Digital Forensics

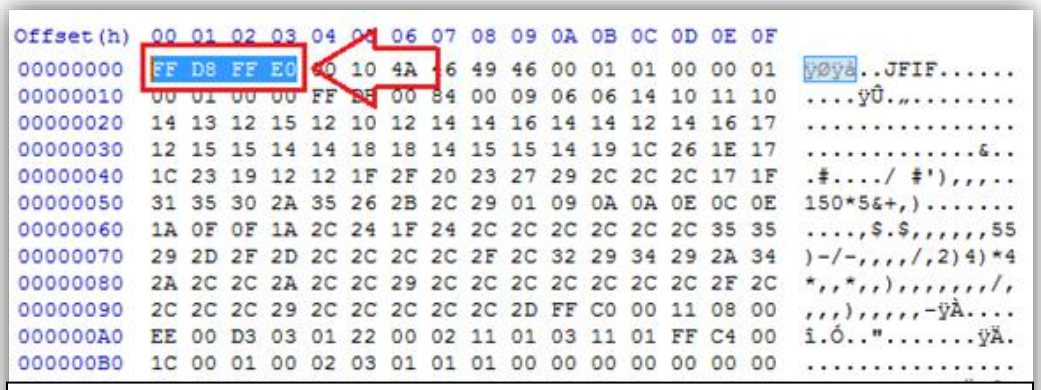
The Basics and Importance of Digital Forensics in a Tactical Environment

Sgt Ben Sutton
SPECIAL OPERATIONS CAPABILITY SPECIALIST – ALL SOURCE ANALYST | DIRECT SUPPORT TEAM DELTA

When conducting media exploitation (MEDEX), there is essentially two lines of effort that you are supporting: The forensic acquisition of information that can be used to prosecute individuals through the justice system, and information that can be used to enhance the intelligence picture of the battlespace. There is a fine line between the two, and understanding the difference is important. For collecting information for legal prosecution, it is important that everything is acquired forensically and the chain of custody is followed and well documented. If the chain of custody is neglected, the defense can use that as leverage and discredit the prosecution. However, if you find information that is of intelligence value, it is less important to follow the chain of custody, and the priority becomes disseminating the information. To help delineate the difference, think about a raid. If a raid is conducted on a house, and an HVI (High Value Individual) is being questioned, you as an analyst can look through his computer to support tactical questioning with the information you see. This is a very quick turn around, and quickly looking through the files for financial records or anything else incriminating becomes the priority due to not knowing how much time you will have with the target and trying to get enough information to detain them is the priority. This differs from a lab setting which would ideally be in a controlled environment and have much more time to carefully document what files were where and what processes you used. This is not to say that collection of evidence for legal action and intelligence gathering can't happen at the same time. When collecting information in a lab setting, it is highly likely that the information you gather will be used to support both legal prosecution and enhancing the intelligence picture. It is important to understand the difference in application of the information you are gathering so that you can most effectively disseminate it, weather it is to other intelligence entities in the area, or handing over case files to law enforcement organizations.

Digital forensics and exploitation of digital media is becoming increasingly important as data is becoming easier to store in larger volumes, as well as being easier to hide. Having at least a baseline understanding of how digital forensics and MEDEX function is important because it will enable analysts to understand how to extract information from media, as well as understanding the sophistication of the target you are dealing with. As emerging technologies are rapidly becoming more prevalent, it is becoming easier and easier to hide data for exfiltration, but through careful forensic examination, the steps to obfuscate data can be illuminated, examined and brought forward to use against individuals for public or private prosecution. When criminals and insider threats are attempting to extract data, there are a few parts of the overall process that leads up to getting a court conviction. First, they will have to hide the data (Assuming we are dealing with someone who is knowledgeable on computers and isn't sending propriety data in the open) before sending it so that the traffic won't be flagged as suspect. After hiding and sending the data, they will have to destroy the evidence so that no one will know about how they sent the data, or what data was sent. Once an investigation is opened, a forensic examiner will have to examine the perpetrator's drives and attempt to determine if s/he can prove beyond reasonable doubt that a crime has been committed, and that it is attributable to the individual. After analyzing the problem set, it is important to look at the tools available and choose which tool is best suited for the job at hand. Part of picking the best tool is being able to characterize the sophistication of the target. For example, if s/he is using full device encryption (FDE), it will be almost impossible to 'break' into the target drive, but if they are using a less sophisticated means of securing data, you could feel more confident not diving into the more time consuming methods.

An easy way to make files unreadable is by editing the header or footer[1] of a document with a hexadecimal editing program, which is sometimes called 'poor man's encryption' as it requires very little knowledge of computers and doesn't require passwords to be remembered. A header is one of the three parts of a document that is read by a program. The header is the first thing a computer reads and determines what type of document it is. It is possible to edit hex values in a file and still maintain general integrity of the file. Especially in larger files, where there is more data to confuse and distract the analyst and anyone else attempting to look at the content. The hex values are a hexadecimal code that determines the content of the file. It is essentially the script that computer programs read to understand what output should be displayed or rendered in a readable format by people. The values in a hexadecimal format are 16 digits long, start at 0 and go up to 9, and then from A to F; a full string going from start to finish increasing one value each time would look like 0-1-2-3-4-5-6-7-8-9-A-B-C-D-E-F. This code can be manipulated using a hex-editing program, one of which is called HexEdit. For example, a Microsoft Word header will look like D0 CF 11 E0, which can loosely be read as DOC FILE if you use some imagination. By changing the header to something else, like the header of a JPEG (FF D8 FF E0), the file will try to be read as a JPEG (Comparable to changing the extension form .doc to .jpeg). The rest of the file isn't formatted as a JPEG, so it will be unreadable. When using Viking, one of the output options is mismatched header/footer, which will flag files that look like certain files, but have the header/footer of another file.



Example of a hex editor highlighting the header of a .jpeg file

The data that is changed in the hex editor (HexEdit is part of the freeware package) can be outside of the hexadecimal range and can be used to hide short messages or passwords within the hex values without compromising the integrity of the file (As viewed normally by the end-user without software). When changing the hex values, it is important that you only change the data in the blank hex values, where there aren't any values. It is also important to keep the content of the messages short, because changing the script can change the content of the file or render it unreadable if too much is changed. Advanced software can spread messages widely throughout the hex values in a specific pattern so that it isn't able to be seen or recognized. Once the file is encoded with the message, it is sent and then decoded by another user with the same software. This poses a challenge because unless you know exactly what pattern to look for, you won't be able to detect the message. An example of hiding data is putting files and images within a picture. This can help in an examination of evidence by potentially revealing how the individual is hiding files if you are examining his/her workstations and you find a number of files that are

unable to be opened (Which is an indicator of poor-man's encryption) but have mismatched headers and footers from what the extension type is.

The practice of hiding files and information within images is called steganography. It is possible to hide images within images, which can be undetectable to the human eye due to the high pixel density of modern pictures. Russian spy network were using steganography in 2005 to pass information to and from deep cover spies. When viewed (Even forensically) it will appear as a normal picture in all aspects. Depending on the size of the data that is being hid, the file size may only be a few kilobytes off. If someone is hiding comparatively massive amounts of data, a small picture that is normally a few hundred kilobytes to one or two megabytes may be a few hundred megabytes. If you find images that are grossly larger than you would expect them to be, you should start to look if there are any signs of steganography or check to see if there is any hint to what the password might be. A red flap for identifying steganography is if someone is uploading the same picture multiple times with seemingly no context. An easy way to identify the possibility of steganography is to either look at the file size, or use a hashing program (One of which is HashCalc; hashing will be discussed later) to see if the hash values are different. An older program called Jphide uses steganography to hide .jpg files inside of other .jpg files. The issue with examining files that have steganography implemented with Jphide specifically is that Jphide uses alphanumeric passwords and will only tell you if steganography was applied to the image using that specific program. This means that if you are looking through an E01 file during a forensic investigation, you won't be able to see the hidden files without the password, but you might be able to get a better understanding of the level of sophistication of the person whose drives you are examining.

A more sophisticated method of transporting data while maintaining data integrity is using file encryption software. Encrypting data software such as VeraCrypt or AxCrypt requires that both the creator and all subsequent users of the file have the same software, as well as the password. Another form of encrypting media is to apply encryption to the entire drive, which is called full device encryption (FDE) as mentioned previously. Applying FDE to a drive prevents people from accessing your files with a USB er, which can bypass the standard login, and provide access to the other files and information on a system. As a forensic examiner, it is important to recognize that (While it is quite easy to encrypt your files) if someone goes out of their way to install encryption software, they have an at least baseline knowledge of how data is accessed and stored, and might be worth going more in-depth and taking a more meticulous approach.

If you are supporting an investigation, one of the first things you should look at is if there is evidence of data being deleted or destroyed. There are a few ways to get rid of data. The quickest and least effective is to just delete the files from your computer and empty your recycle bin. Most people who don't have a depth of knowledge over computers and how they work will stop there, but what deleting a file (Or reformatting a drive) does is remove that data from the directory so that it can be written over again (Called unallocated space or cluster tips). The files are still on the drive, but the path to access that file is erased and the containers where the data path was is available to be written over again. Even after deleting the files and using the drive for a long time, the cluster where the original data was may not be fully used and can be extracted with a data carving program.

The only way to prevent analysis by a data carver is to conduct a bit-by-bit wipe of the drive, where each bit of data is written over by a program. This is called wiping a drive and is important to know the difference. Just clicking 'delete' on a file doesn't fully delete the file, and even writing over a drive may

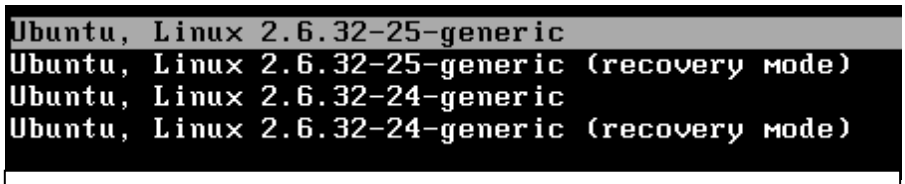
miss some parts of the data. Taking data destruction a step further, in order to fully delete a file, you would have to write over each bit of the file which may take many more passes. A pass is one 'run' over of the wiping software. Whoever wiped Hillary Clinton's E-Mail servers wasn't knowledgeable on what they were doing because they didn't check the box that would have wiped over the cluster tips, which is how her E-Mails were found. Writing over a drive so many times causes damage, as it takes a toll on the life of the drive. It may be possible that you would make the drive become inoperable before you fully delete a file. Understand the difference between wiping files or reformatting a drive is important to a forensic investigator because you may need to look further than the trash can for deleted data, as it may have been partially written over and not viewable without a data-carving tool.

1	(Random)
2	(Random)
3	(Random)
4	(Random)
5	01010101 01010101 01010101
6	10101010 10101010 10101010
7	10010010 01001001 00100100
8	01001001 00100100 10010010
9	00100100 10010010 01001001
10	00000000 00000000 00000000
11	00010001 00010001 00010001
12	00100010 00100010 00100010
13	00110011 00110011 00110011
14	01000100 01000100 01000100
15	01010101 01010101 01010101
16	01100110 01100110 01100110
17	01110111 01110111 01110111
18	10001000 10001000 10001000
19	10011001 10011001 10011001
20	10101010 10101010 10101010
21	10111011 10111011 10111011
22	11001100 11001100 11001100
23	11011101 11011101 11011101
24	11101110 11101110 11101110
25	11111111 11111111 11111111
26	10010010 01001001 00100100
27	01001001 00100100 10010010
28	00100100 10010010 01001001
29	01101101 10110110 11011011
30	10110110 11011011 01101101
31	11011011 01101101 10110110
32	(Random)
33	(Random)
34	(Random)
35	(Random)

Pattern for a 35-pass wiping method called the 'Gutmann method'

The first step in retrieving data from a drive (Especially in a law-enforcement environment) is to make a forensic copy of it with an imaging software such as Forensic Toolkit Imager, or FTK Imager. If you are imaging a thumb drive or other removable media, it is critically important that you do NOT interact with the original media during your forensic examination. You need to use either a hardware or software write-blocker to prevent your computer from touching the media and expose yourself to the risk of changing the hash values. Software write-blockers change registry files and disable write privileges for new removable media that are connected to the computer, so it is important to remember to turn the write-blocker off after you are done. If you have a device already connected, it will not be effected.

Another way of interacting with digital media without changing the data is using a boot-loading[3] software, such as Sumuri's Paladin software. The Paladin software will allow you to boot into a system (Bypass the normal loader, as long as the drive isn't encrypted, or other advanced settings aren't enabled). This will allow you to pull specific file types and other defined parameters from a target drive that you are investigating during a forensic investigation. This software, since it is acting as a separate boot loader and isn't interacting with the host operating system, can provide you with a forensic image of files on the target system since it isn't touching the normal operating system. The hashes will be the same, even on the registry because it is entering the memory of the system from an alternate route. Paladin software is especially good because it can look for and extract specific file types on a drive that is connected to the target system. It won't give you a forensic image, but if you are in the process of ingesting an E01 file and getting the information from that, you can view specific files in the meantime. This is important while investigating media, because if you only have a limited amount of time on target, you can both forensically copy files while simultaneously reviewing what is being extracted. Additionally, bootloaders don't interact with the normal operating system, so there isn't a trace that someone logged in.



Example boot loader menu

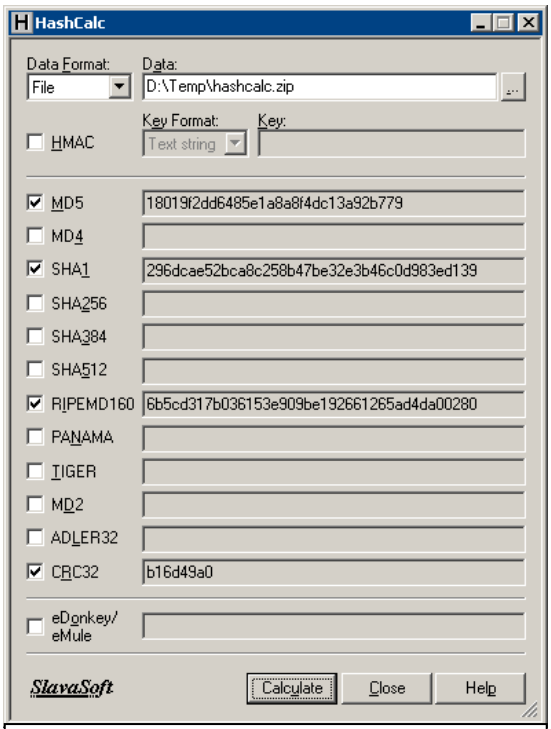
An example of using the Paladin software is hitting a target that is suspected to have footage of a U.S. Person; you could use the Paladin software to triage the material on the target's suspected media and view all videos on the target's computer. Finding illicit material will allow for the individual to be arrested on the spot, and even if you don't find anything illicit, whatever is recovered can be used to conduct a more in-depth search and use what images are pulled up to warrant further action on the target at a later date if more information is found on them. The evidence that is discovered during a search may also be used to get a more refined idea of the sophistication and structure of the target.

Some other tools that are of value during forensic examinations are IrfanView, and VLC Media Player. IrfanView is a powerful image viewing program. Not only does it have an intuitive UI, it also allows for clear 'enhancement'/magnification (When compared to Microsoft's suite of image viewing software), and arguably most importantly, it an integrated '.exif' data viewer. .exif data is metadata, which is information about data. Metadata can be (But is not limited to) the location (Latitude/Longitude) of the where the image was captured, capture date/time group, and possibly author if they have enabled that setting on their device. This is important, especially when the images/videos are of something that is critically important to US interests, such as where insurgents are placing bombs or videos of US persons which would give locational data and start to drive the planning process.

Unfortunately for digital forensics, social media platforms have been stripping metadata from images that are being posted. This is largely due to pedophilic stalkers who would follow and track the locations of children so that they could effectively target them when they were at parks and other public spaces. This is good for the general public, but can impede investigations if, for example, someone posts an incriminating picture that is geo-tagged and would confirm or deny of they were in the area in question if the data was not stripped. A real-life example is in 2017 when an ISIS fighter posted a picture of himself in an ISIS operations room with a geotag, and airstrikes hit the base 22 hours later. This exemplifies how not just law enforcement organizations, but also military, are using metadata to find, correlate and fix individuals who are committing crimes and target them for action.

When examining a targeted/suspect drive for suspect information, you don't want to just focus on the data and metadata, but also the calculated hash[2] algorithms, which was mentioned earlier. There are many types of hash calculators, one of which is HashCalc. When talking about hash lists, there are two different types of lists that you want to have: targeted hash lists and exclusionary hash lists. Exclusionary hash lists are lists of known programs/software/thumbnails that are common to known systems and don't need to be examined (excluded from the results). An example of this is all the default Microsoft emblems and thumbnails that populate registry files. When examining large amounts of data, the default system information can get in the way, and people who are tech-savvy can hide data in a format that may look like default system configuration data to an examiner, but will have a different hash value than the other known system data and will therefore be flagged as non-common data in an area where you know what

'should' be there in a normal system. This will significantly decrease your work load because you won't be digging through all of the system information and you will only get the unique files. The other type of hash list you should have is a targeted hash list. This can be used in military targeting if you know that there is a certain document that has a plan of attack on friendly locations that you have calculated the hash value for, and you find that specific document on targets drive. It can also be used in law enforcement targeting as some dark web sites that solicit the sale of child pornography will list the MD5/SHA hashes to ensure that their clients are receiving the correct file. If you have common hashes of files that are spread via the dark web and you get a hit on a targeted system, you can act on the target faster and more efficiently than without that hash lists. Combining both an exclusionary hash list and a targeted hash list will take a little longer for the computer to compute, but will only highlight the files you want, and allow you to just look at the unique files of the target drive.



User interface of HashCalc with hash values

Looking for hash lists and other specific characteristics can help you build a profile of what the suspect has been doing. Another aspect that you should analyze is the sophistication of the target. For example, if the drive you are examining isn't password protected and has all his/her files on the desktop, you probably won't waste your time crawling through the hex values trying to find hidden messages. However, if you find an encrypted file and a VeraCrypt executable on the desktop, you might want to do a more thorough search, as this specific target is more likely to have a higher level of technical knowledge and be able to take more steps to hide and secure data.

One of the first things you should look for in a target is the level of encryption. If there is none, that just makes your job easier. If there is poor man's encryption, it could indicate at least some level of technical knowledge and would warrant a second look at some of the files. If there were inaccessible files with encryption software such as TrueCrypt/VeraCrypt, I would spend more time looking around the target site (Where the media was recovered) to find the password because there's no use trying to break into a device using advanced encryption software. It would also give you an idea of the importance of the suspect. If the suspect doesn't have knowledge of his encryption/how his devices operates, the question has to be asked of who did it for them, which can illuminate organizational infrastructure.

An important step in both gathering intelligence information and information that is to be used in a prosecution, understanding the chain of custody is important. The chain of custody is the bread crumb trail that depicts who has controlled the item, where it came from, and in some cases what was some to the data (If any initial triage was conducted). Properly documenting the chain of custody is important for gathering intelligence information, especially for highly sensitive targets and for captured enemy material because if the information is lost in translation, it can be difficult to tie the information gained back to the context that it was acquired from. For example, if a hard drive is dropped off at a lab without any documentation or context, and there are attack plans on the drive, it would be difficult to tie the

intelligence to a particular cell or area where an attack is going to occur. Chain of custody is equally, if not more, important in a law enforcement investigation because a court of law is going to require that documentation is filled out correctly, and the investigation could get thrown out if the chain of custody isn't properly completed.

In closing, there are many ways to hide illegal activities in varying levels of sophistication. From poor man's encryption or FDE, criminals will use software to their advantage to attempt to cover their tracks. It is up to law enforcements organizations and their supporting analysts to illuminate the means and methods, the data, and at the end of the day, link the illicit activity to certain individuals in order to get criminals off the streets and make sure the guilty stay guilty and ensure that the innocent aren't wrongly accused. Ensuring that the proper steps are taken to comb through data, and taking the correct measures for the target set (not spending hours looking through hex values when it's not a known TTP) will ensure you are effectively spending your time analyzing data. Also, knowing what you're looking for and prioritizing your efforts for intelligence gathering of time-sensitive requirements will garner the best results. Understanding your requirements and the best method for using specific tools is key to being a proficient forensic analyst.

Glossary

Header/Footer: A header and footer are what programs read to open files in the correct format, and provide computers with all the set-up information for the specific file type. It is important to understand what certain file types look like so that you can manually repair documents if you suspect poor man's encryption. Headers are the first part of the hex values and the footer is the last part, much like on a word document.

Hash: A hash is a unique calculation of all the data about a particular file, such as the author, exact time saved, words used, font, color of text, and anything else that can be changed. It is important to understand hash calculations because it is impossible to recreate hashes, which means if someone has a matching has, they have the exact same file as the one you are referencing. This can be used to see who is sharing files with who and break down communication flow and information sharing. An important exception is with .txt files. .txt files are so small and basic that they don't retain any information like a word file does. No author, saves, fonts etc., which means that a hash value will be the same as long as the content is the same.

E01: An E01 file is the type of file that is ingested into forensic analysis software such as Viking, Encase or autopsy. If the information is too large to fit in the E01 file, succeeding files will be created E02, E03, E04 etc., but only the E01 will be able to be ingested, and all of the other files must be in the same directory as the E01 file when ingesting it into forensic software.

Boot-loader: Requires manual intervention from the user when turning the computer on. For windows computers, during the initialization phase, the user is prompted to press F7 to open the boot menu. Selecting the boot-loader will bypass the BIOS (Basic in/out system) and allow the user access to the computer without going through the normal loading process (Which would require username/passwords). USB boot-loaders can be turned off in system setting if the administrator is knowledgeable enough, and the data on the drive won't be able to be viewed if FDE is being used.