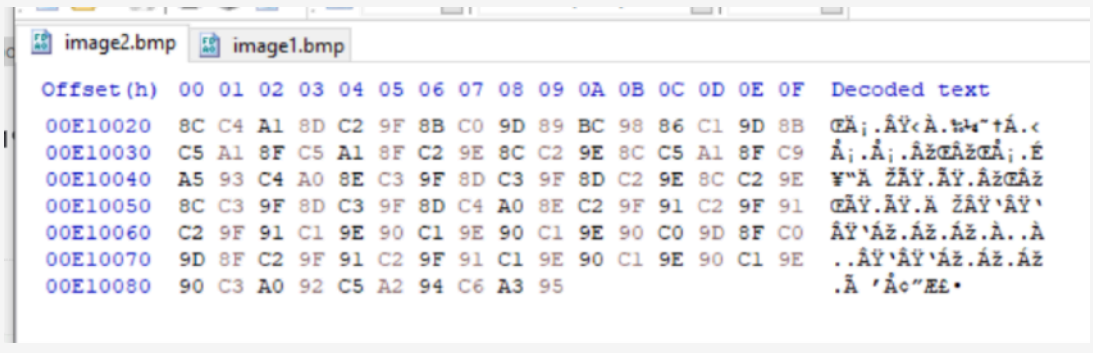
File signature analysis

#	JPG
	File 1
Start	
End	
	File 2
Start	
End	

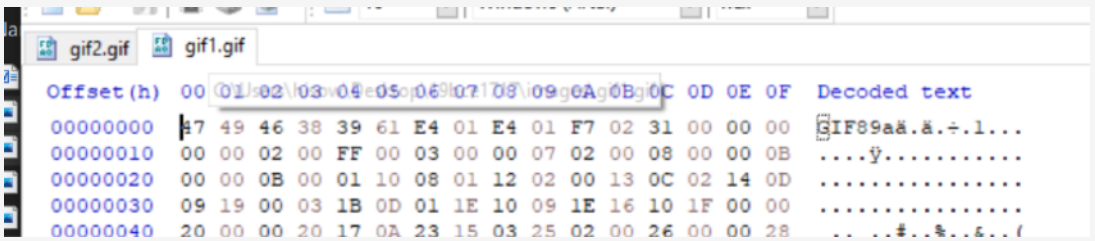
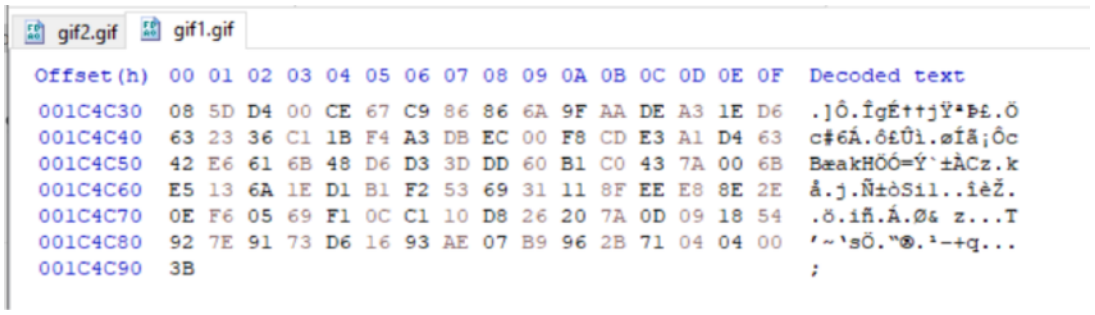
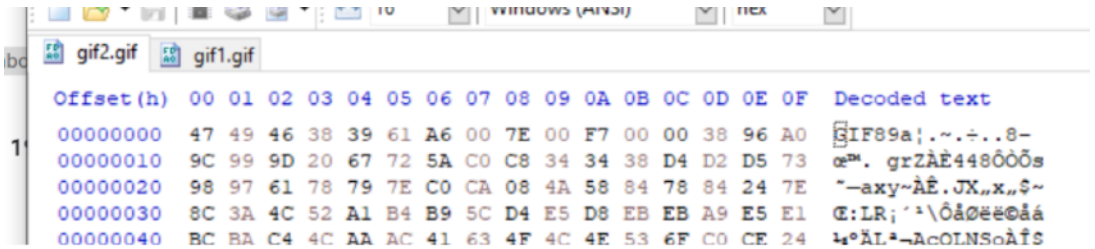
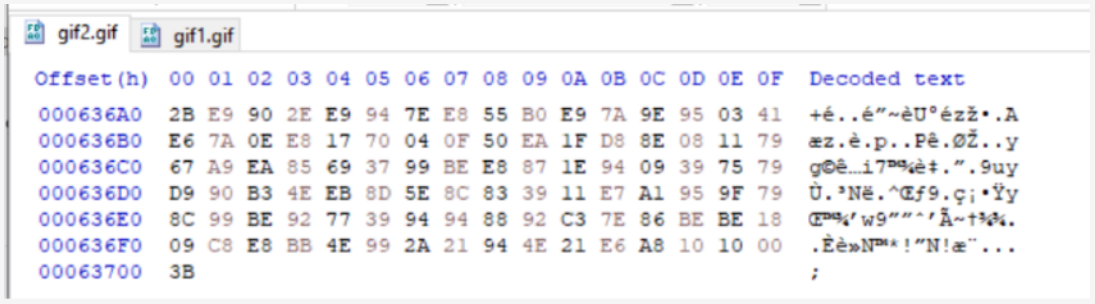
The files image1.jpg and image2.jpg have the same few starting bytes (FF D8 FF E0 10 …) and this can be viewed in the decoded text section on the right. This is known as the header of the JPG file, this indicates that this is a JPG File. They also have the same bytes at the end, this is known as the 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	
End	
	File 2
Start	
End	

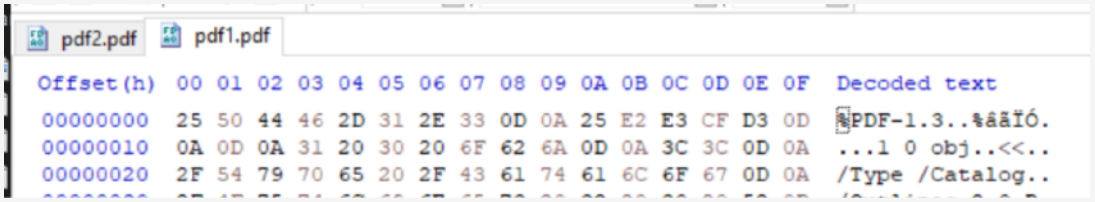
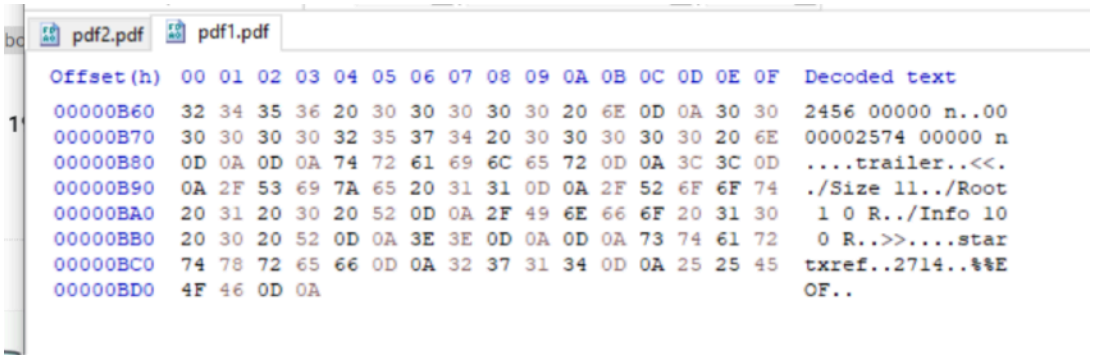
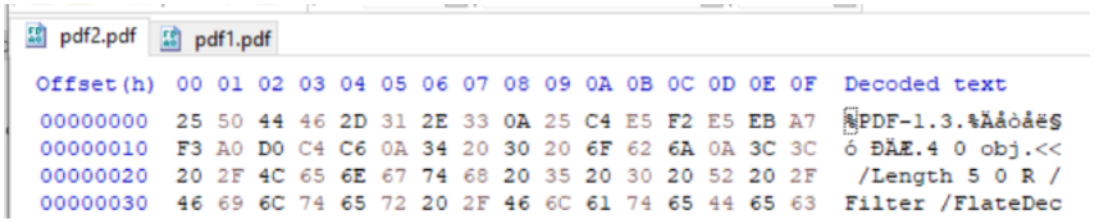
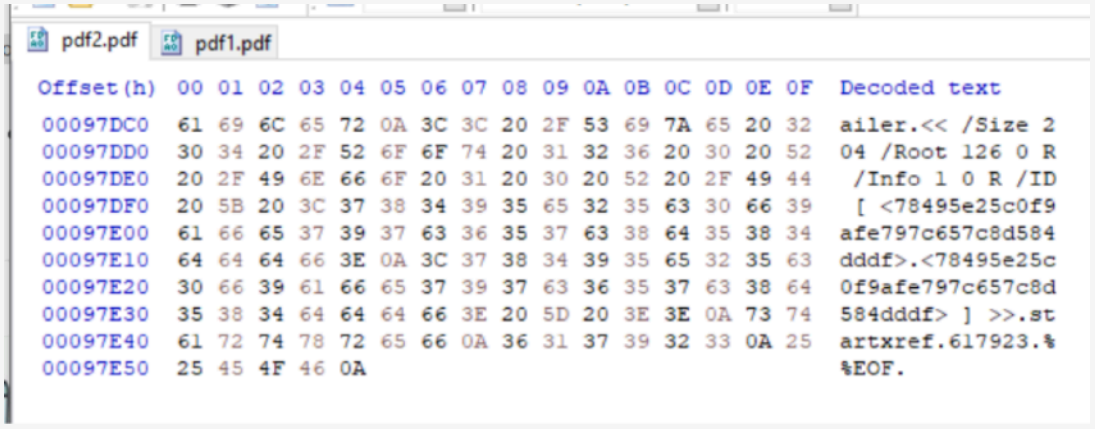
The files image1.png and image2.png have the same few starting bytes (89 50 4E 47 0D …) and this can be viewed in the decoded text section on the right. This is known as the header of the PNG file, this indicates that this is a PNG File. They also have the same bytes at the end, this is known as the 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.

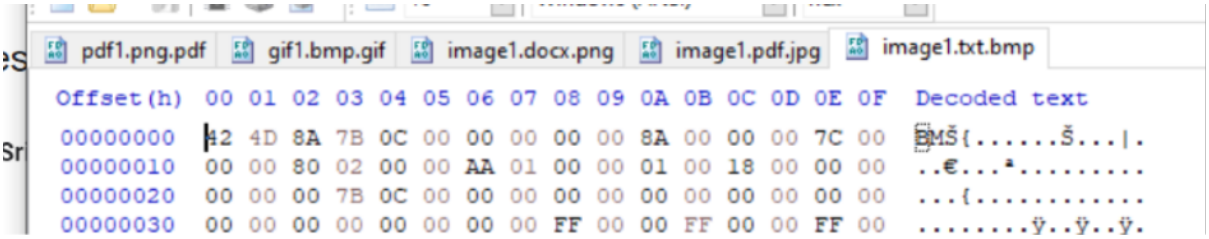
#	GIF
	File 1
Start	
End	
	File 2
Start	
End	

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.
