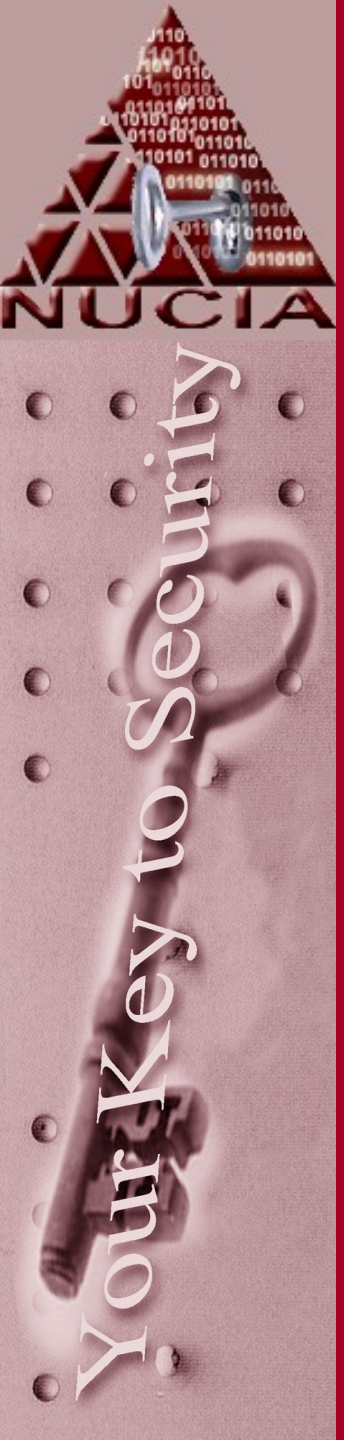


Cyber-Forensics The Basics

CERTConf2006

Tim Vidas

Because you have to start somewhere

Who are we?

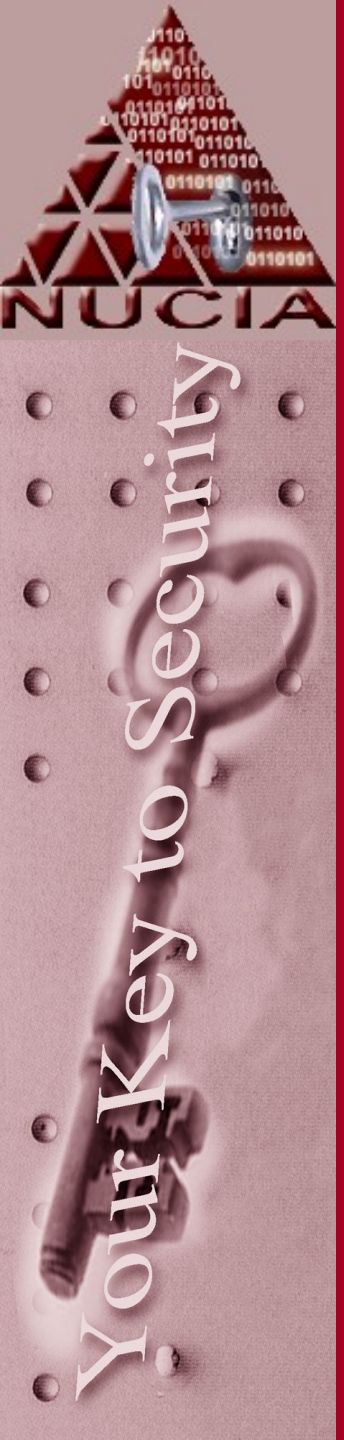
- Tim Vidas
  - Sr. Tech. Research Fellow
  - UNO/PKI/NUCIA
  - Certs: CISSP, 40xx, Guidance, AccessData etc.
  - Instructor: UNO, Guidance, LM RRCF
- Joe Wilson
  - Recent Graduate (MS/MIS)
  - RRCF

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.

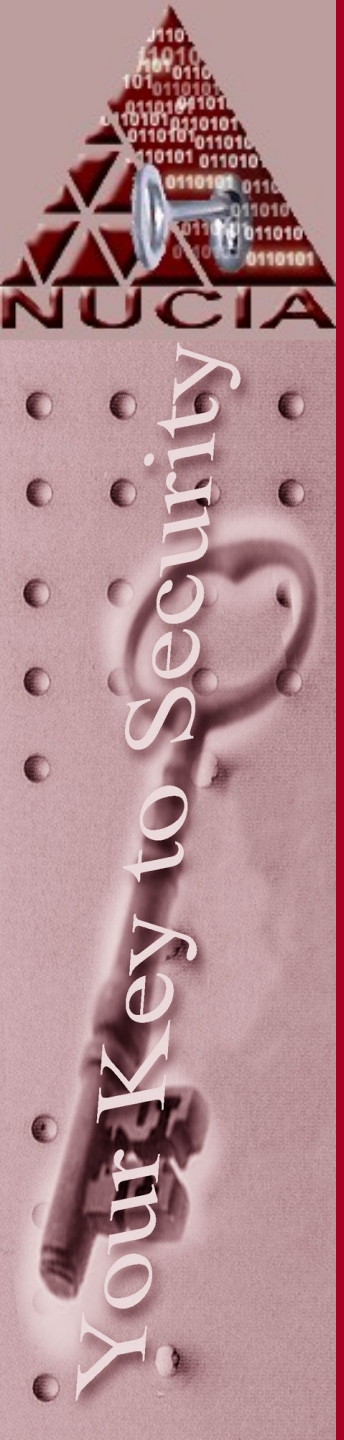
Who are you?

- 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.

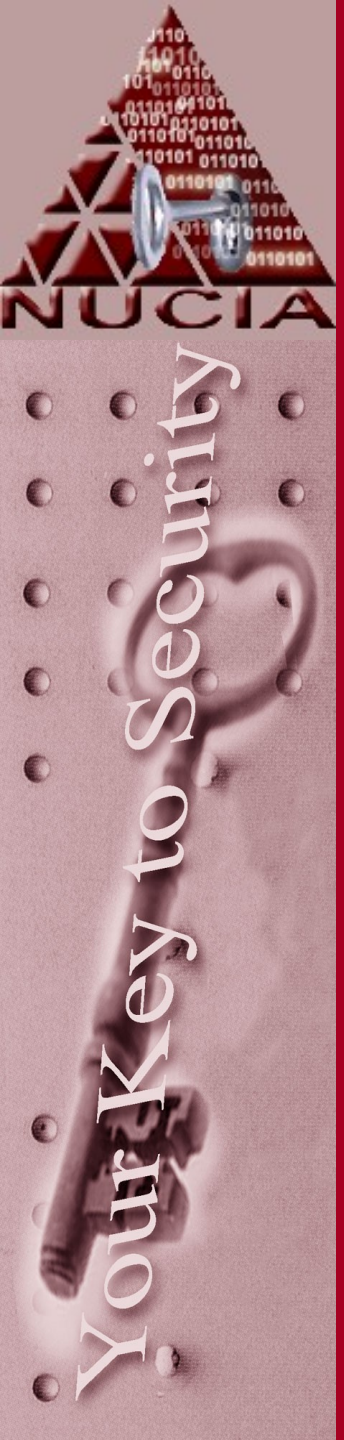
_____ forensics

- 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



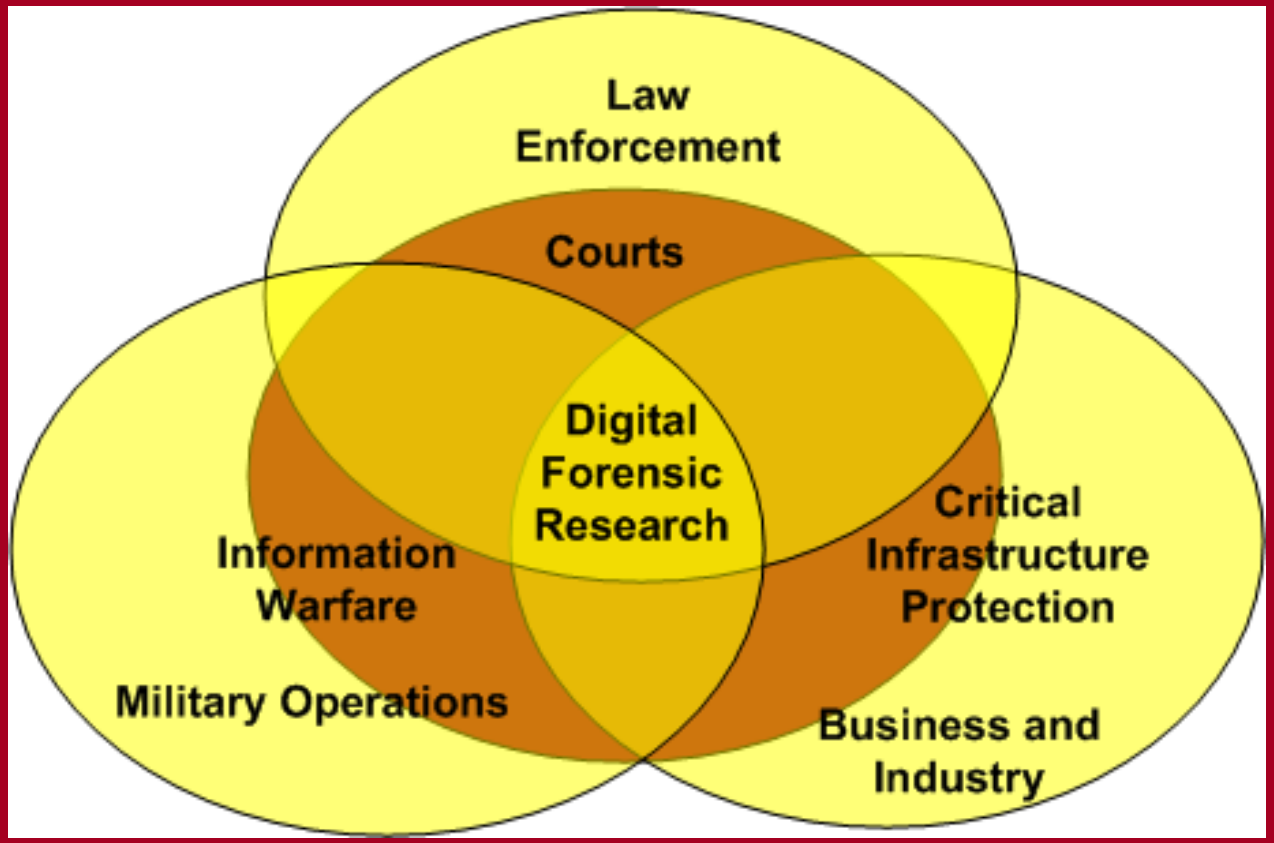
Rapid-Response Cyberforensics

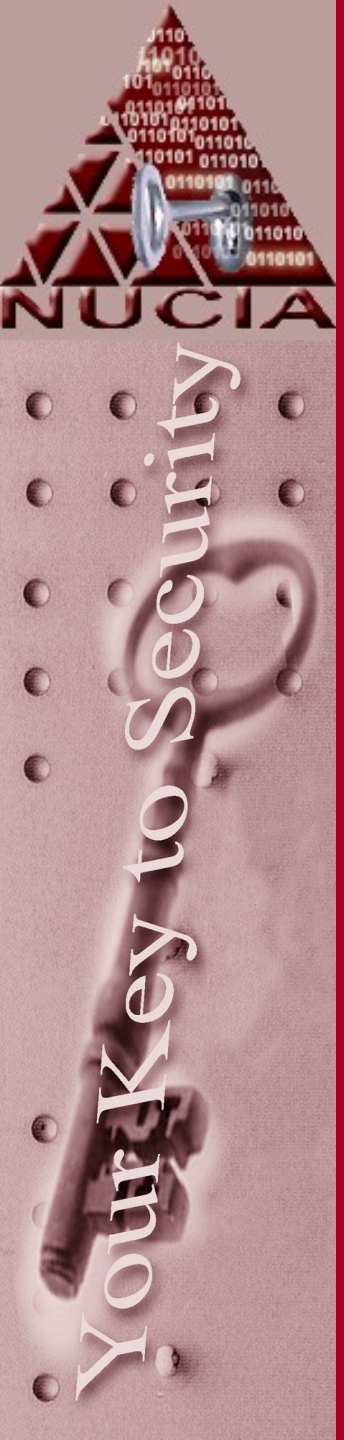
- Characterized by:
  - Live-response
  - Military-type contexts
    - But not of necessity
  - Judicious *a priori* planning
    - Prior strategic incident response planning
    - Requisite training in
      - Basic forensic procedures
      - Live-response
      - Network forensics
    - Continued updating of skills as technology changes
    - Technically adept with a diversity of tools & toolkits

Viewpoint

- 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


Law Enforcement
Courts
Digital Forensic Research
Information Warfare
Military Operations
Critical Infrastructure Protection
Business and Industry

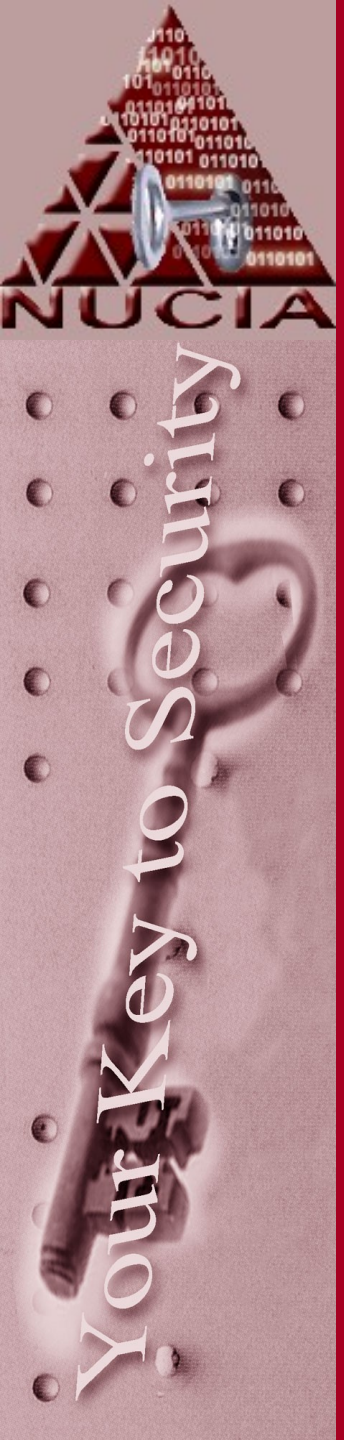


Viewpoint

- 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

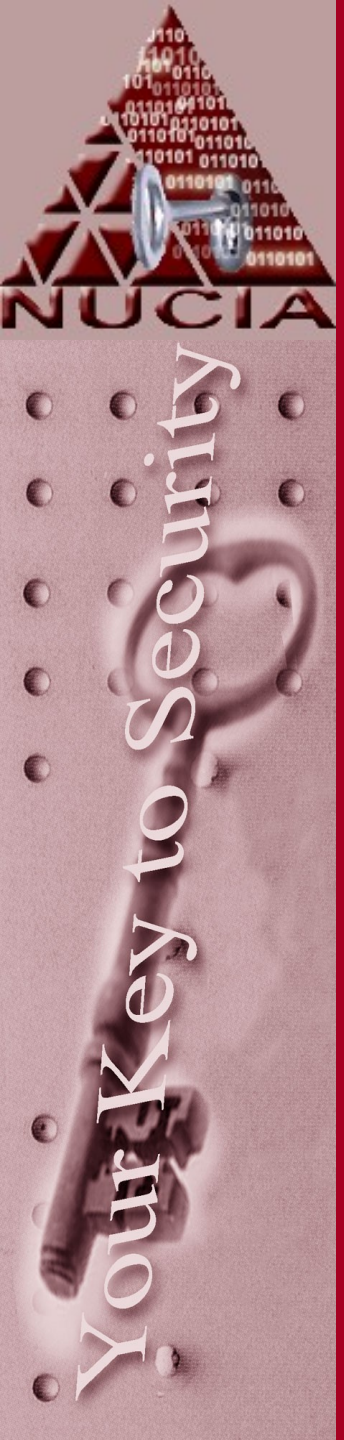
Coverage from OS perspective

- 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

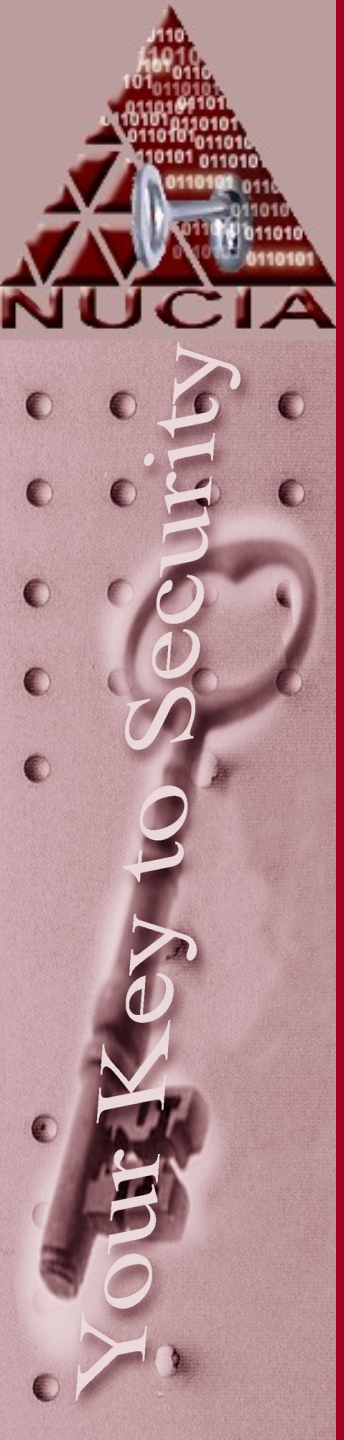
- "Information warfare specialists at the Pentagon estimate that a properly prepared and well-coordinated attack by fewer than 30 computer virtuosos strategically located around the world, with a budget of less than $10 million, could bring the United States to its knees."

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."

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



Scope of the Problem

- In1990 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

Stephenson, 2001.

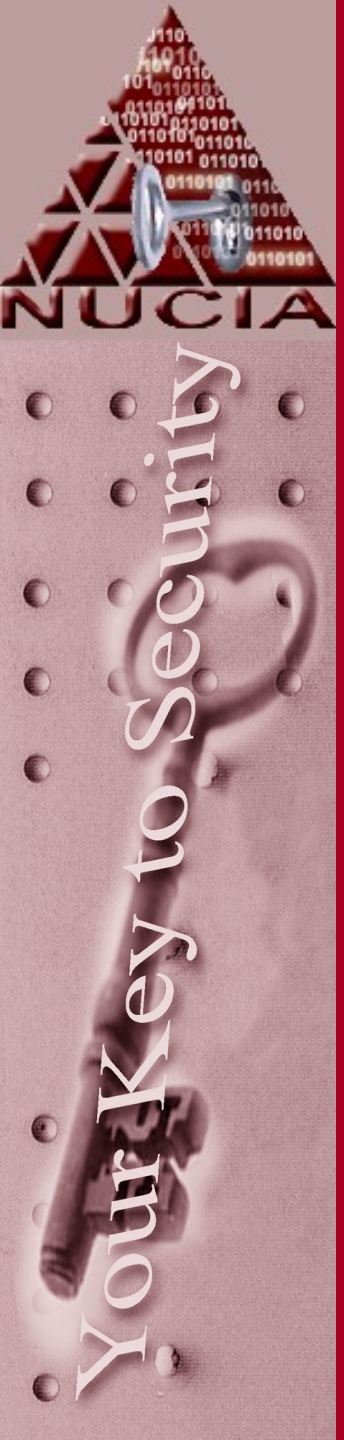
Computer Crimes

- 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

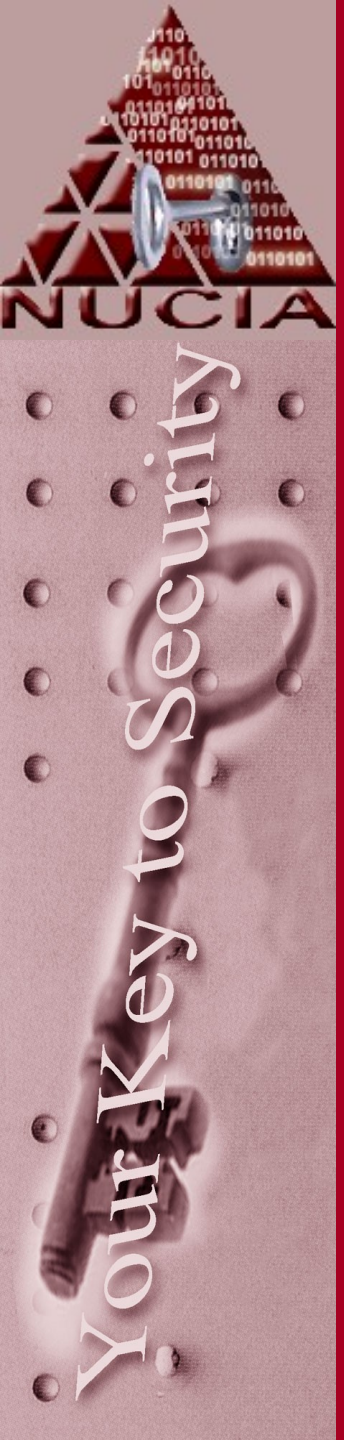
Taxonomy of Computer Crime Scenes

- 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

Incident?

Create Response Strategy

Secure System

Begin Evidence Acquisition

Duplication Required?

Duplicate

Begin Investigation

Report

Feedback to Preparation & Secure System

Is this Correct?

Adapted from Mandia & Prosise 2003

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

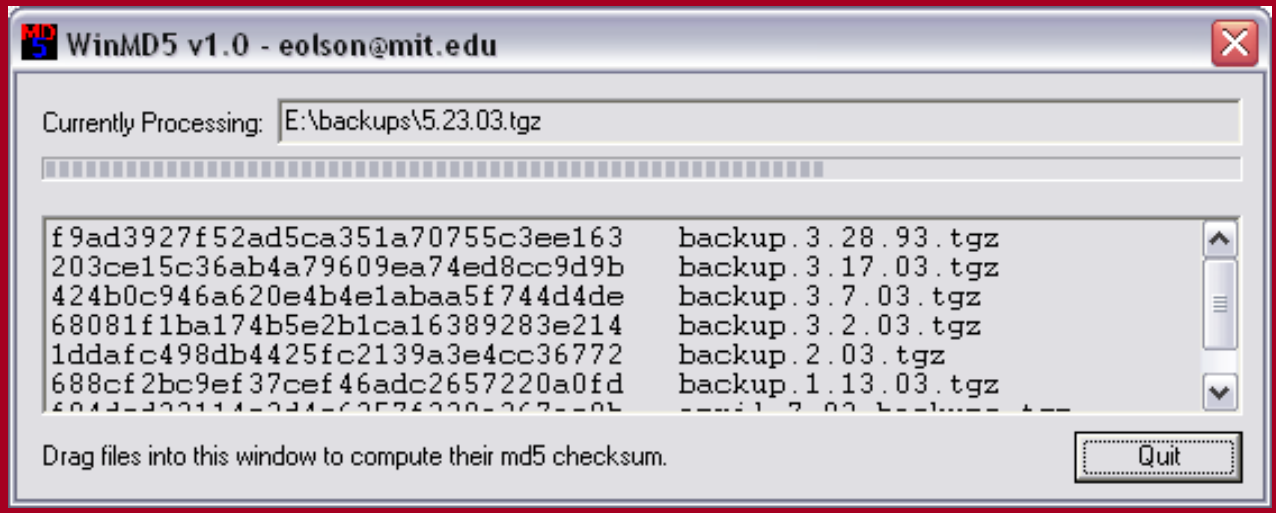
Chain-of-custody

- 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

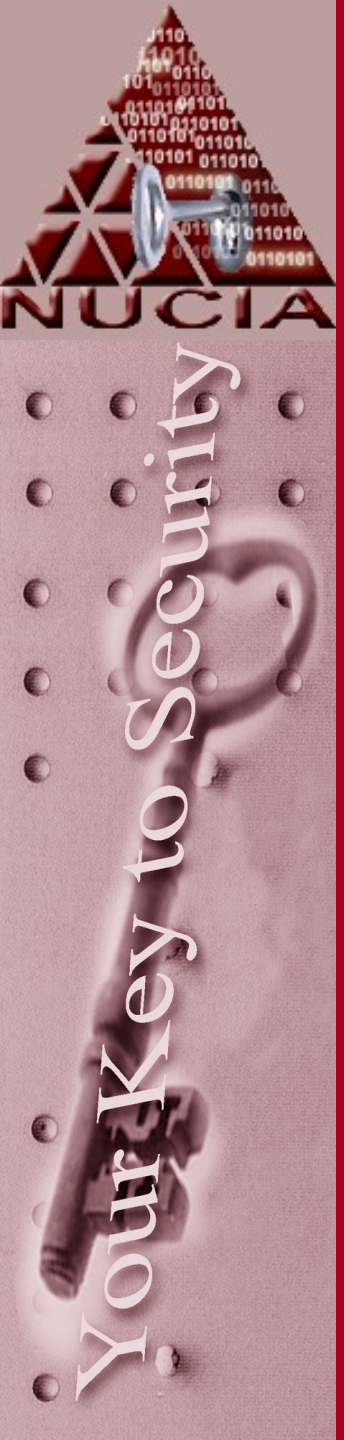
Item	Date	Time	Location	Name	Reason
Dell Inspiron 8000 SN# 4005	10/8/2002	13:05	Locked up in STEAL Lab cabinet	Vidas	Safekeeping
Dell Inspiron 8000 SN# 4005	10/9/2002	8:02	Removed	Vidas	Analysis
Dell Inspiron 8000 SN# 4005	10/9/2002	15:33	Locked up in STEAL Lab cabinet	Vidas	Safekeeping
Dell Inspiron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

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?

Kruse & Heiser, 2002



Chain-of-Custody

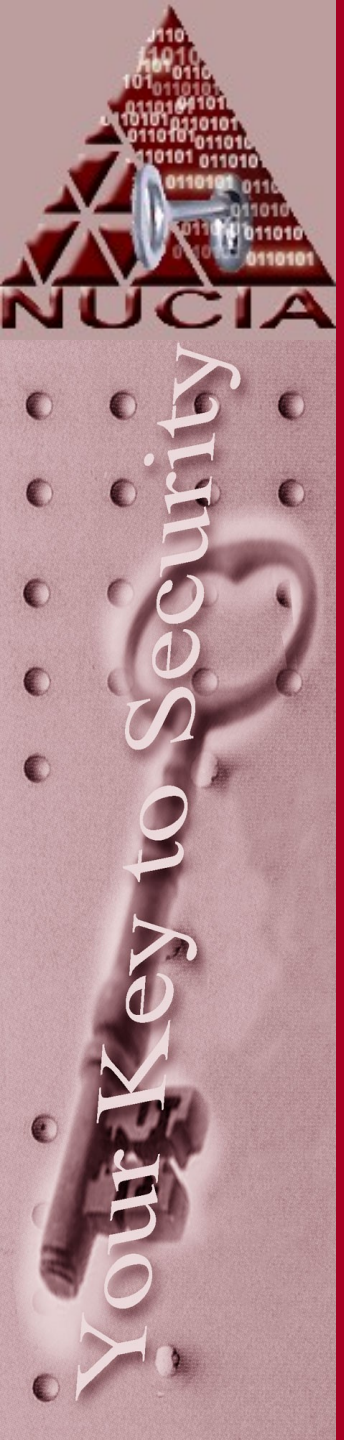
- Be meticulous
  - Defense attorney will cross-reference with other documents to determine any inconsistencies
- Fewer people who have access to your evidence or locker room, the better.
  - Defense attorneys will argue otherwise…

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

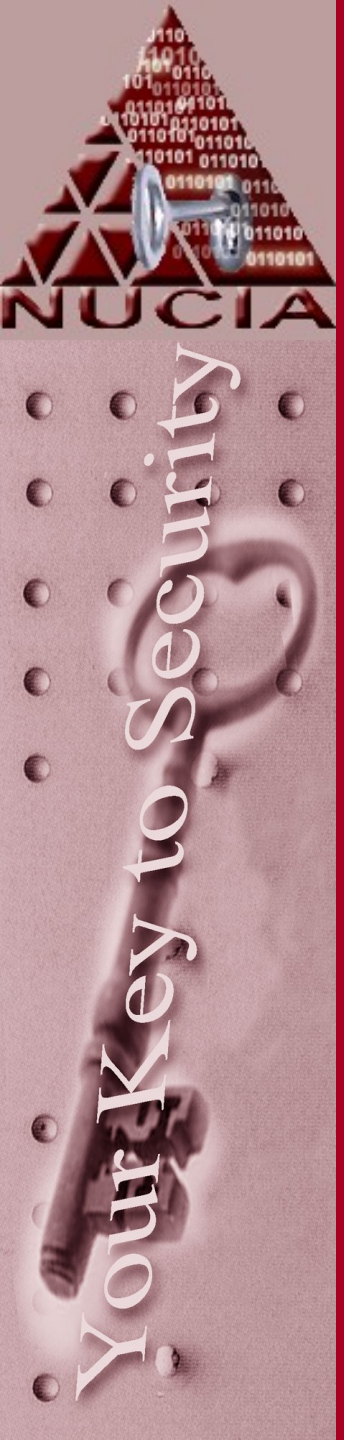
- Incident:
  - Unauthorized use
- Example
  - KPorn surfing from company workstation
- Strategy
  - Perform forensic duplication
  - Offline analysis
  - Interview user
- Likely outcome
  - Suspect identified and evidence collected for disciplinary action.

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.

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…

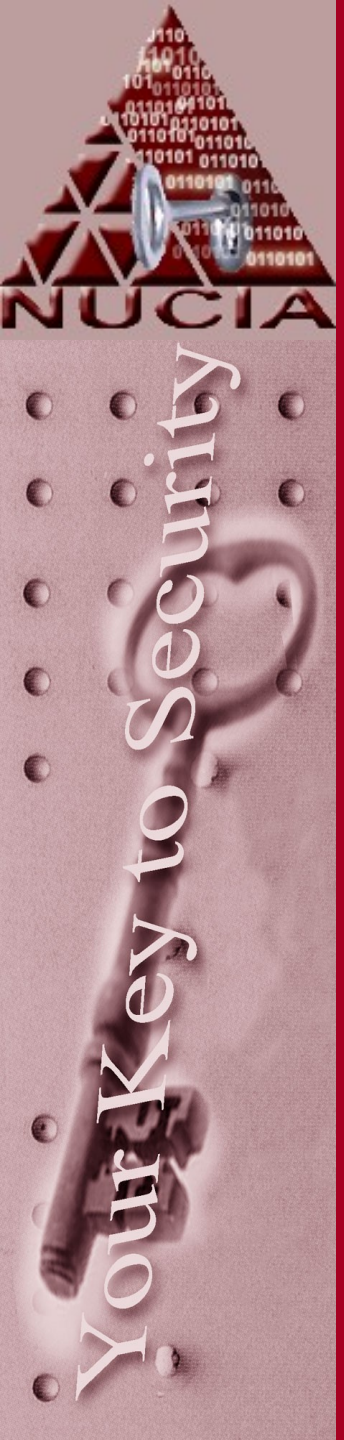
Authenticate the Evidence

- 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?

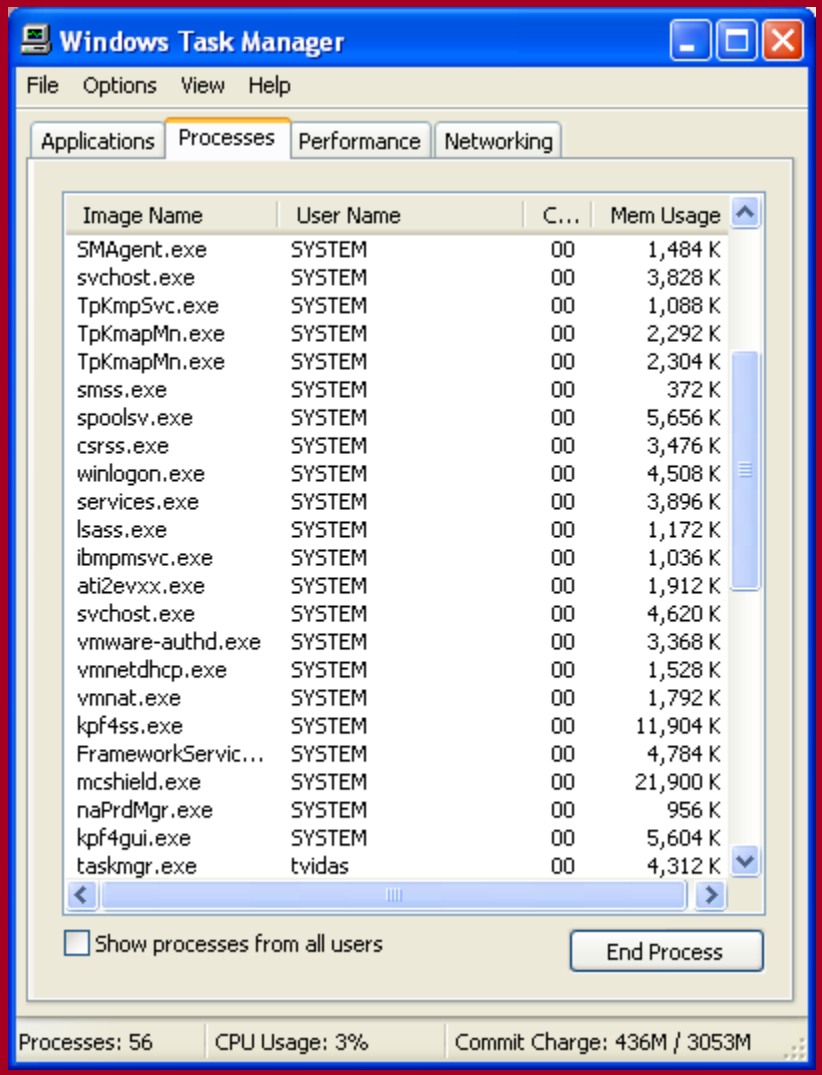
- Failed login attempts
- Logins into dormant and default accounts
- Activity during nonworking hours
- Presence of new accounts not created by the systems administrator
- Unfamiliar files or programs

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

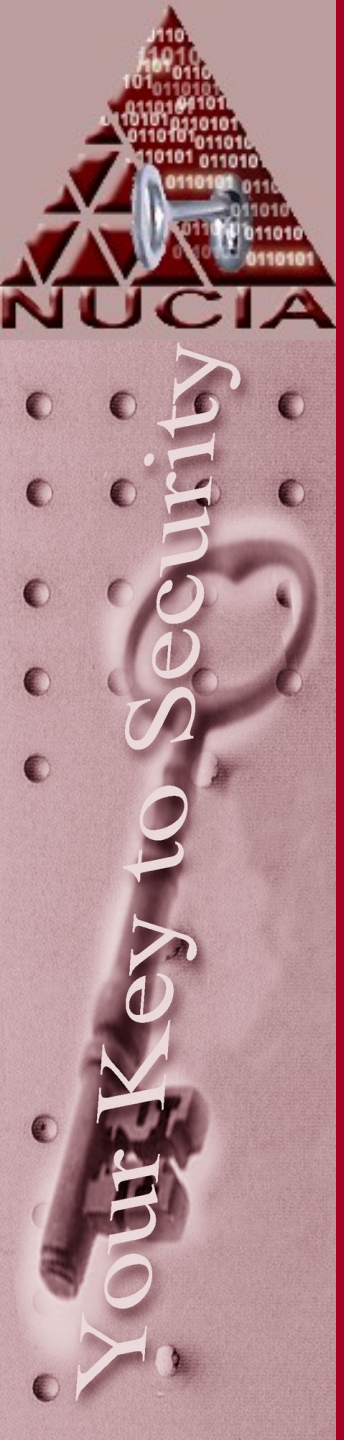
- Changes in DNS tables or router or firewall rules that cannot be accounted for.
- Unusually slow system performance
- System crashes
- Social engineering attempts

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

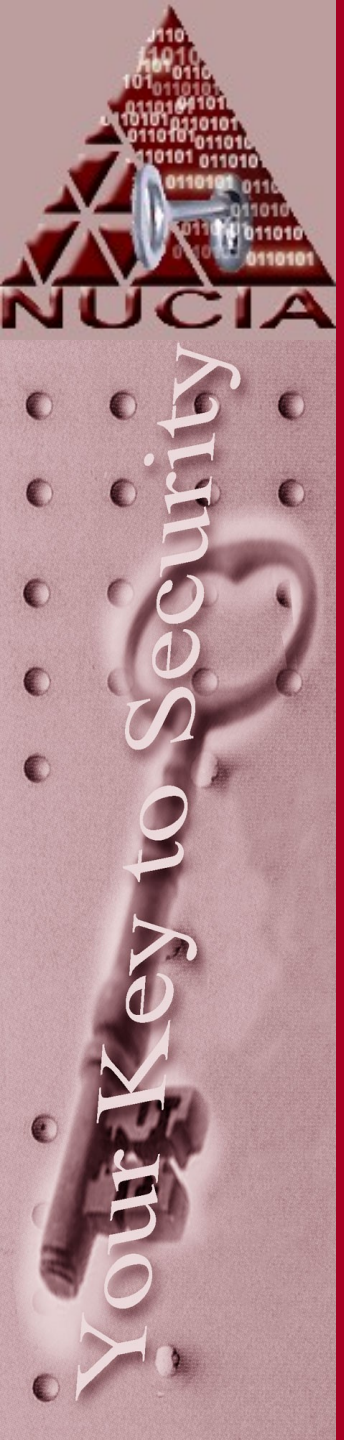
The Initial Assessment

- 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



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:



Hex

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

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F

BIN %	OCT	DEC	HEX 0x
1	1	1	1
10	2	2	2
11	3	3	3
100	4	4	4
101	5	5	5
110	6	6	6
111	7	7	7
1000	10	8	8
1001	11	9	9
1010	12	10	A
1011	13	11	B
1100	14	12	C
1101	15	13	D
1110	16	14	E
1111	17	15	F
10000	20	10	10
10001	21	11	11

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:

```
1 * 1000 = 1000
2 *  100 =  200
3 *   10 =   30
4 *    1 =    4
```

```
1 * 10^3
2 * 10^2
3 * 10^1
4 * 10^0
```

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 * 16^1 + 11 * 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 * 16^2 + 15 * 16^1 + 1 * 16^0 = 2801

What about binary?

A = 1010 F = 1111 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 address 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:

AAAAAAAA.BBBBBBBB.CCCCCCCC.xxxxxxxx

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 lnk points to (deleted or not) is stored in the lnk.
- 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?

Shortcut to WinHex.exe.lnk

Offset	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	10	11	12	13	14	15	16	17	18	19	1A	1B	1C	1D	1E	1F
00000000	4C	00	00	00	01	14	02	00	00	00	00	00	C0	00	00	00	00	00	00	46	9B	00	00	00	20	00	00	00	C0	09	99	88
00000020	F4	CB	C4	01	50	33	9A	CC	18	CC	C4	01	00	30	6C	CC	AC	BA	C4	01	00	82	0E	00	00	00	00	00	01	00	00	00
00000040	00	00	00	00	00	00	00	00	00	00	00	00	DB	00	14	00	1F	50	E0	4F	D0	20	EA	3A	69	10	A2	D8	08	00	2B	30
00000060	30	9D	19	00	2F	43	3A	5C	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	36	00	31	00	00
00000080	00	00	00	70	31	D0	7E	10	00	74	6F	6F	6C	73	0																	
000000A0	00	00	00	74	00	6F	00	6F	00	6C	00	73	00	00	0																	
000000C0	65	78	00	1E	00	03	00	04	00	EF	BE	70	31	C9	7																	
000000E0	00	46	00	32	00	00	82	0E	00	59	31	00	81	20	0																	
00000100	00	EF	BE	70	31	CA	7E	70	31	CB	7E	14	00	00	0																	
00000120	00	65	00	00	00	1A	00	00	00	46	00	00	00	1C	0																	
00000140	00	45	00	00	00	11	00	00	00	03	00	00	00	CC	0																	
00000160	65	78	5C	57	69	6E	48	65	78	2E	65	78	65	00	0																	
00000180	00	5C	00	74	00	6F	00	6F	00	6C	00	73	00	5C	0																	
000001A0	00	78	00	2F	00	65	00	78	00	65	00	0C	00	43	0																	

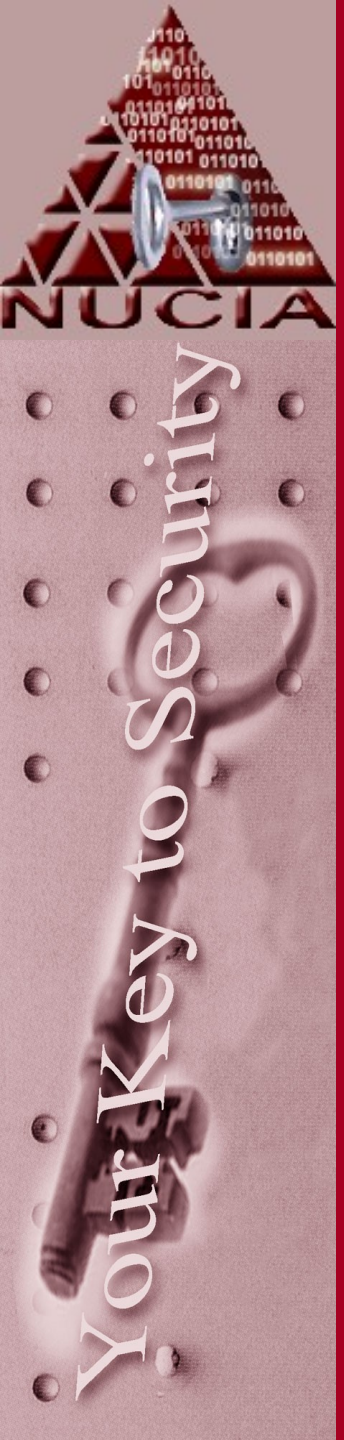
DCode Date v2.07.000 written by Craig Wilson

Time Zone: UTC -06:00 — Window on top

Decode Format: Windows: 64 bit Hex Value - Little Endian

Value to Decode: C0099988F4CBC401

Date & Time: Tue, 16 November 2004 09:54:19 -0600



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

```
$ xxd SavedEmail.txt
0000000: 4861 6269 622c 0d0a 4f75 7220 706c 616e  Habib,..Our plan
0000010: 7320 6172 6520 696e 206d 6f74 696f 6e20  s are in motion
0000020: 616e 6420 616c 6c20 6973 2077 656c 6c2e  and all is well.
0000030: 2020 486f 6d65 6c61 6e64 2053 6563 7572    Homeland Secur
0000040: 6974 7920 7375 7370 6563 7473 206e 6f74  ity suspects not
0000050: 6869 6e67 2c20 7468 6520 6578 706c 6f73  hing, the explos
0000060: 696f 6e20 7769 6c6c 2062 6520 6772 616e  ion will be gran
0000070: 642e 2020 4174 7461 6368 6564 2061 7265  d.  Attached are
0000080: 2074 6865 2063 6f6f 7264 696e 6174 6573   the coordinates
0000090: 206f 6620 7468 6520 6174 7461 636b 2073   of the attack s
00000a0: 6176 6564 2069 6e20 7468 6520 7573 7561  aved in the usua
00000b0: 6c20 7761 792e 0d0a 4c6f 7961 6c6c 792c  l way...Loyally,
00000c0: 0d0a 5361 6d69 720d 0a                   ..Samir..
```

SavedEmail.txt

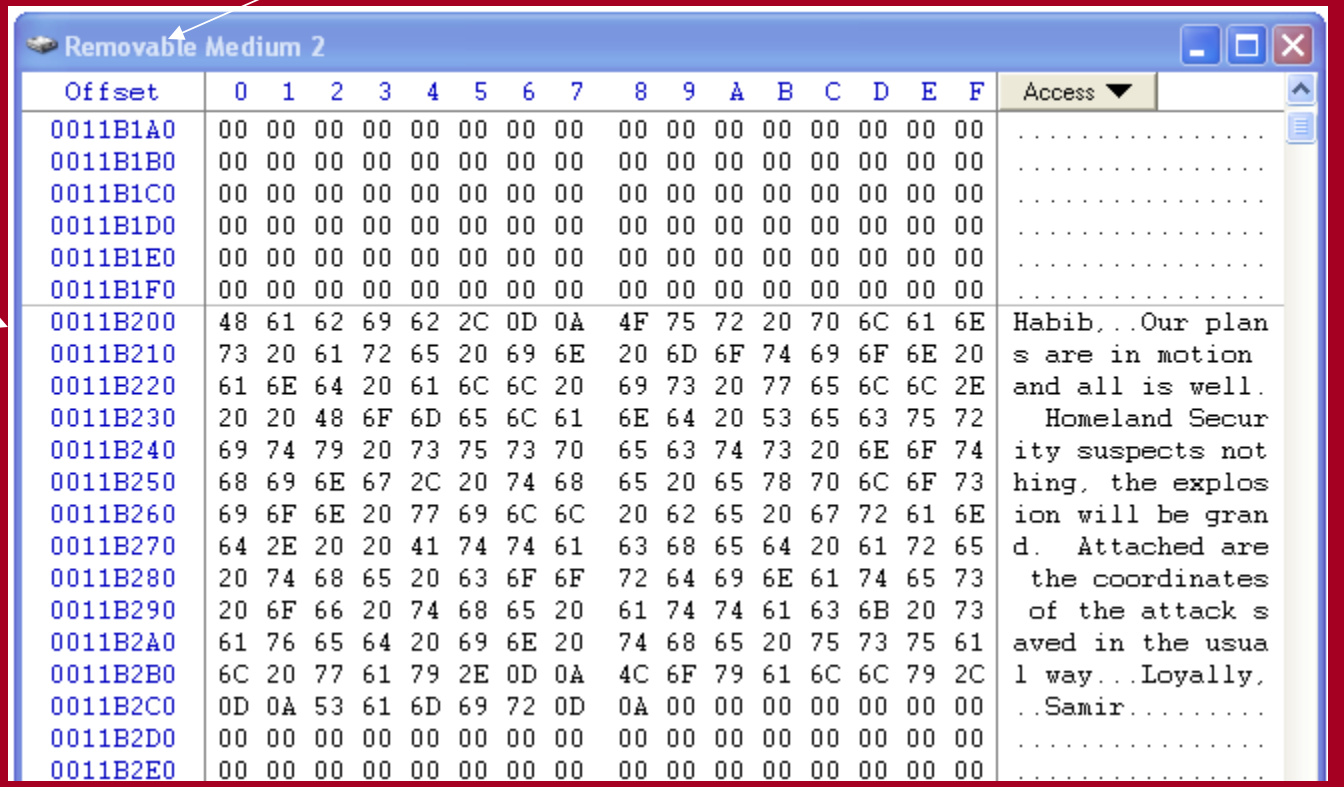
Offset	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	
00000000	48	61	62	69	62	2C	0D	0A	4F	75	72	20	70	6C	61	6E	Habib,..Our plan
00000010	73	20	61	72	65	20	69	6E	20	6D	6F	74	69	6F	6E	20	s are in motion
00000020	61	6E	64	20	61	6C	6C	20	69	73	20	77	65	6C	6C	2E	and all is well.
00000030	20	20	48	6F	6D	65	6C	61	6E	64	20	53	65	63	75	72	Homeland Secur
00000040	69	74	79	20	73	75	73	70	65	63	74	73	20	6E	6F	74	ity suspects not
00000050	68	69	6E	67	2C	20	74	68	65	20	65	78	70	6C	6F	73	hing, the explos
00000060	69	6F	6E	20	77	69	6C	6C	20	62	65	20	67	72	61	6E	ion will be gran
00000070	64	2E	20	20	41	74	74	61	63	68	65	64	20	61	72	65	d. Attached are
00000080	20	74	68	65	20	63	6F	6F	72	64	69	6E	61	74	65	73	the coordinates
00000090	20	6F	66	20	74	68	65	20	61	74	74	61	63	6B	20	73	of the attack s
000000A0	61	76	65	64	20	69	6E	20	74	68	65	20	75	73	75	61	aved in the usua
000000B0	6C	20	77	61	79	2E	0D	0A	4C	6F	79	61	6C	6C	79	2C	l way...Loyally,
000000C0	0D	0A	53	61	6D	69	72	0D	0A								..Samir..

Hex Editor

- How is this different?

Viewing a DISK

Contents do not start at 0



Files

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

Files

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"

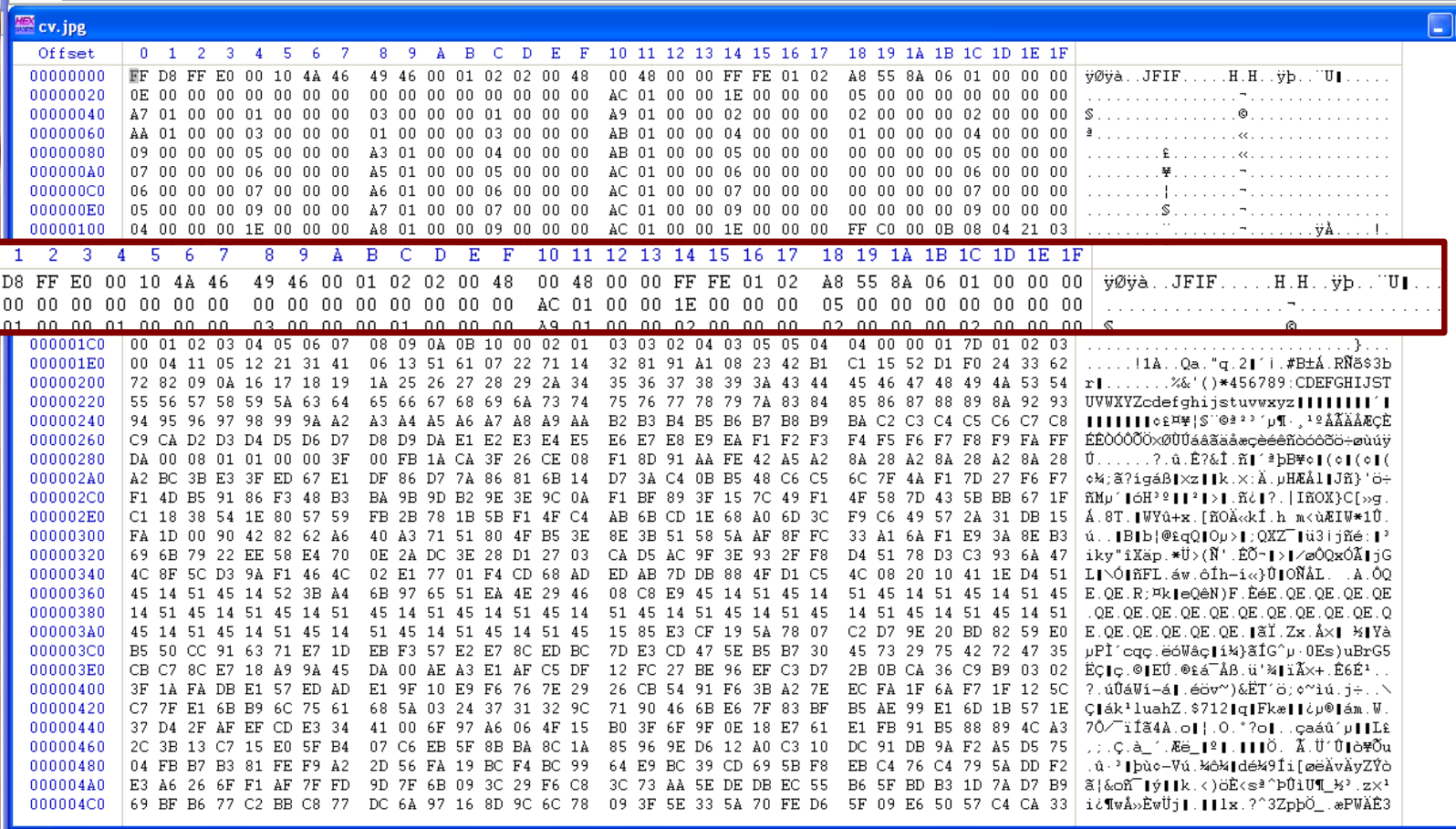
; JPEG Bitmap graphics JPE=0x00:0xFF,0xD8,0xFF,0xE0,0x00,0x10,"JFIF"

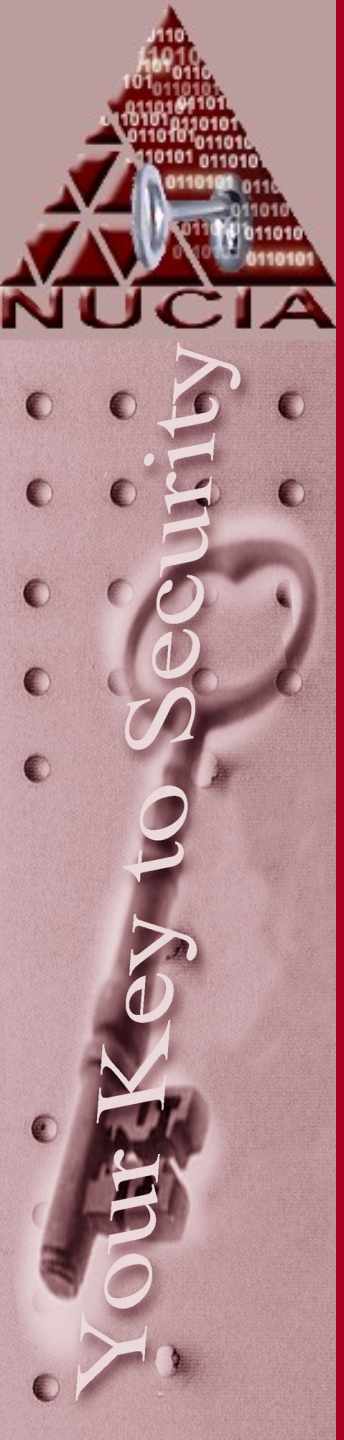
; JPEG Bitmap graphics JPG=0x00:0xFF,0xD8,0xFF,0xE0,0x00,0x10,"JFIF" JS=0x00:"/"

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

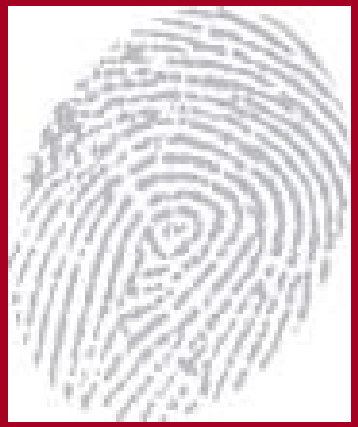
- In some cases you may find portions or fragments of a file. If you suspect that the fragment may be part of what used to be JPEG for example (because near where the header should be you found “FIF” and you know that jpeg headers contain “JFIF”) you can attempt to recover the file by editing the correct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

Hashing

```
$ md5sum.exe SavedEmail.txt
b13e0863a5ab9f329b59a0d15519f9ea *SavedEmail.txt
```

```
$ sha1sum.exe SavedEmail.txt
017f2def39d04cb5f42cb0562b5bc4b41b2eebc8 *SavedEmail.txt
```

```
[root@localhost root]# md5sum SavedEmail.txt ; sha1sum SavedEmail.txt
04851fe8301dde9086e1838e9d564b72  SavedEmail.txt
7fac072552ac7368eb42e54eae90f35987bcaf08  SavedEmail.txt
[root@localhost root]#
```

- The algorithm is independent of OS – the same hash is produced from the same file on Linux and Windows

Your Key to Security

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

```
[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?

– http://en.wikipedia.org/wiki/Hash_collision

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

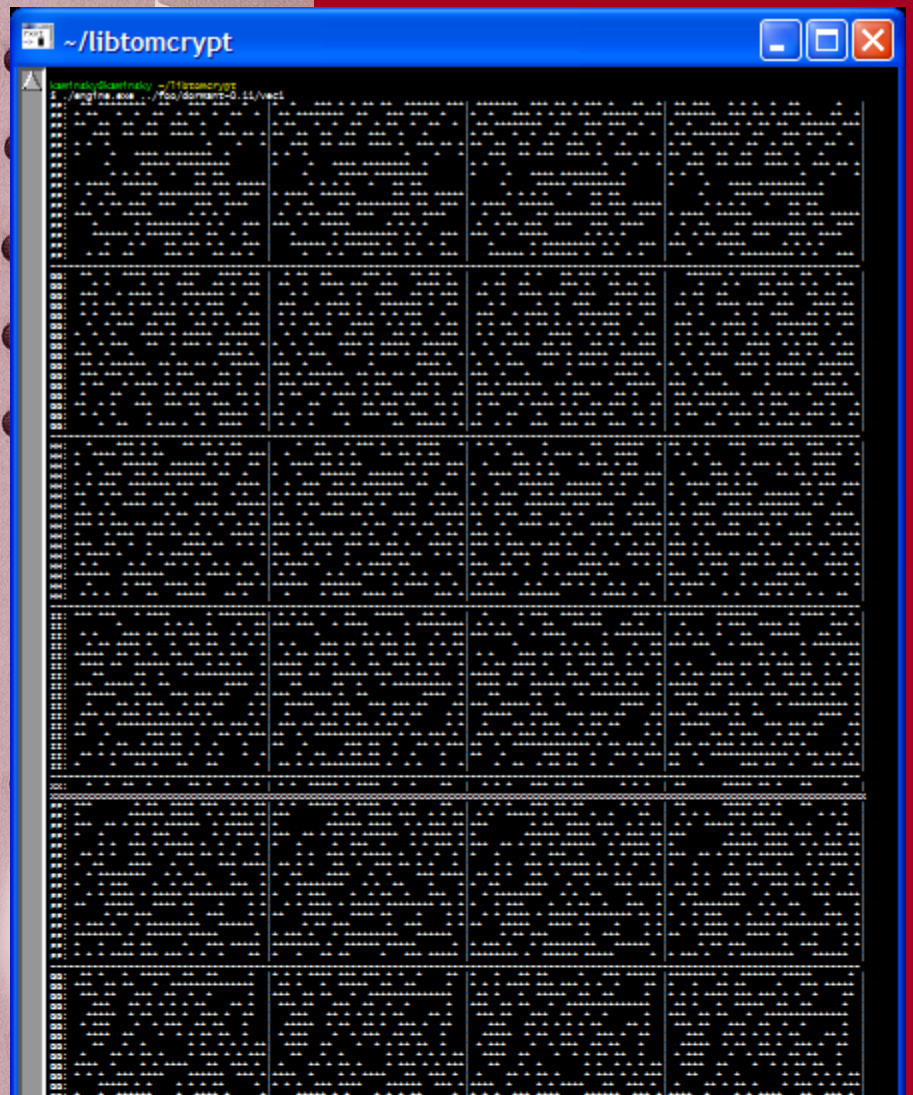
- Tools like stripwire can actually create 2 files that have the same md5…and very quickly

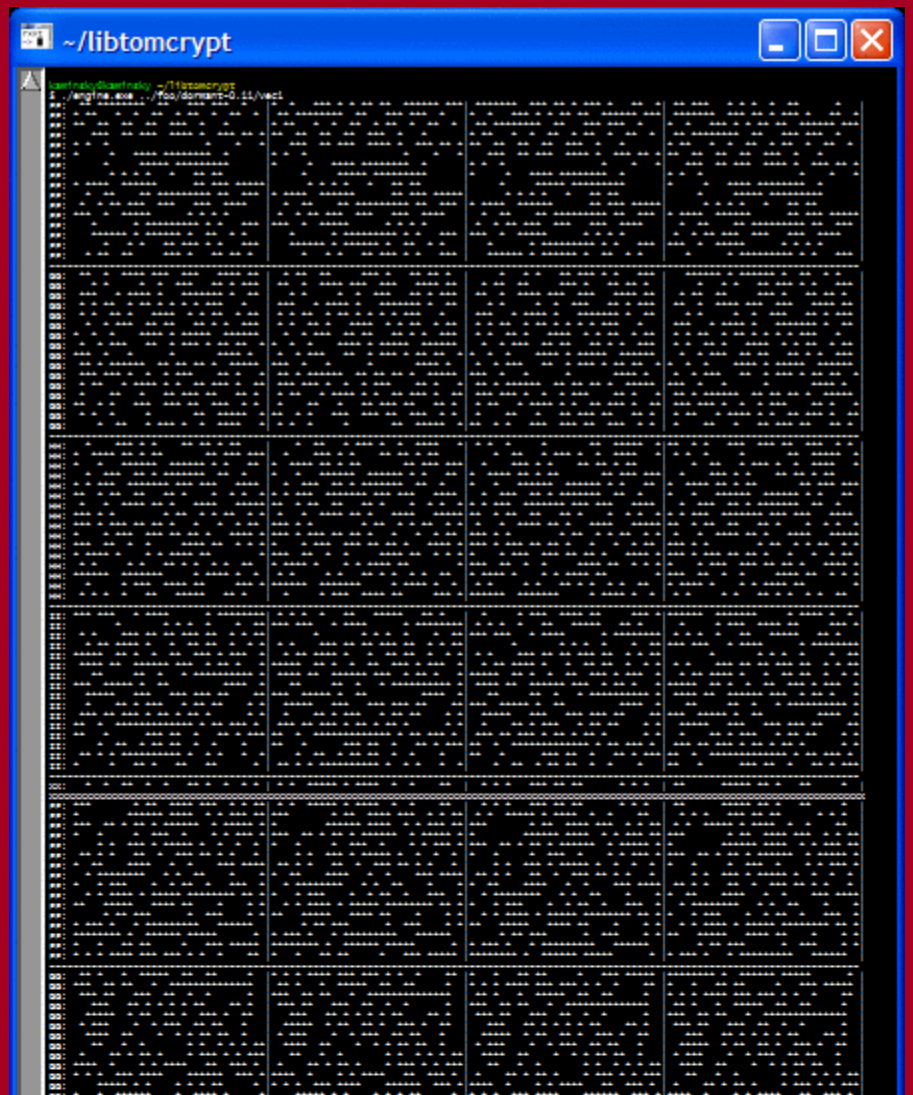
```
stripwire $VERSION:  Conflation Attack Using Colliding MD5 Test
Vectors
      Author:                       Dan Kaminsky(dan\@doxpara.com)
     Example:   ./stripwire.pl -v -b test.pl -r fire.bin
     Options:   -b [file.pl]     :  Build encrypted archives of this
                                    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 Collisoins

- 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 disscussion 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 34 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

- Casey, E. (2001). *Digital Evidence and Computer Crime*. Academic Press.
- Casey, E. (2002). *Handbook of Computer Crime Investigation: Forensic Tools and Technology*. Academic Press.
- Kruse, W.G. III, & Heiser, J.G. (2002). *Computer Forensics : Incident Response Essentials*. Addison-Wesley.
- Mandia, K., Prosise, C., & Pepe, M. (2003). *Incident Response: Investigating Computer Crime*. Osborne.

References

- Stephenson, P. (2001). *Investigating Computer-Related Crime*. CRC Press.
- Center for Strategic & International Studies (CSIS) http://www.csis.org/pubs/cyberfor.html
- **http://www.ascld-lab.org/**
- **http://www.Dcci.gov**