Cyber-Forensics
The Basics
CERTConf2006
Tim Vidas

Because you have to start somewhere
Who are we?

- **Tim Vidas**
  - Sr. Tech. Research Fellow
  - UNO/PKI/NUCIA
  - Certs: CISSP, 40xx, Guidance, AccessData etc.
  - Instructor: UNO, Guidance, LM RRCF

- **Joe Wilson**
  - Recent Graduate (MS/MIS)
  - RRCF
NUCIA

- Nebraska University Consortium on Information Assurance
- IA full time
- Traditional university coursework in IA, Crypto, Forensics, Secure Administration, Certification and Accreditation, etc
- STEAL Labs
- “Other work”
- Most of us are ‘around’ CERTconf.
Who are you?

- Who are you?
- Where do you work?
- What do you do?
- How many of you are planning on attending all “Forensics” sessions?
- What are you expecting to get out of them? (I’ll try to be accommodating)
The learning theory:

• A technical, practical, hands-on approach
• ‘Technical’ means the class(es) will either require or provide a significant amount of technical expertise.
• ‘Practical’ implies that the information covered should provide you with the capacity to conduct many “cyberforensic activities”.
• ‘Hands-on” the additional hands-on component provides an active experience in which you will immersed in related exercises.
  — The best way to learn is by doing.
Disclaimer

• Even though this class touches on quite a few legal topics – nothing should be construed as advice or legal instruction.

• Before performing many of the skills learned this week on a computer other than your own, you may need to seek permission (possibly written) and or seek advice from your own legal counsel.
Whereas **computer forensics** is defined as “the collection of techniques and tools used to find evidence in a computer”,

**digital forensics** has been defined as “the use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitation or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations”
What is Cyberforensics?

• This really depends on the point of view…
• Traditionally Cyber forensics involves the
  – preservation,
  – collection,
  – validation,
  – identification,
  – analysis,
  – interpretation,
  – documentation and
  – presentation
• …of computer evidence stored on a computer.

• “Forensics is the application of science to the legal process.”
  – Jim Christy, DCCI
Rapid-Response Cyberforensics

• Characterized by:
  – Live-response
  – Military-type contexts
    • But not of necessity
  – Judicious *a priori* planning
    • Prior strategic incident response planning
    • Requisite training in
      – Basic forensic procedures
      – Live-response
      – Network forensics
    • Continued updating of skills as technology changes
    • Technically adept with a diversity of tools & toolkits
According to the CFEWG curriculum group there are three perspectives of cyberforensics:

- Law enforcement
  - FBI/IRS
- Business/Industry
  - Cisco
- Military/counterintelligence
  - AF OSI/NSA

Although not mutually exclusive, each can have its own thrust.

“academia” is becoming a fourth...
Viewpoint
Viewpoint

• Each perspective has different objectives, even though there is overlap, the approaches of each remain mainly ah-hoc and uncoordinated
• Technology is vendor-driven
• No industry certification
• No standards
  – ASCLAD – for labs
• Interesting situations with the court system
  – Who is more believable?
  – Evidence isn’t questioned
Coverage from OS perspective

• Windows
  – 95% of cases involve Windows (FBI)
  – Topics
    • File systems: FAT & NTFS
    • Multiple tools:
      – Commercial
      – Freeware
        » Windows & Linux
    • Live response
    • Network forensics
Topics we will cover…

• We’re going to start by establishing a basis for cyber-forensics
  – Hexadecimal notation
  – Traditional “post-mortem” forensics
    • Duplication
    • Analysis
    • File systems
    • Footprints
    • Etc

• Then build upon this foundation and explore other avenues

• Generally speaking, if you don’t know how a particular tool is working ‘behind the scenes’ you might not be able to hold weight on the witness stand (or corporate report, or ______)
Cybercrime & Cyberwarfare

• “Information warfare specialists at the Pentagon estimate that a properly prepared and well-coordinated attack by fewer than 30 computer virtuosos strategically located around the world, with a budget of less than $10 million, could bring the United States to its knees.”

Center for Strategic & International Studies (CSIS)
http://www.csis.org/pubs/cyberfor.html
Cybercrime & Cyberwarfare

• “Such a strategic attack, mounted by a cyberterrorist group, either substate or nonstate actors, would shut down everything from electric power grids to air traffic control centers.”
Scope of the Problem

- In 1990 a computer hard drive seized in a criminal investigation would contain approximately 50,000 pages of text.
- The same hard drives now contain 5 million to 50 million pages of data.
  - But the ability of these agencies to retain computer talent is seriously jeopardized by the compensation packages offered by the private sector.

Center for Strategic & International Studies (CSIS)
http://www.csis.org/pubs/cyberfor.html
Computer Crime

• Sample of computer crimes from 2001
  – Demoted employee installs a logic bomb, which later deactivates hand-held computers used by the sales force.
  – eBay
    • User advertises goods, but on receiving payment never ships the goods.
    • Advertised collectibles turn out to be fakes
  – Disgruntled student sends threatening emails, leading to school closing down.
  – Ring of software pirates use web site to distribute pirated software

Computer Crimes

- Software company employee is indicted for altering a copyright program to overcome file reading limitations
- Hacker accesses 65 U.S. Court computers and downloads large quantities of private information.
- Hacker accesses bank records, steals banking and personal details.
- 15 year old boy runs scripts that invoke DOS against eBay, Yahoo!, AOL, etc.

• Moral: NO SUCH THING AS TYPICAL COMPUTER CRIME.
• Must be flexible in your response

Taxonomy of Computer Crime Scenes

• Computer Crime
  – Computer used to conduct crime
    • Examples?
  – Computer is target of crime
    • Examples?

• Response
  – Live: real-time
  – After the fact
Introduction

• Computer forensics involves
  – Preservation
    • Evidence changed, court case is gone
  – Identification
    • Of the 100,000 files, what is evidence of a crime?
  – Extraction
    • Take the evidence off the hard drive for presentation
  – Documentation
    • Document what you found to present in court
  – Interpretation
    • Interpret the evidence in light of the charges

• As much art as science
Goals (Questions) of Forensic Analysis

• Identify root cause of an event to ensure it won’t happen again
  – Must understand the problem before you can be sure it won’t be exploited again.

• Who was responsible for the event?

• Most computer crime cases are not prosecuted
  – Consider acceptability in court of law as our standard for investigative practice.
  – Ultimate goal is to conduct investigation in a manner that will stand up to legal scrutiny.
  – Treat every case like a court case!

Kruse & Heiser, 2002
Cyberforensics Procedures

- Cyberforensics is a large, complex problem composed of various flexible steps
- Each step has an input and an output
Preparation → Detection → IR Team Informed → First Response
Secure System → Create Response Strategy
Begin Evidence Acquisition → Incident?
Duplicate → Duplication Required?
Begin Investigation → Report
Feedback to Preparation & Secure System

Adapted from Mandia & Prosise 2003
Is this Correct?
Preparation

• What to do before the incident
  – Incident response plan
    • What to do in case of
      – User incident
        » User or customer reports problem
      – Application incident
        » Web page changed, etc.
      – System incident
        » Virus
        » Server down
      – Denial-of-service attack
      – Hostile code
      – Unauthorized access
      – Network probes
Preparation

• What to do before the incident
  – Incident response team
    • Systems administrators
    • Forensic analysts
    • Users
    • Managers
  – May have to wear more than one hat
Detecting Incidents

• You detect something you believe to be an incident
  – Something outside the scope of normal operation
  – Now what?

• DO UNTIL DONE
  – Document everything
  – Document everything
  – Document everything
Incident Detecting

• Follow a well-defined methodology
  – Care and due diligence must proceed with each case
  – TREAT EACH CASE AS IF IT MAY END UP IN COURT
  – Don’t begin analysis, decide you have a problem, THEN start handling it as evidence
    • TOO LATE by then, because you have changed the “scene of the crime”.
    • Defense attorney won’t care whether this was done accidentally or not.
Incident Detection

• How to document
  – Create a notification checklist
  – Assure you won’t miss any details
  – Facts to include:
    • Time & Date
    • Who or what is reporting the incident
      – User, sysadmin, IDS…
    • When incident is suspected to have occurred
    • Hardware/software
    • POC
Chain-of-custody

- CRITICAL that documentation regarding how evidence is handled.
  - Establishes continuity of who/what/where RE evidence
    - Who collected the evidence
    - What comprises the evidence
    - When evidence was collected
    - If hardware (take a photo)
      - Make, model, serial #
    - Description of the evidence, technical information
    - Name and signature of individual receiving evidence
    - Case number & tag (bag & tag)
    - If electronic, cryptographic hashes

Mandia & Prosise, 2003
Chain-of-custody

- How to bag & tag electronic evidence?
  - Cryptographic hash of the electronic file
  - More on this stuff a bit later
  - Time & date stamp before and after capture
## Evidence Checkout Log

<table>
<thead>
<tr>
<th>Item</th>
<th>Date</th>
<th>Time</th>
<th>Location</th>
<th>Name</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dell Inspiron 8000 SN# 4005</td>
<td>10/8/2002</td>
<td>13:05</td>
<td>Locked up in STEAL Lab cabinet</td>
<td>Vidas</td>
<td>Safekeeping</td>
</tr>
<tr>
<td>Dell Inspiron 8000 SN# 4005</td>
<td>10/9/2002</td>
<td>8:02</td>
<td>Removed</td>
<td>Vidas</td>
<td>Analysis</td>
</tr>
<tr>
<td>Dell Inspiron 8000 SN# 4005</td>
<td>10/9/2002</td>
<td>15:33</td>
<td>Locked up in STEAL Lab cabinet</td>
<td>Vidas</td>
<td>Safekeeping</td>
</tr>
<tr>
<td>Dell Inspiron 8000 SN# 4005</td>
<td>10/11/2002</td>
<td>7:30</td>
<td>Removed</td>
<td>Nicoll</td>
<td>Analysis</td>
</tr>
<tr>
<td>Dell Inspiron 8000 SN# 4005</td>
<td>10/11/2002</td>
<td>12:00</td>
<td>Locked up in STEAL Lab cabinet</td>
<td>Nicoll</td>
<td>Safekeeping</td>
</tr>
</tbody>
</table>

Adapted from Kruse & Heiser, 2002
Handling Evidence

• Chain-of-Custody
  – Goal is to protect the integrity of your evidence
  – Make it difficult for the defense attorney to successfully argue that the evidence was tampered with it while it was in your custody

• Document following questions
  – Who collected the evidence?
  – How was it collected? From where was it collected?
  – Who took possession of it?
  – How was it stored and protected in storage?
  – Who took it out of storage and why?
Chain-of-Custody

• Be meticulous
  – Defense attorney will cross-reference with other documents to determine any inconsistencies

• Fewer people who have access to your evidence or locker room, the better.
  – Defense attorneys will argue otherwise…
Chain-of-Custody

• What does a typical CoC list look like?

• CoC is quite a bit different with digital evidence these days…
First Response

• You’ve detected an incident, now what?
  – Verify incident and related information
  • Initiate network monitoring if appropriate
    – IDS
    – Sniffer
  – Users involved
  – Business impact if any
Formulate/Execute Response Strategy

- Your response strategy should be driven by your incident response plan
  - If you don’t have one, you must develop on the fly.
  - Select the most appropriate strategy
    • Best if you have thought about this beforehand
  - Context determines whether to do a **live response** or a **off-line media analysis** after forensic duplication
    • Big difference between the two, notwithstanding legal implications
Formulate/Execute Response Strategy

• Determine
  – How serious the problem is
  – Sensitivity of the compromised information
  – Potential offenders
  – Whether the incident is public or private
    • Internal network vs. web page
  – Level of access gained by intruder
  – Skill of the intruder
  – Level of tolerable downtime
    • Determines live response vs. offline
  – $$$$$ lost
Formulate/Execute Response Strategy

• Incident:
  – DOS

• Example
  – SMURF attack

• Strategy
  – Reconfigure router to minimize effect of flooding
  – Establishing perpetrator too costly

• Likely outcome
  – Reconfiguration reduces effect of flooding

Mandia, Prosise & Pepe, 2003
Formulate/Execute Response Strategy

• Incident:
  – Unauthorized use

• Example
  – KPorn surfing from company workstation

• Strategy
  – Perform forensic duplication
  – Offline analysis
  – Interview user

• Likely outcome
  – Suspect identified and evidence collected for disciplinary action.

Mandia, Prosise & Pepe, 2003
Formulate/Execute Response Strategy

• Incident:
  – Computer intrusion

• Example
  – Buffer-overflow gives intruder root access to critical system

• Strategy
  – Monitor intruder activities
  – Isolate the machine, reduce problem scope
  – Secure and recover the system

• Likely outcome
  – Vulnerability identified, system recovered.

Mandia, Prosise & Pepe, 2003
Formulate/Execute Response Strategy

• Incident:
  – Stolen information

• Example
  – Stolen CC numbers from company database

• Strategy
  – Issue public statement
  – Perform forensic duplication & analysis
  – Contact LE

• Likely outcome
  – LE agents participate in investigation
  – Systems offline until problem resolved.
Considerations

• Presenting strategies to management, consider
  – Downtime
    • Network/system
    • User
  – Legal liability
    • e.g., downstream liability
    • Stolen CC
  – Publicity
    • Most intrusions are not reported
  – Theft of IP
Forensic Duplication

• Your strategy is to take the system offline.
  – Case may go to court or high-cost damage
  – Need to perform a bit-level copy of the system
  – WHY?
  – Two types of analysis
    • Logical
    • Physical
Forensic Duplication

- Your strategy is to take the system offline.
  - Can’t do a physical analysis on a mere logical ‘copy’ of the hard drive
    - Misses ‘ambient’ data that may contain a wealth of evidence
    - Must access each sector of the HD
    - ‘Ambient’ data found in areas no privy to the user
Forensic Duplication

• Your strategy is to take the system offline.
  – Offline analysis allows you to preserve the system ‘as-is’, i.e., like putting yellow police tape around the scene of a crime
  – Offline analysis doesn’t affect the integrity of the evidence because you are doing analyses on copies of the evidence.
  – Of course you’ll likely loose all volatile data by shutting down the machine
Forensic Duplication

• More on this a little later…
Authenticate the Evidence

• It is difficult to show that evidence of any kind collected is the same as what was left behind by a criminal
  – Computer drives deteriorate slowly
  – Child pornography and Taliban terror plans don’t show up randomly on a HD
  – Chain of custody and other handling rules assure the jury that no unanticipated or introduced changes occurred.
  – “prove who was at the keyboard” problem

Kruse & Heiser, 2002
Investigation

• Answers
– Who, what, when, where, how…
– How you perform the investigation determined by whether you have a forensic duplicate, or whether you are conducting a live response.
  • IE..Can’t get certain portions of a hard disk if working with live-response
  • …Can’t do a string search on a swap file under live-response
Investigation

• What is the goal?
  – Search for appropriate types of information
    • Graphics/images
    • Text
  – Problems:
    • There are hundreds or thousands of files
    • “Needle in a stack of needles” problem
    • Files can be hidden
      – Kiddy porn graphic saved as “myhomework.doc”
      – Steganography or alternate data streams
      – Files deleted
      – .files
      – Hidden areas of disk
      – obfuscation
Common Mistakes

- Altering time and date stamps.
- Killing rogue processes.
- Patching the system before the investigation.
- Not recording commands executed on the system.
- Using untrusted commands and binaries.
- Writing over potential evidence by:
  - Installing software on the evidence media
  - Running programs that store their output on the evidence media.

Kruse & Heiser, 2002
How do you know something is wrong?

- Failed login attempts
- Logins into dormant and default accounts
- Activity during nonworking hours
- Presence of new accounts not created by the systems administrator
- Unfamiliar files or programs
Running Processes

What is this?

What’s ‘wrong’ with this?

Linux: top, ps
Other:

• Event Log
• Computer management (mmc)
• Open Shares (mmc)
• Network connections (netstat)
• Services (mmc)
• Connected users (mmc)

• MMC, administrative tools, and 3rd party applications are all going to be valuable
Detection

• Unexplained changes in file and directory permissions
• Unexplained elevation or use of privileges
• An altered web page
• Presence of pornographic images on a system
Detection

• Use of commands or functions not normally associated with a user’s job
• Presence of ‘contraband’ utilities – (cracking, hacking, crypto, obfuscating, etc)
• Gaps in or erasure of system logs
Detection

- Changes in DNS tables or router or firewall rules that cannot be accounted for.
- Unusually slow system performance
- System crashes
- Social engineering attempts
Where do I find this evidence?

• It depends on the OS
• For Windows
  – it will likely be in various GUI-based utilities
  – Or in highly obfuscated portions of specific files
• For UNIX/Linux, it will likely be in various text files
The Initial Assessment

• What probably happened?

• Best response?
  – Investigator must assess scene and respond accordingly
  – Difference between
    • Someone lying on the ground bleeding at scene of the crime
    • Someone lying on the ground dead at the scene of the crime
  – Response differ depending on circumstances

Mandia & Prosise, 2003
Incident Notification Checklist

• Who called:
  – Time/Date
  – Phone
• Nature of incident
• When did it occur?
• How was it detected?
• When was it detected?
• Immediate and future impact to client:

Mandia & Prosise, 2003
Always practice safe hex.
Why HEX?

• While hex is less readable than ascii text, it is more readable than code the machine understands…
  – The number 65535 would be written down as 16 ones, or $1111111111111111_2$
  – Prone to error…was that 16 or 17 1’s?
  – To condense the same information we use a base 16 system, called hexadecimal.
What is HEX?

- Hex uses decimals first, followed by alphabetic characters.
- It is fairly straightforward to convert back and forth from binary to hex.
<table>
<thead>
<tr>
<th>BIN %</th>
<th>OCT</th>
<th>DEC</th>
<th>HEX 0x</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>100</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>101</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>110</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>111</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>1000</td>
<td>10</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>1001</td>
<td>11</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>1010</td>
<td>12</td>
<td>10</td>
<td>A</td>
</tr>
<tr>
<td>1011</td>
<td>13</td>
<td>11</td>
<td>B</td>
</tr>
<tr>
<td>1100</td>
<td>14</td>
<td>12</td>
<td>C</td>
</tr>
<tr>
<td>1101</td>
<td>15</td>
<td>13</td>
<td>D</td>
</tr>
<tr>
<td>1110</td>
<td>16</td>
<td>14</td>
<td>E</td>
</tr>
<tr>
<td>1111</td>
<td>17</td>
<td>15</td>
<td>F</td>
</tr>
<tr>
<td>10000</td>
<td>20</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>10001</td>
<td>21</td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>
Converting

• If you write down 1234, (base 10) you are talking about the number one thousand, two hundred and thirty four.

• This can be rewritten as:

\[
\begin{align*}
1 & \times 10^3 = 1000 \\
2 & \times 10^2 = 200 \\
3 & \times 10^1 = 30 \\
4 & \times 10^0 = 4
\end{align*}
\]
Converting

- It is the same in all other bases, each place represents a power of the base:

  %1010 would be
  
  1 * 2^3
  0 * 2^2
  1 * 2^1
  0 * 2^0

  0x1234 will be
  
  1 * 16^3
  2 * 16^2
  3 * 16^1
  4 * 16^0
Converting

• What is 0xCB in Decimal?
  C = 12 and B = 11 so
  $12 \times 16^1 + 11 \times 16^0 = 203$

What about binary?

C = 12 = 1100  B = 11 = 1011
CB = 1100 . 1011
so 0xCB = %11001011
Converting

• What is 0xAF1 in Decimal?

• A = 10, F = 15 so
  \[10 \times 16^2 + 15 \times 16^1 + 1 \times 16^0 = 2801\]

What about binary?

A = 1010 \quad F = 1111 \quad 1 = 0001

AF1 = 1010 . 1111 . 0001

so 0xAF1 = %101011110001
Practical bits

• Netmask:
  – So most people just type in: 255.255.255.0
  – What does this mean?
Practical bits

- Netmask:
  IP addresses are ‘dotted quad’, basically the dots just break up bits to make them easier to read.

How many bits does it take to represent 256 (base 10)?
Practical bits

• Netmask:
  11111111 = 255, so 8 bits for 256 unique values

Therefore, 255.255.255.0 is ‘decimal dotted quad” for the base 2 number:

11111111.11111111.11111111.00000000

This is also sometimes referred to as as /24 network because there are 24 1’s

Netmasks *almost* always start with sequential 1’s and end with sequential 0’s
...slight diversion now...

- Netmask: 
  11111111 . 11111111 . 11111111 . 00000000
  network (subnet) host

So this particular netmask (/24) allows for 256 different hosts...(well actually a bit less – but lets just say 256) on one subnet. Every time you add a bit to the netmask, you get more subnets and less hosts per subnet.

Example: 
192.168.100.0 – 192.168.100.255
...slight diversion now...

- Netmask: $11111111.11111111.11111111.1 \ 0000000$
  
  network (subnet)  host

So this particular netmask (/25) has 2 subnets...

Example:

- 192.168.100.0 – 192.168.100.127  subnet1
- 192.168.100.0 – 192.168.100.255  subnet2

So /26 has 4 subnets, /27 has 8 subnets, all the way through /30 which has 64 subnets (4 hosts per)
...slight diversion now...

• Netmask:
Looking at Netmasks that ‘lower’ than /24 get into Class A,B,C type discussions and are definitely out of scope here…
Basically each fourth of the dotted quad controls a class, so using letters to represent the class a bit belongs to:
AAAAAAA.ABBBBBBBBBBBB.CCCCCCCCCC.xxxxxxxxx
Class D is used for broadcasting
Class E is “Experimental” is basically a leftover from bureaucratic / political “design by committee” fallout
Practical Bits

• Netmask:
  – What’s the mask actually do?
  – Used for Bitwise AND with a host’s address
  – If my computer is 137.48.112.123
     and my netmask is 255.255.255.0
     
    | 10001001 | 00110000 | 01110000 | 01111011 |
    | 11111111 | 11111111 | 11111111 | 00000000 |

    AND 10001001.00110000.01110000.00000000

    so for the very common /24 netmask the result
    may be familiar then the last number (123) is the
    host id, and the others 137.48.112 is the network.
Why does all this matter?

• So as a forensic examiner you might not be overly concerned with netmasks, or the class of a particular network
• And you may not be able to decode machine language when you see it
• But you should understand what it is and realize that decoding it correctly could change data into information…
Why does all this matter?

• In the physical world if an investigator found a letter at a crime scene he would not throw it away just because the crime was committed in Nebraska and the letter was written in Chinese.
Why does all this matter?

• A set of 1’s and 0’s that translates into an peculiar set of Hex characters may appear to be gibberish, but upon proper decoding, it may reveal an MIME encoded message (for example)

• Just because the data isn’t in a particularly useful form, doesn’t mean that it’s not valuable.
Peek into the Future:

• Windows stores all kinds of data in all kinds of places…

• And interesting example are .lnk files

• And extension of .lnk means?
Peek into the Future:

• Turns out that the date / time information for the original file the Ink points to (deleted or not) is stored in the Ink.

• Starting at byte offsets 28, 36, and 44 you can gleam creation, last access and last modification times..

• These are Windows “Date/Time” values – 64 bit little Endian
Peek into the Future:

• CST is 6 hours behind GMT…
• Notice the highlighted portion of hex in winhex
• What happened on 11/16/04 at about 9:54 AM?
Peek into the Future:

• What are the odds?
• Every time a document is accessed a lnk is created in the hidden system folder RECENT
• This folder exists for all users individually
  – Obviously this knowledge has a variety of uses
Encoding is not Encrypting

- It is also important to note the different between Encoding and Encrypting
- Encoding is done primarily to make information EASY to interpret
- Encrypting is done primarily to make information HARD to interpret
Encoding is not Encrypting

• The very fact that data has been encrypted is sometimes enough to raise ‘red flags’

• Depending on circumstances the existence of encrypted files may create, or be a contributing factor for Probable Cause

• This is not the case with encoded files
The Hex Editor

- In windows you may find a tool such as winhex, frHed, or Hackman valuable:
- In Linux maybe something like xxd, Heme, SHED, gHex, KHexEdit or some other abstraction (Autopsy for example has a hex view option).
Hex Editor

You can use these hex tools at varying granularity…by file:

Viewing a FILE

NOTICE the offset starts at 0
Hex Editor

• How is this different?

Viewing a DISK

Contents do not start at 0

Removable Medium 2

| Offset | 0   | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | A   | B   | C   | D   | E   | F   | Access |
|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------|
| 0011B1A0 | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  |       |
| 0011B1B0 | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  |       |
| 0011B1C0 | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  |       |
| 0011B1D0 | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  |       |
| 0011B1E0 | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  |       |
| 0011B1F0 | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  |       |
| 0011B200 | 48  | 61  | 62  | 69  | 62  | 2C  | 0D  | 0A  | 4F  | 75  | 72  | 20  | 70  | 6C  | 61  | 6E  |       |
| 0011B210 | 73  | 20  | 61  | 72  | 65  | 20  | 69  | 6E  | 20  | 6D  | 6F  | 74  | 69  | 6F  | 6E  | 20  |       |
| 0011B220 | 61  | 6E  | 64  | 20  | 61  | 6C  | 6C  | 20  | 69  | 73  | 20  | 77  | 65  | 6C  | 6C  | 2E  |       |
| 0011B230 | 20  | 20  | 48  | 6F  | 6D  | 65  | 6C  | 61  | 6E  | 64  | 20  | 53  | 65  | 63  | 75  | 72  |       |
| 0011B240 | 69  | 74  | 79  | 20  | 73  | 75  | 73  | 70  | 65  | 63  | 75  | 73  | 70  | 65  | 63  | 75  |       |
| 0011B250 | 68  | 69  | 6E  | 67  | 2C  | 20  | 74  | 68  | 65  | 20  | 65  | 78  | 70  | 6C  | 6F  | 73  |       |
| 0011B260 | 69  | 6F  | 6E  | 20  | 77  | 69  | 6C  | 6C  | 20  | 62  | 65  | 20  | 67  | 72  | 61  | 6E  |       |
| 0011B270 | 64  | 2E  | 20  | 20  | 41  | 74  | 74  | 61  | 63  | 68  | 65  | 64  | 20  | 61  | 72  | 65  |       |
| 0011B280 | 20  | 74  | 68  | 65  | 20  | 63  | 6F  | 6F  | 72  | 64  | 69  | 6E  | 61  | 74  | 65  | 73  |       |
| 0011B290 | 20  | 6F  | 66  | 20  | 74  | 68  | 65  | 20  | 61  | 74  | 74  | 61  | 63  | 6B  | 20  | 73  |       |
| 0011B2A0 | 61  | 76  | 65  | 64  | 20  | 69  | 6E  | 20  | 74  | 68  | 65  | 20  | 75  | 73  | 75  | 61  |       |
| 0011B2B0 | 6C  | 20  | 77  | 61  | 79  | 2E  | 0D  | 0A  | 4C  | 6F  | 79  | 61  | 6C  | 6C  | 79  | 2C  |       |
| 0011B2C0 | 0D  | 0A  | 53  | 61  | 6D  | 69  | 72  | 0D  | 0A  | 00  | 00  | 00  | 00  | 00  | 00  | 00  |       |
| 0011B2D0 | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  |       |
| 0011B2E0 | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  | 00  |       |

Habib... Our plans are in motion and all is well. Homeland Security suspects nothing, the explosion will be grand. Attached are the coordinates...
Files

- Many “low level” things can be determined at the Hex level
- Files always have particular header information (this is different than file-extensions like .doc or .jpeg)
- This is often called a file signature or Magic numbers
When considering “graphics files”

; Windows Bitmap graphics BMP=0x00:"BM" ; Compressed BM? File BM_=0x00:"SZDD"
; Graphics Interchange Format bitmap graphics GIF=0x00:"GIF8"
; Graphics Interchange Format bitmap graphics (GIF 87a) GIF87A=0x00:"GIF87a"
; Graphics Interchange Format bitmap graphics (GIF 89a) GIF89A=0x00:"GIF89a"
; JPEG Bitmap graphics JPE=0x00:0xFF,0xD8,0xFF,0xE0,0x00,0x10,"JFIF"
; JPEG Bitmap graphics JPG=0x00:0xFF,0xD8,0xFF,0xE0,0x00,0x10,"JFIF"
JS=0x00:"/

These are standard types, the information is widely available, these particular lines came from drivespy.ini
Files

This is the hex representation of a jpg:
Files

• If files are simply stored in ‘hidden’ areas, like unallocated, slack, or interpartition space, they will still have header information.

• If files are enciphered some way (like stereography) then there is no header information.

• If files are encrypted / compressed, there may not be header information about the file, but there will typically be header information about the encryption / compress for decryption / decompression purposes.
Files

• In some cases you may find portions or fragments of a file. If you suspect that the fragment may be part of what used to be JPEG for example (because near where the header should be you found “FIF” and you know that jpeg headers contain “JFIF”) you can attempt to recover the file by editing the correct header information back to the disk.
Hashing

No. Not that kind.
Hashing

• One of the best ways to describe hashing is to describe a hash as a fingerprint [of an image].
• Fingerprints uniquely identify a much larger object (human) from a much smaller object (the fingerprint)
Hashing

• ...similarly, a digital hash is a unique representation of a larger object – like an image

• This hash is a file that is completely separate from the image that it is fingerprinting and has a fixed length – like 128 or 160 bits.

• A 1 MB file and a 1 GB files will both produce hashes of the same length
Hashing

• The general idea is that a very small (any) change in the source file will result in a very large change in the hash

• The hashes we are referring to are “one-way” hashes
Hashing

• There are many automated tools that provide hashing components.
• Most *nix distributions provide hashing tools by default, for windows you’ll have to download software
The algorithm is independent of OS – the same hash is produced from the same file on Linux and Windows.
Hashing

• The same software typically provides the means to check to see if the hash of a given file has changed ... in this case a \(-c\).

```bash
[root@localhost root]# md5sum SavedEmail.txt > SavedEmail.md5
[root@localhost root]# md5sum -c SavedEmail.md5
SavedEmail.txt: OK
[root@localhost root]# sha1sum SavedEmail.txt > SavedEmail.sha1
[root@localhost root]# sha1sum -c SavedEmail.sha1
SavedEmail.txt: OK
[root@localhost root]#  
[root@localhost root]# md5sum -c SavedEmail.md5
SavedEmail.txt: FAILED
md5sum: WARNING: 1 of 1 computed checksum did NOT match
[root@localhost root]# sha1sum -c SavedEmail.sha1
SavedEmail.txt: FAILED
sha1sum: WARNING: 1 of 1 computed checksum did NOT match
[root@localhost root]#  
```
Hashing

- MD5 – 128 bits
- Sha1 – 160 bits
- Sha256 – 256 bits
- …sha384, sha512
Hashing

Use something like md5deep or sha1deep for recursion:

```
C:\tools\md5deep>md5deep -r * > c:\file.dat
C:\tools\md5deep>md5deep -rX c:\file.dat *
C:\tools\md5deep>md5deep -rX c:\file.dat *
b31540d38eb675b77c2b417b374bada5  C:\tools\md5deep\README.txt
C:\tools\md5deep>
```
Side Rant: Hashing DLs

• When downloading software a hash is often provided along with the download.
  – What purpose does this hash serve?
MD5 hash collisions

• What’s all this hoopla about?
• Who has heard of this?
  – explain
Hash Collisions

• Hash Collision (n): a term in computer programming for a situation that occurs when two distinct inputs into a hash function produce identical outputs.

• What does this mean to us forensically?

Hash Collisions

• It’s all computation time relative
  – 1) Create bad file
  – 2) Gen MD5
  – 3) Was is the MD5 I wanted? (no)
  – 4) Mod bad file in some way
  – 5) go to 2 (until done)

• Turns out that if you can produce, locate, etc two strings of the same arbitrary length that happen to hash to the same MD5, then you can do some ‘interesting’ things
Hash Collisions

• With getting too into it…
  – Message Digest 5 uses Merkle-Damgard construction rounds
  – Starting at 128 bits then ‘adding’ (processing?) in 512 more bits at a time
  – Unfortunately, at time+X for two arbitrary files going through rounds if at any given round in either file the ‘current’ hash matches, then arbitrary data can be appended afterward….and the resultant hashes will match
Hash collisions

• Tools like stripwire can actually create 2 files that have the same md5...and very quickly

stripwire $VERSION: Conflation Attack Using Colliding MD5 Test Vectors

Author: Dan Kaminsky(dan@doxpara.com)
Example: ./stripwire.pl -v -b test.pl -r fire.bin
Options: -b [file.pl] : Build encrypted archives of this perl code
-r [file.bin] : Attempt to self-decrypt and execute this file
-v : Increase verbosity.
-a : Rename active payload (fire.bin)
-i : Rename inactive payload (ice.bin)
Hash Collisions

• What does this mean to us?
• Actually very little!
  – Since we are creating two new files with the same MD5 this doesn’t even effect Known Hash Set Lists, like KFF or NIST / NSRL
  – This whole discussion was on MD5, but does/may apply to other hashing algorithms
    • This can be mitigated by simply storing ‘dual hashes’ or in a weaker sense by storing other metadata like filesize
Difference between the 2
Difference between the 2

- d1 31 dd 02 c5 e6 ee c4 69 3d 9a 06 98 af f9 5c
- 2f ca b5 87 12 46 7e ab 40 04 58 3e b8 fb 7f 89
- 55 ad 34 06 09 f4 b3 02 83 e4 88 83 25 71 41 5a
- 08 51 25 e8 f7 cd c9 9f d9 1d bd f2 80 37 3c 5b
- d8 82 3e 31 56 34 8f 5b ae 6d ac d4 36 c9 19 c6
- dd 53 e2 b4 87 da 03 fd 02 39 63 06 d2 48 cd a0
- e9 9f 33 42 0f 57 7e e8 ce 54 b6 70 80 a8 0d 1e
- c6 98 21 bc b6 a8 83 93 96 f9 65 2b 6f f7 2a 70

- d1 31 dd 02 c5 e6 ee c4 69 3d 9a 06 98 af f9 5c
- 2f ca b5 07 12 46 7e ab 40 04 58 3e b8 fb 7f 89
- 55 ad 34 06 09 f4 b3 02 83 e4 88 83 25 f1 41 5a
- 08 51 25 e8 f7 cd c9 9f d9 1d bd 72 80 37 3c 5b
- d8 82 3e 31 56 34 8f 5b ae 6d ac d4 36 c9 19 c6
- dd 53 e2 b4 87 da 03 fd 02 39 63 06 d2 48 cd a0
- e9 9f 33 42 0f 57 7e e8 ce 54 b6 70 80 28 0d 1e
- c6 98 21 bc b6 a8 83 93 96 f9 65 ab 6f f7 2a 70
Bit Rot

• Has anyone ever heard of “Bit Rot”? 
Bit Rot

- There is actually a lot of DA backlash about hashing as part of Chain of Custody
- Over time, MTBF kicks in and a bit mysteriously flips on an HD.
- This one bit obliterates the hash
  - What are the legal repercussions to a re-opened case?
Resources

- http://md5deep.sourceforge.net/
- www.doxpara.org
- En.wikipedia.org
References


References

- Center for Strategic & International Studies (CSIS) [http://www.csis.org/pubs/cyberfor.html](http://www.csis.org/pubs/cyberfor.html)
- [http://www.asclld-lab.org/](http://www.asclld-lab.org/)