CSN11121/CSN11122
System Administration and Forensics

File metadata and analysis

03/11/2011

r.ludwiniak@napier.ac.uk

Lecture Objectives

1. File metadata and analysis
2. Data hiding
3. Data recovery
   - File signature
   - File carving

METADATA

MFT List of possible attributes

- Defined in $AttrDef entry of MFT, but default is:
  - 0x10 STANDARD_INFORMATION
  - 0x20$ATTRIBUTE_LIST
  - 0x30$FILE_NAME0
  - X40 (NT) $VOLUME_VERSION (2K) $OBJECT_ID
  - 0x50 $SECURITY_DESCRIPTOR
  - 0x60$VOLUME_NAME
  - 0x70 $VOLUME_INFORMATION
  - 0x80$DATA
  - 0x90$INDEX_ROOT
  - 0xA0$INDEX_ALLOCATION
  - 0xB0$BITMAP
  - 0xC0 (NT) $SYMBOLIC_LINK, (2K) $REPARSE_POINT
  - 0xD0$EA_INFORMATION
  - 0xE0$EA0xF0NT$PROPERTY_SET
  - 0x100 (2K) $LOGGED_UTILITY_STREAM

Date-Time Stamps Significance

- File Created
  - This date-time stamp usually shows when a file or folder was created
  - When an existing file is copied, the File Created date-time stamp of the new copy is set to the current time
  - When a file is moved onto a different volume using the Windows command line or drag-and-drop feature, the File Created date-time stamp of the new copy is set to the current time
  - When a file is moved onto a different volume using the Cut and Paste menu options, the File Created date-time stamp remains unchanged (the Last Accessed and Entry Modified date-time stamps would most likely change).
- Modified
  - This date-time stamp represents the last time the $DATA attribute of a file was altered.

Date-Time Stamps Significance

- Last Accessed
  - This date-time stamp represents the most recent time a file or folder was accessed by the file system. This date-stamp does not necessarily indicate that a file was opened; simply placing the mouse over the filename in Windows Explorer can update the last accessed date.

- SIA Modified
  - This date-time stamp represents the last time any attribute in the MFT record for the file or folder was modified. Reasons for an update to this date-time stamp can include changing a file's location on the disk, another data stream being added to the file, or a change in the file's name.

Metadata Analysis Considerations

- Directory Entry time Values
  - Times are stored with respect to time zones
  - Last access and created times are optional
    - Corroborate with Application-Level data

Metadata Analysis Considerations

- Create time
  - Resolution of 10 Milliseconds
- Write Time
  - Resolution of 2 Seconds
- Access Time
  - 1 Day

METHODS OF HIDING DATA

Methods Of Hiding Data

- To human eyes, data usually contains known forms, like images, e-mail, sounds, and text. Most Internet data naturally includes gratuitous headers, too. These are media exploited using new controversial logical encodings: steganography and marking.
  - **Steganography**: The art of storing information in such a way that the existence of the information is hidden.

Methods Of Hiding Data

- **T**o **h**uman **e**yes, **d**ata **u**sually **c**ontains **k**nown **f**orms, **l**ike **i**mages, **e**-mail, **s**ounds, **a**nd **t**ext. **M**ost **I**nternet **d**ata **n**aturally **i**ncludes **g**ratuitous **h**eaders, **t**oo**.** **T**hese **a**re **m**edia **e**xploited **u**sing **n**ew **c**ontroversial **l**ogical **e**ncodings: **s**teganography **a**nd **m**arking.

- ***The duck flies at midnight. Tame uncle Sam***
  - Simple but effective when done well

Methods Of Hiding Data

- **Watermarking**: Hiding data within data
  - Information can be hidden in almost any file format.
  - File formats with more room for compression are best
    - Image files (JPEG, GIF)
    - Sound files (MP3, WAV)
    - Video files (MPG, AVI)
  - The hidden information *may* be encrypted, but not necessarily
  - Numerous software applications will do this for you: Many are freely available online

Methods Of Hiding Data

- Hard Drive/File System manipulation
  - Slack Space is the space between the logical end and the physical end of file and is called the file slack. The logical end of a file comes before the physical end of the cluster in which it is stored. The remaining bytes in the cluster are remnants of previous files or directories stored in that cluster.
    - Slack space can be accessed and written to directly using a hex editor.
    - This does not add any “used space” information to the drive
  - Partition waste space is the rest of the unused track which the boot sector is stored on – usually 10s, possibly 100s of sectors skipped
    - After the boot sector, the rest of the track is left empty

Methods Of Hiding Data

- Hard Drive/File System manipulation cont…
  - Hidden drive space is non-partitioned space in-between partitions
    - The File Allocation Table (FAT) is modified to remove any reference to the non-partitioned space
    - The address of the sectors must be known in order to read/write information to them
  - Bad sectors occur when the OS attempts to read info from a sector unsuccessfully. After a (specified) # of unsuccessful tries, it copies (if possible) the information to another sector and marks (flags) the sector as bad so it is not read from/written to again
    - users can control the flagging of bad sectors
    - Flagged sectors can be read to /written from with direct reads and writes using a hex editor

Methods Of Hiding Data

- Hard Drive/File System manipulation cont…
  - Extra Tracks: most hard disks have more than the rated # of tracks to make up for flaws in manufacturing (to keep from being thrown away because failure to meet minimum #).
    - Usually not required or used, but with direct (hex editor) reads and writes, they can be used to hide/read data
  - Change file names and extensions – i.e. rename a .doc file to a .dll file

METHODS OF DETECTING/RECOVERING DATA

Methods Of Detecting/Recovering Data

- Steganalysis - the art of detecting and decoding hidden data
  - Hiding information within electronic media requires alterations of the media properties that may introduce some form of degradation or unusual characteristics
  - The pattern of degradation or the unusual characteristic of a specific type of steganography method is called a signature
  - Steganalysis software can be trained to look for a signature

Methods Of Detecting/Recovering Data

- Steganalysis Methods - Detection
  - Human Observation
    - Opening a text document in a common word processor may show appended spaces and “invisible” characters
    - Images and sound/video clips can be viewed or listened to and distortions may be found
      - Generally, this only occurs if the amount of data hidden inside the media is too large to be successfully hidden within the media (15% rule)
  - Software analysis
    - Even small amounts of processing can filter out echoes and shadow noise within an audio file to search for hidden information
    - If the original media file is available, hash values can easily detect modifications

Methods Of Detecting/Recovering Data

- Steganalysis Methods – Detection cont...
  - Disk analysis utilities can search the hard drive for hidden tracks/sectors/data
  - RAM slack is the space from the end of the file to the end of the containing sector. Before a sector is written to disk, it is stored in a buffer somewhere in RAM. If the buffer is only partially filled with information before being committed to disk, remnants from the end of the buffer will be written to disk. In this way, information that was never "saved" can be found in RAM slack on disk.
  - Firewall/Routing filters can be applied to search for hidden or invalid data in IP datagram headers

Methods Of Detecting/Recovering Data

- Steganalysis Methods – Detection cont...
  - Statistical Analysis
    - Most steganographic algorithms that work on images assume that the Least Significant Bit (LSB) is random
    - If a filter is applied to an image, the LSB bits will produce a recognizable image, so the assumption is wrong
    - After inserting hidden information into an image, the LSB is no longer non-random (especially with encrypted data). If you apply the same filter, it will no longer produce a recognizable image
    - Statistical analysis of the LSB will tell you if the LSB bits are random or not
    - Can be applied to audio files as well (using LSB)
  - Frequency scanning
    - Software can search for high, inaudible frequencies

Methods Of Detecting/Recovering Data

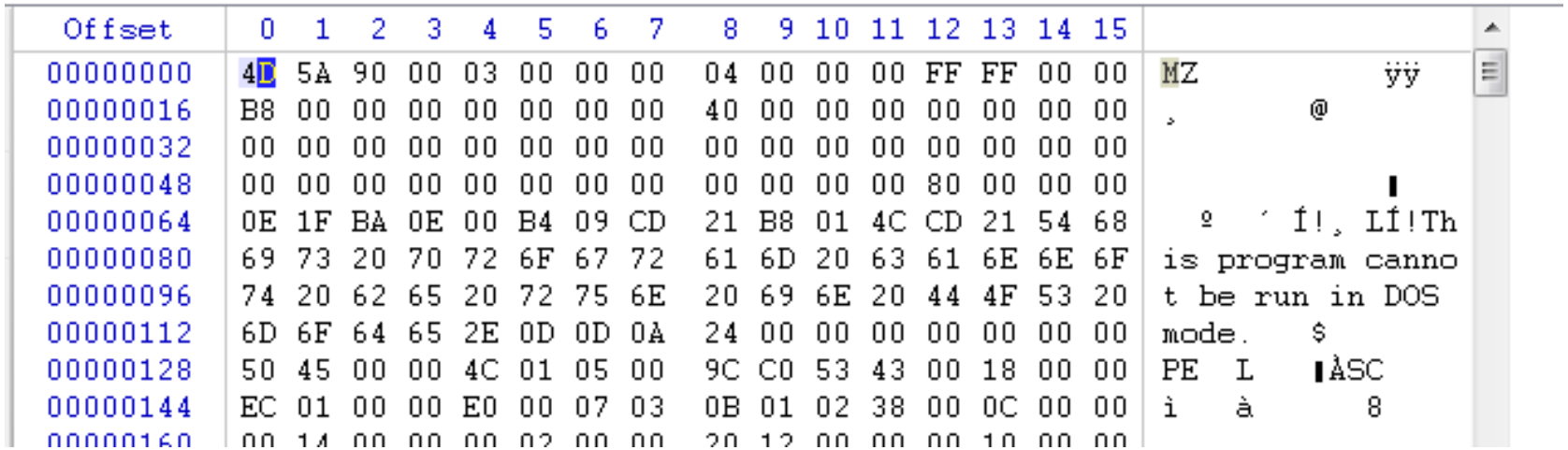
- Steganalysis Methods – Recovery
  - Recovery of watermarked data is extremely hard
    - Currently, there are very few methods to recover hidden, encrypted data.
  - Data hidden on disk is much easier to find. Once found, if unencrypted, it is already recovered
  - Deleted data can be reconstructed (even on hard drives that have been magnetically wiped)
  - Check swap files for passwords and encryption keys which are stored in the clear (unencrypted)
  - Software Tools
    - Scan for and reconstruct deleted data
    - Break encryption
    - Destroy hidden information (overwrite)

Files

- Windows uses file extensions to figure out how to open a file
  - e.g. .pdf
- However, files contain information inside them to allow other OS to process them
  - File Headers

File Header

- Example:
  - Executables have the header MZ (0x4D)



Offset	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
00000000	4D	5A	90	00	03	00	00	00	04	00	00	00	FF	FF	00	00	MZ ÿÿ
00000016	B8	00	00	00	00	00	00	00	40	00	00	00	00	00	00	00	¸ @
00000032	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
00000048	00	00	00	00	00	00	00	00	00	00	00	00	80	00	00	00	▌
00000064	0E	1F	BA	0E	00	B4	09	CD	21	B8	01	4C	CD	21	54	68	º ´ Í!¸ LÍ!Th
00000080	69	73	20	70	72	6F	67	72	61	6D	20	63	61	6E	6E	6F	is program canno
00000096	74	20	62	65	20	72	75	6E	20	69	6E	20	44	4F	53	20	t be run in DOS
00000112	6D	6F	64	65	2E	0D	0D	0A	24	00	00	00	00	00	00	00	mode. $
00000128	50	45	00	00	4C	01	05	00	9C	C0	53	43	00	18	00	00	PE L ▌ÀSC
00000144	EC	01	00	00	E0	00	07	03	0B	01	02	38	00	0C	00	00	ì à 8
00000160	00	14	00	00	00	02	00	00	20	12	00	00	00	10	00	00	

File Signatures

- Prime target for hiding data
  - E.g. hiding image files as dll's in a system folder
- Files also contain end regions, or footers
- A combination of file extension, headers and footers can be used for file recovery

File Signatures

4D 5A	MZ
COM, DLL, DRV, EXE, PIF, QTS, QTX, SYS	Windows/DOS executable file
ACM	MS audio compression manager driver
AX	Library cache file
CPL	Control panel application
FON	Font file
OCX	ActiveX or OLE Custom Control
OLB	OLE object library
SCR	Screen saver
VBX	VisualBASIC application
VXD, 386	Windows virtual device drivers

25 50 44 46	%PDF
PDF, FDF	Adobe Portable Document Format and Forms Document file **Trailers:** 0A 25 25 45 4F 46(.%%EOF) 0A 25 25 45 4F 46 0A(.%%EOF.) 0D 0A 25 25 45 4F 46 0D 0A(..%%EOF..) 0D 25 25 45 4F 46 0D(.%%EOF.)

[512 byte offset] EC A5 C1 00	[512 byte offset] ì¥Á.
DOC	Word document subheader (MS Office)

File Signatures

49 44 33	ID3	
MP3	MPEG-1 Audio Layer 3 (MP3) audio file	
52 49 46 46 xx xx xx xx 41 56 49 20 4C 49 53 54	RIFF.... AVI LIST	
AVI	Resource Interchange File Format -- Windows Audio Video Interleave file	
[29,152 byte offset] 57 69 6E 5A 69 70	[29,152 byte offset] WinZip	
ZIP	WinZip compressed archive	

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

File Hash Searching

- Databases of known good files or known bad files can be used to rapidly detect content
- `hfind`
  - Requires a database
  - HASH FILENAME
- `md5deep`
  - Creates an output file of directory and contents

File Carving

- Carving is the process of discovering and extracting files based on their header and footer signatures
- `scalpel`

	extension	case sensitive	size	header	footer
#	gif	y	5000000	\x47\x49\x46\x38\x37\x61	\x00\x3b
#	gif	y	5000000	\x47\x49\x46\x38\x39\x61	\x00\x3b
#	jpg	y	200000000	\xff\xd8\xff\xe0\x00\x10	\xff\xd9