Digital Forensics
Lecture Week 5

Hex View of Data
File Metadata

Nelson – Ch 8

Readings

- In December 2012, anti-virus programmer John McAfee was arrested in Guatemala while fleeing from persecution in Belize, over the border.
- Vice (magazine) had published an exclusive interview with McAfee "on the run" that included a photo of McAfee with a Vice reporter taken with an iPhone 4S smart phone.
- The photo's metadata included GPS coordinates locating McAfee in Guatemala, and he was captured two days later.

The Panama papers

The Panama Papers was a collection of leaked documents that showed the Pakistani prime minister had a large fortune that exceed what he got legitimately.

A team was set up to determine where the money came from, so he gave them a document that appeared to show that the money had been acquired through legitimate methods.

The issue was the font used in the document was Microsoft Calibri, a font that was released to the public in 2007 but the document's date was 2006.

Objectives

- To look at Disk Bytes as Hex
- To understand metadata
- To examine image metadata
- To examine file metadata
- To examine metadata in some documents
- To understand hashing

Hex Editors

- Editor apps expect the file to contain certain coding
- Notepad expects ascii (txt)
- WordPad expects rich text (rtf)
- Open Office expects open document format (odt)
- MS Word expects Office XML (docx)
- How do we read raw data? (no coding at all)
- In particular, disk sectors
- We need a Hex Editor

Hex Editors

- We Use Hxd
- Edits disk sectors, memory and files of any size
- http://mh-nexus.de/en/hxd/
- There is also WinHex (paid)
- http://www.x-ways.net/winhex/
- Notepad++ can also edit hex
- http://notepad-plus-plus.org/
- http://en.wikipedia.org/wiki/Hex_editor

Hex Viewers

- In Linux use xxd

```
root@kali64:~# whatis xxd
xxd (1)                  - make a hexdump or do the reverse.
```

- There are other hex viewer tools
- There are python versions

Hex Displays

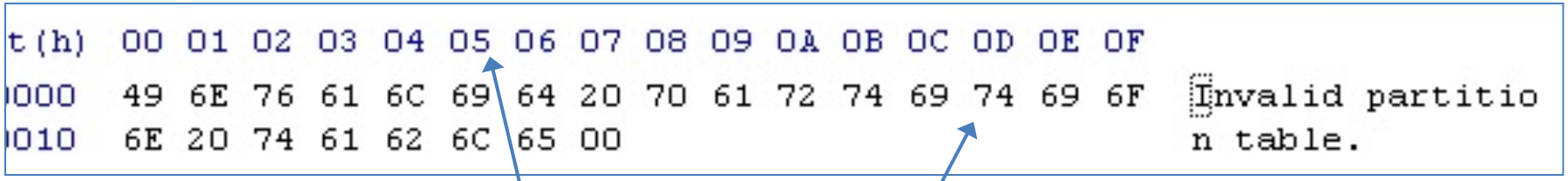
- The unit of data is the byte – 8 bits
- This can contain $2^8 = 256$ combinations
- These combinations can be represented in base 16 notation (Hex)

Decimal	0	1	2	3	…	10	11	12	..	15
Hex	0	1	2	3	…	A	B	C	..	F

- So the range of data 0 – 255 is now 00 – FF in Hex

The Hex view of data

- By using a hex editor, we see data as hex

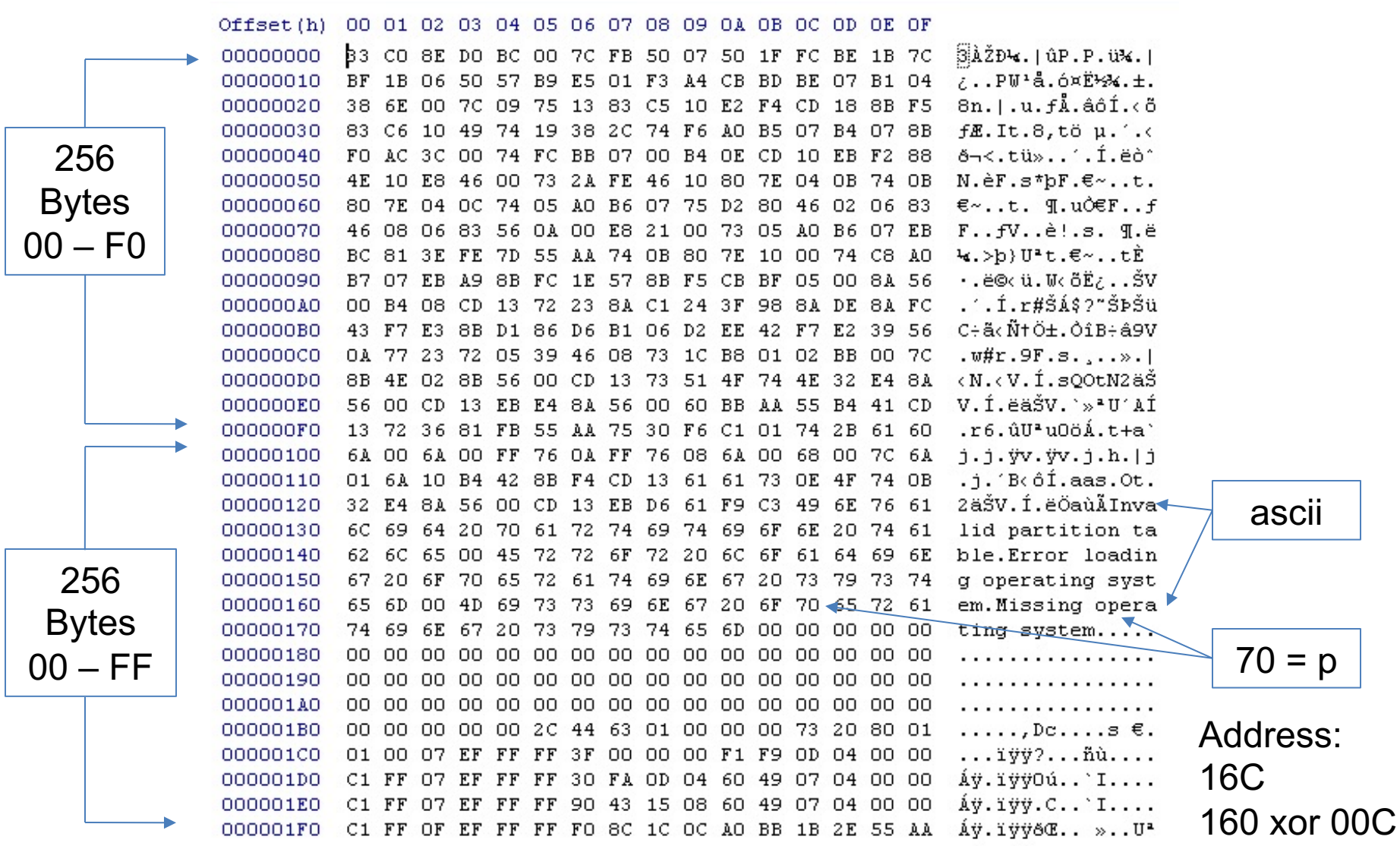


- The data address (or location) is in hex
- The data value (or content) is also in hex
- We express the content in a way that has the most meaning, here it is an error message
- hex → ascii → error message

Hex Displays #2

- To display 256 bytes we use a 16 x 16 array
- In Hex this is a (00 – 0F) x (00 - F0) array
- To display 512 bytes we use a 16 x 32 array
- In Hex this is a (00 – 0F) x (000 - 1F0) array
- The hex editor will usually display an ascii view as well

Sample 512 byte Hex Display



Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
00000000	33	C0	8E	D0	BC	00	7C	FB	50	07	50	1F	FC	BE	1B	7C	3ÀŽÐ¼.|ûP.P.ü¾.|
00000010	BF	1B	06	50	57	B9	E5	01	F3	A4	CB	BD	BE	07	B1	04	¿..PW¹å.ó¤Ë½¾.±.
00000020	38	6E	00	7C	09	75	13	83	C5	10	E2	F4	CD	18	8B	F5	8n.|.u.ƒÅ.âôÍ.‹õ
00000030	83	C6	10	49	74	19	38	2C	74	F6	A0	B5	07	B4	07	8B	ƒÆ.It.8,tö µ.´.‹
00000040	F0	AC	3C	00	74	FC	BB	07	00	B4	0E	CD	10	EB	F2	88	ð¬<.tü»..´.Í.ëòˆ
00000050	4E	10	E8	46	00	73	2A	FE	46	10	80	7E	04	0B	74	0B	N.èF.s*þF.€~..t.
00000060	80	7E	04	0C	74	05	A0	B6	07	75	D2	80	46	02	06	83	€~..t. ¶.uÒ€F..ƒ
00000070	46	08	06	83	56	0A	00	E8	21	00	73	05	A0	B6	07	EB	F..ƒV..è!.s. ¶.ë
00000080	BC	81	3E	FE	7D	55	AA	74	0B	80	7E	10	00	74	C8	A0	¼.>þ}Uªt.€~..tÈ
00000090	B7	07	EB	A9	8B	FC	1E	57	8B	F5	CB	BF	05	00	8A	56	·.ë©‹ü.W‹õË¿..ŠV
000000A0	00	B4	08	CD	13	72	23	8A	C1	24	3F	98	8A	DE	8A	FC	.´.Í.r#ŠÁ$?˜ŠÞŠü
000000B0	43	F7	E3	8B	D1	86	D6	B1	06	D2	EE	42	F7	E2	39	56	C÷ã‹Ñ†Ö±.ÒîB÷â9V
000000C0	0A	77	23	72	05	39	46	08	73	1C	B8	01	02	BB	00	7C	.w#r.9F.s.¸..».|
000000D0	8B	4E	02	8B	56	00	CD	13	73	51	4F	74	4E	32	E4	8A	‹N.‹V.Í.sQOtN2äŠ
000000E0	56	00	CD	13	EB	E4	8A	56	00	60	BB	AA	55	B4	41	CD	V.Í.ëäŠV.`»ªU´AÍ
000000F0	13	72	36	81	FB	55	AA	75	30	F6	C1	01	74	2B	61	60	.r6.ûUªu0öÁ.t+a`
00000100	6A	00	6A	00	FF	76	0A	FF	76	08	6A	00	68	00	7C	6A	j.j.ÿv.ÿv.j.h.|j
00000110	01	6A	10	B4	42	8B	F4	CD	13	61	61	73	0E	4F	74	0B	.j.´B‹ôÍ.aas.Ot.
00000120	32	E4	8A	56	00	CD	13	EB	D6	61	F9	C3	49	6E	76	61	2äŠV.Í.ëÖaùÃInva
00000130	6C	69	64	20	70	61	72	74	69	74	69	6F	6E	20	74	61	lid partition ta
00000140	62	6C	65	00	45	72	72	6F	72	20	6C	6F	61	64	69	6E	ble.Error loadin
00000150	67	20	6F	70	65	72	61	74	69	6E	67	20	73	79	73	74	g operating syst
00000160	65	6D	00	4D	69	73	73	69	6E	67	20	6F	70	65	72	61	em.Missing opera
00000170	74	69	6E	67	20	73	79	73	74	65	6D	00	00	00	00	00	ting system.....
00000180	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000190	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000001A0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000001B0	00	00	00	00	00	2C	44	63	01	00	00	00	73	20	80	01	.....,Dc....s €.
000001C0	01	00	07	EF	FF	FF	3F	00	00	00	F1	F9	0D	04	00	00	...ïÿÿ?...ñù....
000001D0	C1	FF	07	EF	FF	FF	30	FA	0D	04	60	49	07	04	00	00	Áÿ.ïÿÿ0ú..`I....
000001E0	C1	FF	07	EF	FF	FF	90	43	15	08	60	49	07	04	00	00	Áÿ.ïÿÿ.C..`I....
000001F0	C1	FF	0F	EF	FF	FF	F0	8C	1C	0C	A0	BB	1B	2E	55	AA	Áÿ.ïÿÿðŒ.. »..Uª

Objectives

- To use Linux tools on Windows
- To understand metadata
- To look at Disk Bytes as Hex
- To examine image metadata
- To examine file metadata
- To examine metadata in some documents
- To understand hashing

Data

- What is data?
  - a collection of facts
- Examples
  - Numbers
  - Words
  - Measurements
  - Observations
  - Descriptions
- Presented as
  - Qualitative (description)
  - Quantitative (number)
  - Discrete (limited values)
  - Continuous (a range)
- What is Metadata?
  - Data about the data

Metadata examples

- Images
  - Author, Location, camera type
- Files
  - Dates, author, size, folder path
- Email
  - Source ip, server, dates,
- NTFS Disks
  - Dates, update count
- Phone calls
  - Duration, location

Metadata

- The purpose of Metadata is to support the data
  - Camera metadata helps image editing apps
- Metadata can be physical or electronic
- Metadata can be indexed for searching
  - The book library ISBN
- Metadata can be scrubbed from documents
  - Law firm reports for clients
- Collecting metadata will leave a metadata trace
  - See Locard Week 2

Metadata #2

- Structural metadata describes the data container
  - How the data is structured
  - An example is the Tag:Value pair
- Descriptive metadata describes individual instances
  - the data content of this data sample is “city":"Chatswood"
- Administrative Metadata describes how the data is used
  - Origin, Category, Access rights
- This week we will look at metadata in files
- We will start with graphic files

Objectives

- To look at Disk Bytes as Hex
- To understand metadata
- To examine image metadata
- To examine file metadata
- To examine metadata in some documents
- To understand hashing

Graphic formats

- Pictures (images) can be saved as a file
- There are two ways to do this
- Bitmap – dots (pixels) as a 2D array
- Vector Graphics – define shapes such as lines (length + direction)
- Bitmap editors – MS paint, Photoshop, gimp
- Vector editors – CorelDraw, adobe illustrator
- Bitmap viewers – MS Office
- Vector viewers – pdf-Xchange, Adobe Reader

Graphic files

- Black and White Bitmaps have one bit per pixel.
- A 640x480 pixel image requires 307200 bits = 0.37MB
- Real colour has 24 colour bits for each pixel.
- A real colour 640x480 image needs 7.37 MB
- A vector graphic of a simple image is much smaller than the pixel version.

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

Scaling images

- Bitmap files do not scale.
- Increasing the number of bits using an editor does not increase the image quality – the image looks pixelated
- Decreasing the number of bits loses information.
- Vector files scale well.
- You can scale a full stop dot up to fill a page and it still has a perfect round circumference with no pixelation

Graphic compression

- Graphic compression is like text compression
- Consider a green line of length 100 bits.
- Instead of recording in the file 100 green bits
- we record one green bit and a x 100 instruction
- Some compression is lossless
- Uncompressing retrieves the original
- Some compression is lossy
- Uncompressing retrieves only part of the original.

Graphic file types

- Pixels
  - BMP – no compression, large and lossless, windows
  - TIFF – tagged bitmap, large and lossless
  - GIF , PNG – simple graphics, 8 bit colour
  - JPEG, JPG – small and lossy, used to share photos
  - EXIF – Combines jpeg and tiff for camera metadata
- Vectors
  - SVG – open standard that includes scripting

Graphic File hex signature tags

- A graphic file often has an identifying tag near the start

```
G:\Forensics>xxd -l 16 logo.gif
0000000: 4749 4638 3961 dc00 3200 f700 00ff c35c  GIF89a..2......\

G:\Forensics>xxd -l 32 "MS Office Meta Data.jpg"
0000000: ffd8 ffe0 0010 4a46 4946 0001 0101 0096  ......JFIF......
0000010: 0096 0000 ffdb 0043 0001 0101 0101 0101  .......C........

G:\Forensics>xxd -l 32 IMAG1672a.jpg
0000000: ffd8 ffe0 0010 4a46 4946 0001 0101 0048  ......JFIF.....H
0000010: 0048 0000 ffe1 1242 4578 6966 0000 4d4d  .H.....BExif..MM
```

Objectives

- To look at Disk Bytes as Hex
- To understand metadata
- To examine image metadata
- To examine file metadata
- To examine metadata in some documents
- To understand hashing

Metadata sample – Camera image

Tag	Value
Manufacturer	CASIO
Model	QV-4000
Orientation (rotation)	top - left [8 possible values[21]]
Software	Ver1.01
Date and Time	2003:08:11 16:45:32

Forensics and metadata

- How does the metadata get captured?
- Where is it stored?
- What is the data structure?
- Does it have tag = value pairs?
- What are the tag codes?
- What forensic tools are available?
- What are the current forensic issues?

File Analysis

- A three step process
- #1 Recover any hidden/encrypted/deleted files
  - Look for secret disks
  - Look for secret partitions
  - Look for secret files

File Analysis #2

- #2 Recover basic file system information
  - file system type, metadata location, sector size
  - number of partitions, partition size
  - we use the Sleuthkit tools - mmls, fsstat (see later)
- #3 Recover mismatched file information
  - where the file header does not match the file extension
  - we use the file command on each file
  - file requires a magic file; listing all known file types

file extensions

- What happens when you click on Trade_Secrets.txt?
- by magic the notepad program is loaded and handed the file to open
- How does an operating system know which program to associate with a file?
- The file name extension is used (.txt above) kept in the registry
- What if a suspect alters an extension?
  - trojan.exe → cars.txt
- by more magic, forensics can expose this forgery

magic files

- Many file formats have headers with metadata
  - file authoring
  - camera files have camera settings, gps location, etc
- As there are many file types, there are many headers
- To simplify all this, magic numbers evolved
- Originally these were two bytes at the start of the file
- The identifiers are now stored in a separate magic file

- http://en.wikipedia.org/wiki/File_format#Magic_number

magic files #2

```
root@kali64:~# whatis magic
magic (5)            - file command's magic pattern file
```

- The magic patterns are found in /usr/share/misc/magic
- each file type signature is described as a comment

magic file examples – text file

```
C:\Forensics_Graham>type test.txt
This is a text file
With two lines of text

C:\Forensics_Graham>xxd test.txt
0000000: 5468 6973 2069 7320 6120 7465 7874 2066  This is a text f
0000010: 696c 6520 0d0a 5769 7468 2074 776f 206c  ile ..With two l
0000020: 696e 6573 206f 6620 7465 7874 200d 0a    ines of text ..
```

- 20 = space
- 0D= Carriage Return (CR), 0A = Line Feed (LF) / new line
- The End of Line marker (EOL) is 0D0A in Windows
- This text file ends with a single EOL

magic file examples – graphic files

```
G:\Forensics>file logo.gif
logo.gif: GIF image data, version 89a, 220 x 50

G:\Forensics>file "MS Office Meta Data.jpg"
MS Office Meta Data.jpg: JPEG image data, JFIF standard 1.01

G:\Forensics>file IMAG1672a.jpg
IMAG1672a.jpg: JPEG image data, JFIF standard 1.01
```

- the exif in the last file is not shown

magic file example – MS Word

```
Magic entry for Microsoft Office XML

# start by checking for ZIP local file header signature
0                string          PK\003\004
# make sure the first file is correct
>0x1E            string          [Content_Types].xml
```

```
C:\Forensics>xxd -l 64 Test.docx
0000000: 504b 0304 1400 0600 0800 0000 2100 0924  PK..........!..$
0000010: 8782 8101 0000 8e05 0000 1300 0802 5b43  ..............[C
0000020: 6f6e 7465 6e74 5f54 7970 6573 5d2e 786d  ontent_Types].xm
0000030: 6c20 a204 0228 a000 0200 0000 0000 0000  l ...(..........
```

magic file example – Windows Executable

```
C:\Forensics>xxd -l 144 grep.exe
0000000: 4d5a 9000 0300 0000 0400 0000 ffff 0000  MZ..............
0000010: b800 0000 0000 0000 4000 0000 0000 0000  ........@.......
0000020: 0000 0000 0000 0000 0000 0000 0000 0000  ................
0000030: 0000 0000 0000 0000 0000 0000 8000 0000  ................
0000040: 0e1f ba0e 00b4 09cd 21b8 014c cd21 5468  ........!..L.!Th
0000050: 6973 2070 726f 6772 616d 2063 616e 6e6f  is program canno
0000060: 7420 6265 2072 756e 2069 6e20 444f 5320  t be run in DOS
0000070: 6d6f 6465 2e0d 0d0a 2400 0000 0000 0000  mode....$.......
0000080: 5045 0000 4c01 0700 e45a 5852 0034 0300  PE..L....ZXR.4..
```

- MZ at 00
- Jump to (80) at 3C (compiler dependant)
- DOS ascii text message at 4D
- PE (Portable Executable) at or after 80 (compiler dep)

The file file

- In Linux, the magic file is accessed using file

```
C:\Forensics>file Test.docx
Test.docx: Microsoft Word 2007+
```

- file * # list all files

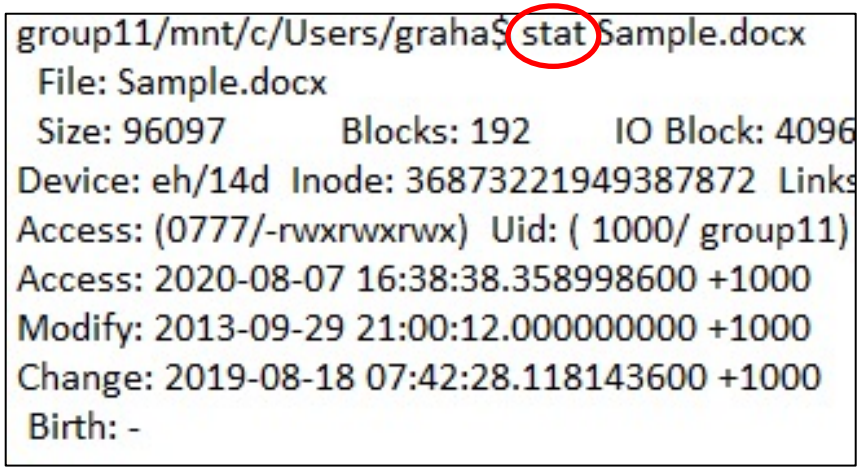
```
trade_secrets.txt: ISO-8859 English text, with
very long lines, with CRLF line terminators

ls.exe: PE32 executable (console) Intel 80386
(stripped to external PDB), for MS Windows

cmarko-tskintro.pdf: PDF document, version 1.4
```

File Date/Time

- In Linux you can stat a file to see the three date/time stamps



- However in Linux a suspect can touch a file to change one or more of these date/time stamps
- However the file header may also contain another date/time stamp as metadata
  - this may be missed by the suspect using touch

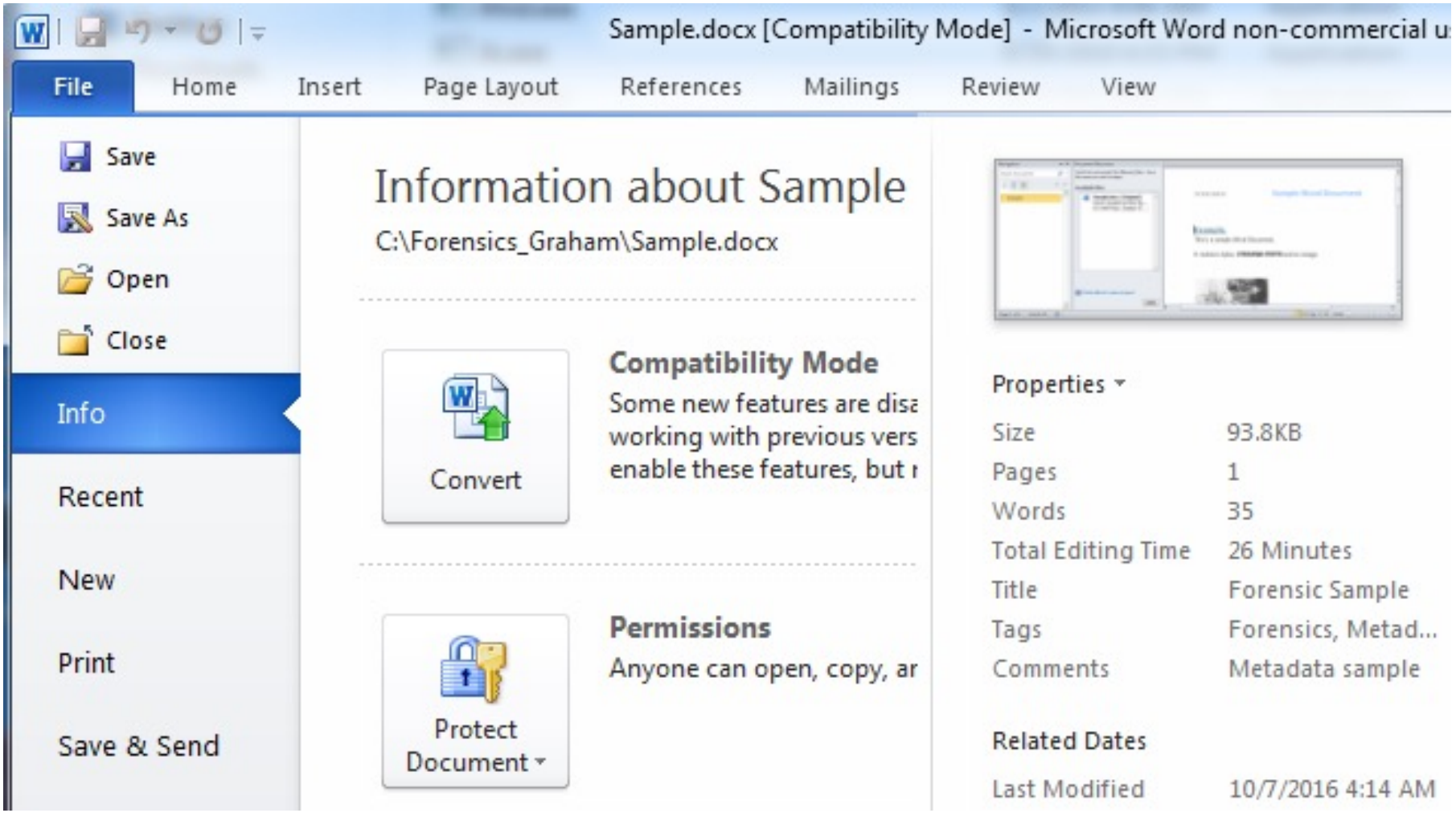
Objectives

- To look at Disk Bytes as Hex
- To understand metadata
- To examine file metadata
- To examine metadata in some documents
- To understand hashing

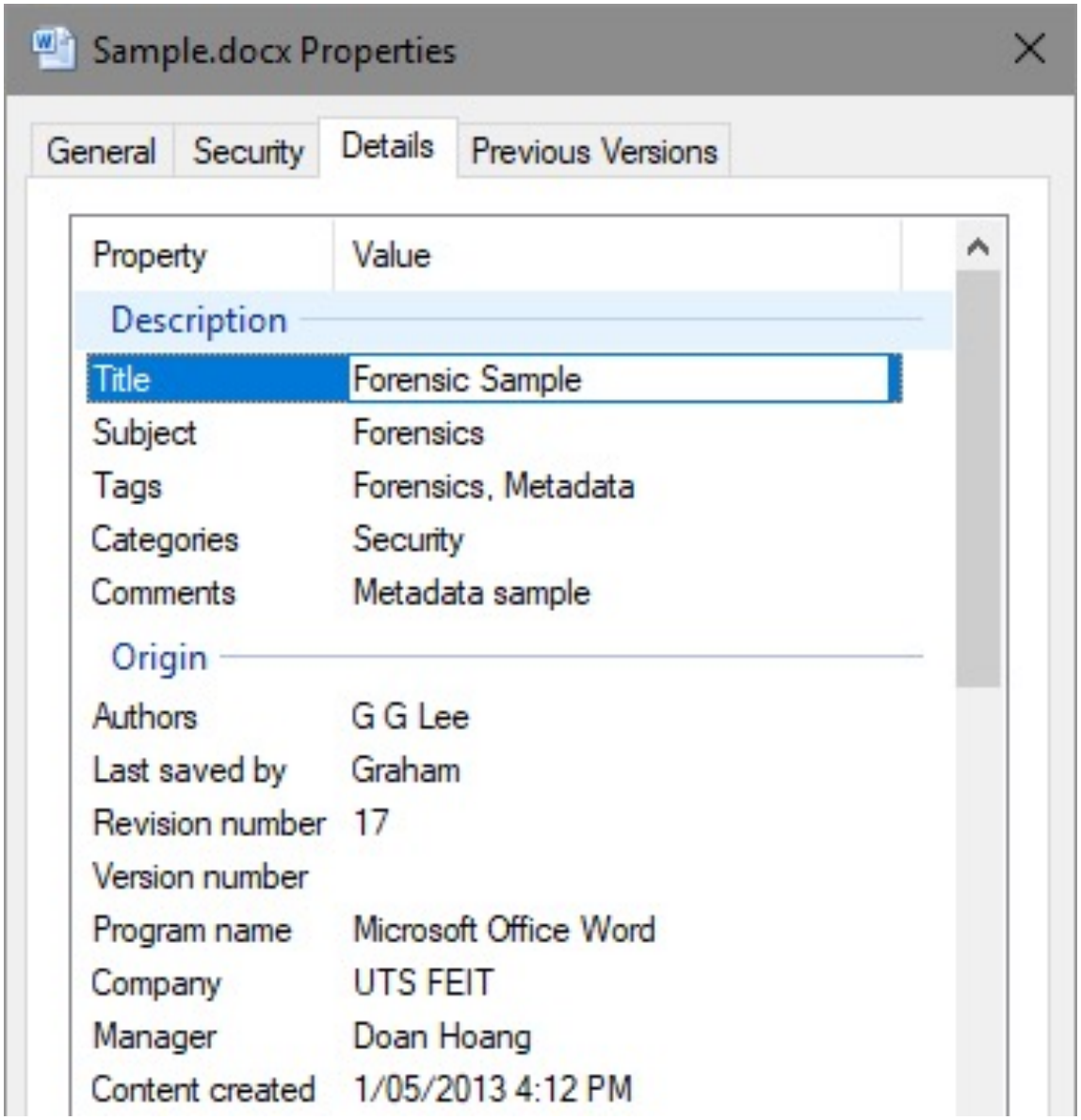
MS Office Files

- MS Office files are three xml zipped files
  - examine with the 7-zip tool.
- app.xml
  - application properties, such as page count
- core.xml
  - author, date altered, print date
- word/media
  - contains any images
- See Readings

docx metadata in MS Word

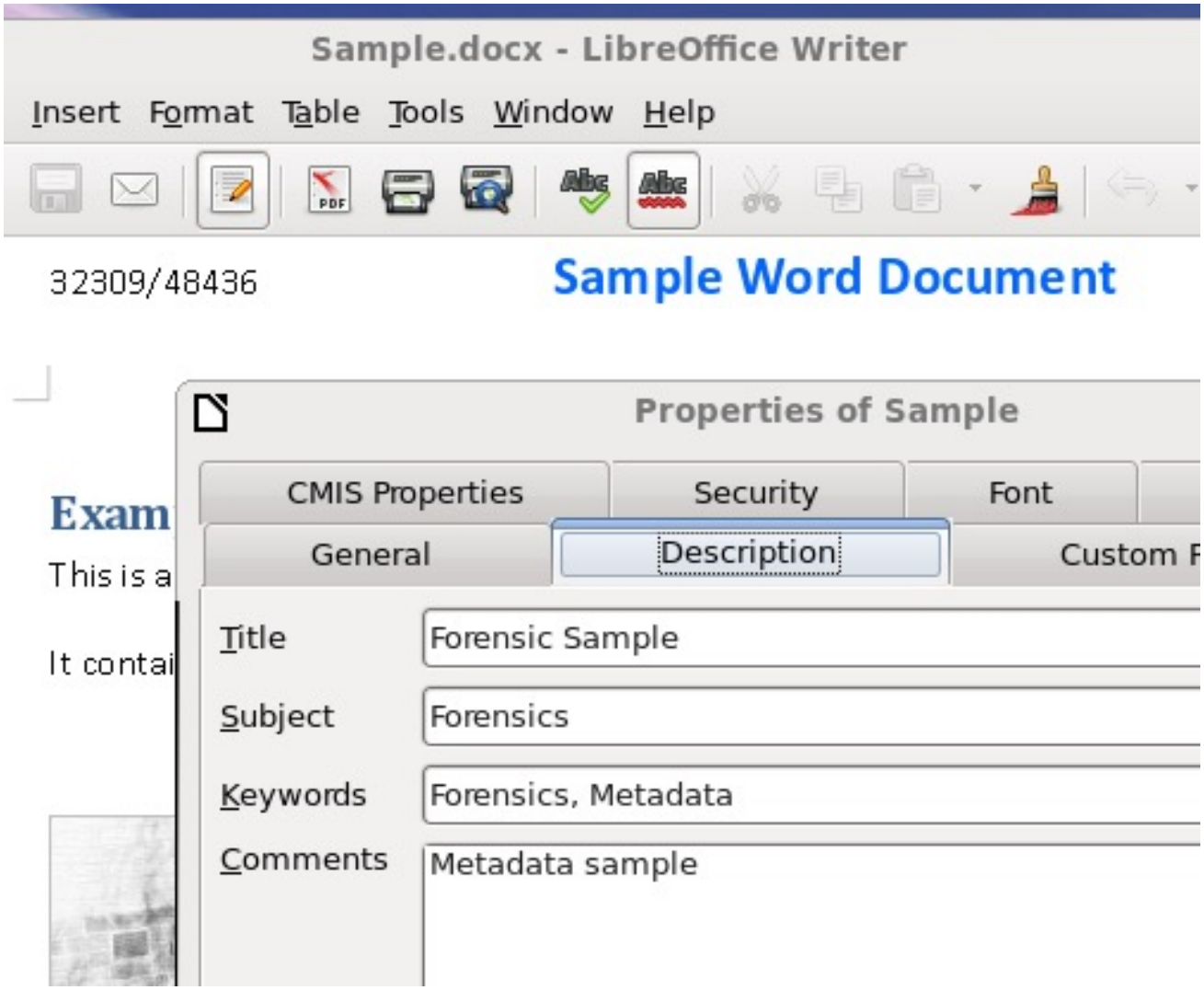


docx metadata in Win 10



docx metadata in LibreOffice

- Select File, Properties, Description



PDF Files

- The PDF structure is complex
- Header, objects and a trailer
- Uses internal scripting commands
- /Launch will launch a program
- /JavaScript will launch JavaScript
- /OpenAction will run a script on open
- http://en.wikipedia.org/wiki/Portable_Document_Format
- See Readings

PDF Headers

- The PDF Header starts with %PDF-1.x

```
C:\Forensics>xxd -l 16 "Evidence ACPO.pdf"
0000000: 2550 4446 2d31 2e35 0d25 e2e3 cfd3 0d0a  %PDF-1.5.%......
```

- Adobe versions also contain 25 e2 e3 cf d3

Google this

- The Trailer contains an End of File marker

```
25 45 4F 46 0D 0A                                  %EOF..
```

Malicious PDFs

- Malware can:
  - embed an exe into the pdf
  - embed malicious JavaScript in the pdf

Thinking like an attacker

- I want to be invisible ⇒ evasion tricks
- I want to kill PDF files and/or Reader ⇒ denial of services
- I want to steal information (read + send) ⇒ information leakage
- I want to corrupt my target ⇒ egg dropping
- I want to overrun the target ⇒ code execution

- https://doi.org/10.1007/s11416-009-0128-2

Camera Files

- Exchangeable image file format (exif)
- Used by cameras, Smartphones and scanners
- Developed from JPEG and TIFF
- Includes Geolocation
- Camera files are exif files, but save as jpegs
- The exif standard has many shortcomings
- Many cameras use their own format instead
- https://en.wikipedia.org/wiki/Exchangeable_image_file_format

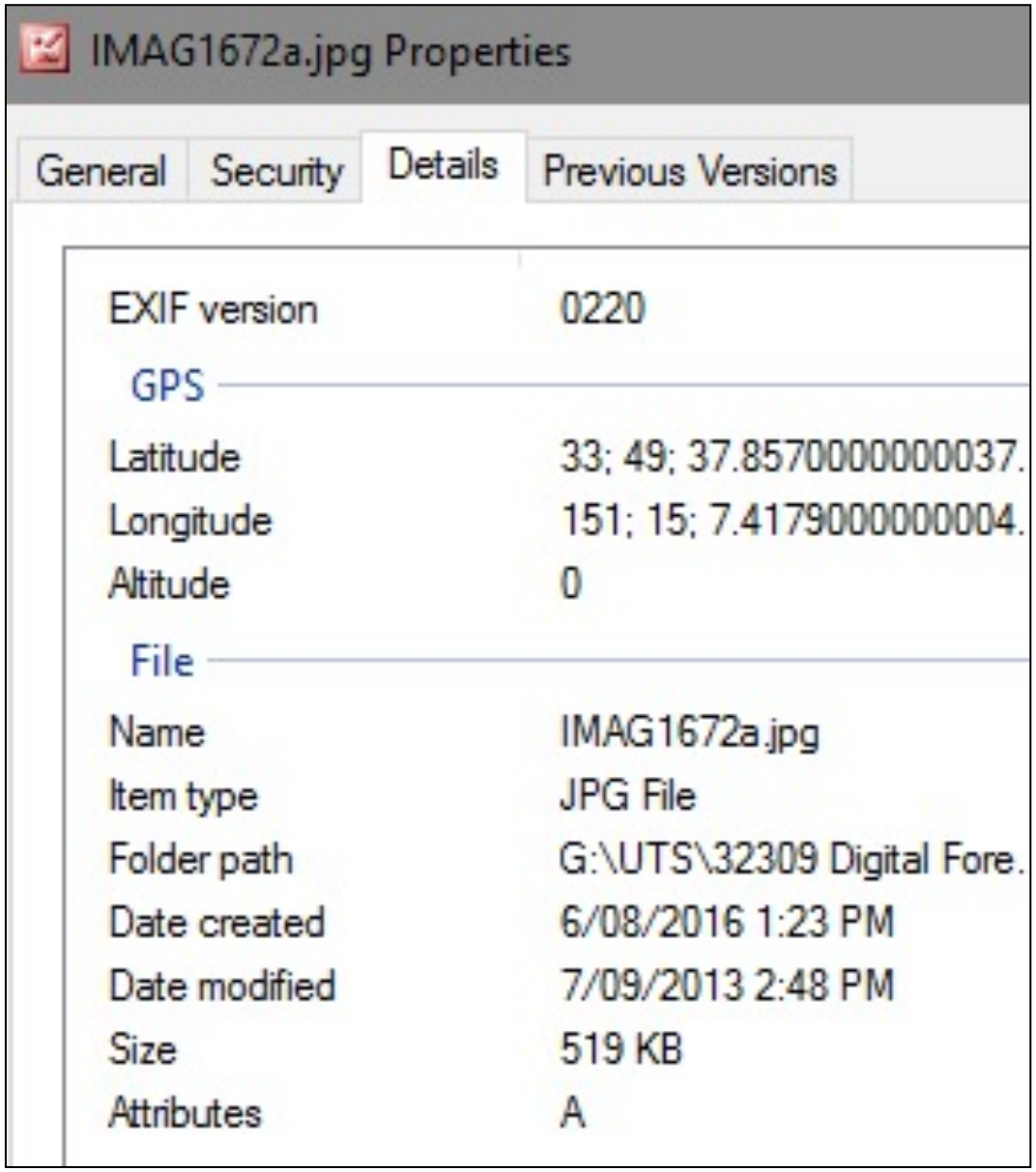
Exif Files

- file thinks this is a jpg
- The stat date/times can be altered
- exif reveals the metadata date

```
Holland America 1.jpg
 ExifVersion  0220
 ComponentsConfiguration
 ExifImageWidth  3264
 DateTimeOriginal  2013:03:16 18:3
 DateTimeDigitized  2013:03:16 18:
 GPSInfo  Lat: -33.0 Long: 151.0
 FlashPixVersion  0100
 ISOSpeedRatings  74
 ExifOffset  2166
 FocalLength  (431, 100)
 ExifImageHeight  2448
 Make  HTC
 Model  HTC Sensation Z710a
 SubsecTimeOriginal  00
 SubsecTimeDigitized  00
 YCbCrPositioning  1
 59932  LÛ
 ColorSpace  1
```

Exif Files #2

- Windows 10 includes a built-in exif viewer under file properties.



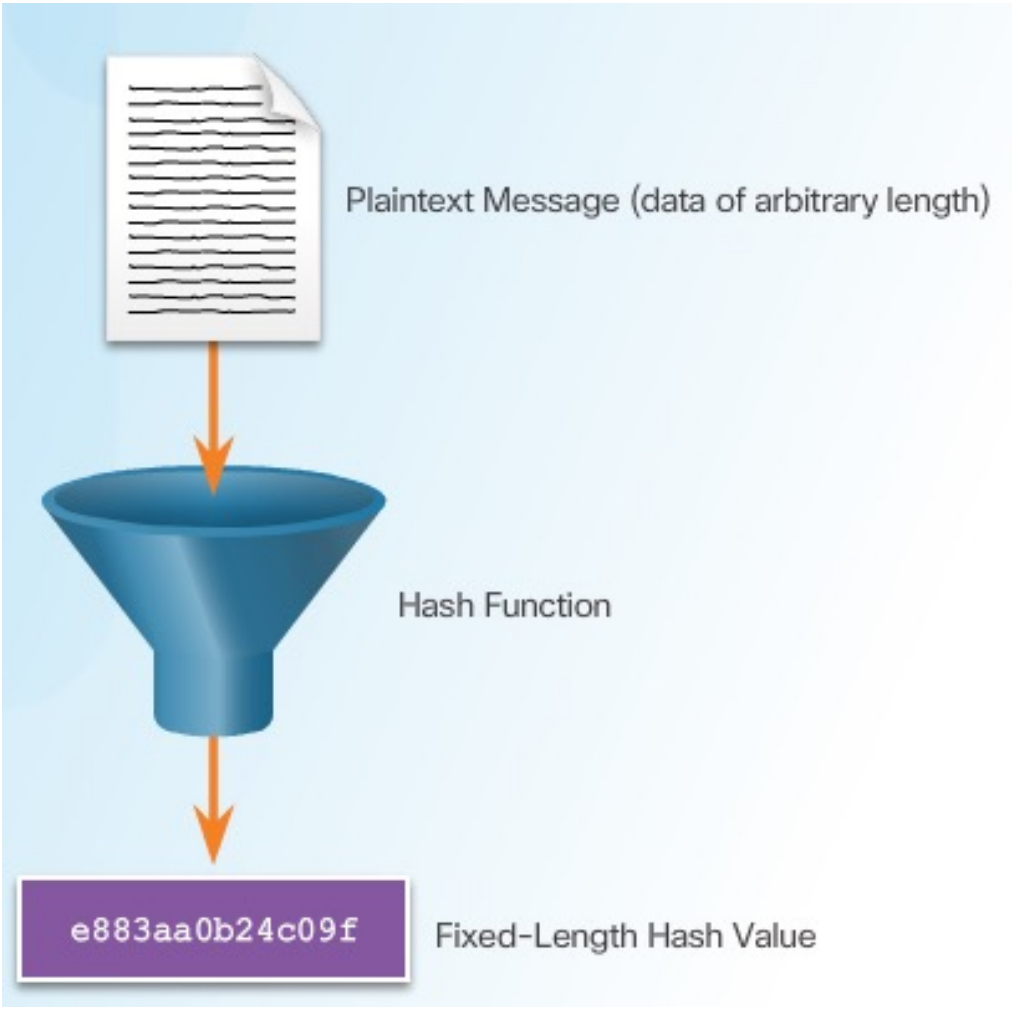
Email metadata

- Email *.EML files can be found on the client.
- Emails contain headers that can contain useful information
  - the mail server ip used to send the email
  - IP address of the person receiving the email
  - IP addresses that the email was passed to
  - email authentication.

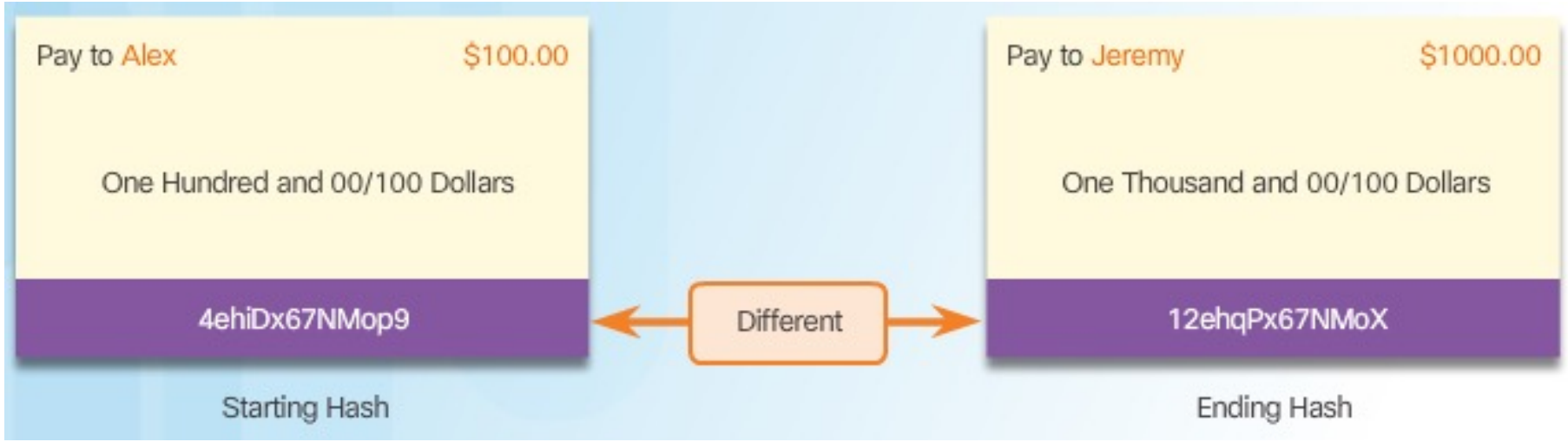
Objectives

- To look at Disk Bytes as Hex
- To understand metadata
- To examine file metadata
- To examine metadata in some documents
- To understand hashing

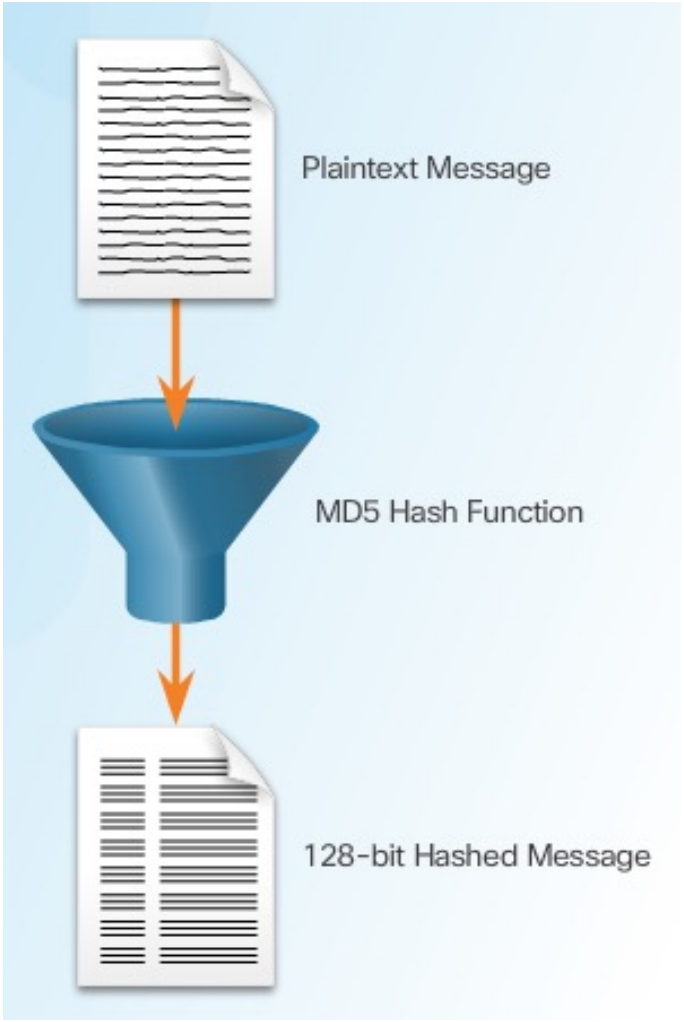
Cryptographic Hash Function



Using a hash protects integrity



Message Digest 5 Algorithm



Secure Hash Algorithm

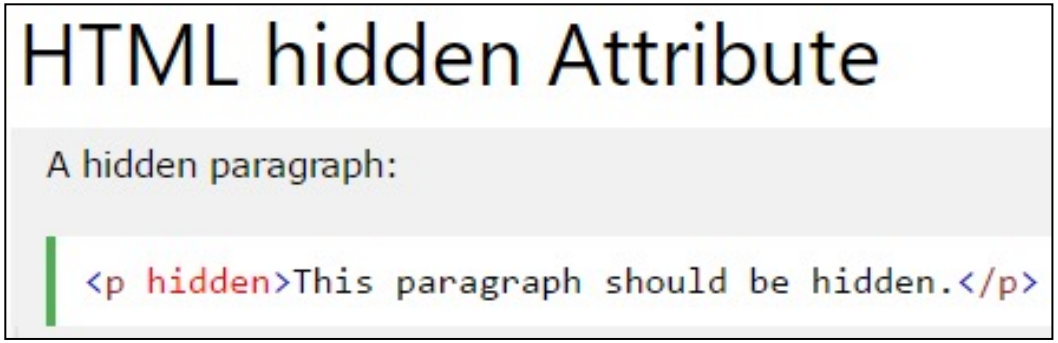
MD5 Versus SHA

Forensic hashing

- We can use hashing to ensure the integrity of evidence
- We can pickup changes made to images for instance by a suspect or a virus
- However if we hash all of a live disk, a second hash will be different.
  - Because a live disk is always changing.
- We use forensic hashing to hash many parts of a disk and keep the results in a table of hashes.

Steganography

- “Hiding in plain sight”
- We insert data into an image file.
- It is invisible to the naked eye
- Two methods: Insertion and substitution
- Insertion - html has a hidden attribute
- The data inside the tag is hidden



Substitution Steganography

- Replaces least significant bits (lsb) in a bitmap image with data
- Can also embed data in mp3 audio files.
  - Mp3 has metadata using the ID3 format
- Other methods adjust spacing characters in text files.
- Common steno programs are password protected
- It is often hard to detect steno using an IDS

Watermarking

- Commercial programs such as Photoshop can watermark an image to detect a copyright infringement
- Watermarks can record:
- the copyright owner
- the distributor
- the distribution chain
- the purchaser of the document, game or music

DRM

- Digital Rights Management (DRM) supply an encryption key for paid downloads such as pdf courseware.
- The document will not open on any other device
- This technique is also used for Gaming

FIN

- Sayōnara