Overview

• Background
• Early Forensics
• Modern Forensics/Trends
• Future
Senior software engineer at Cayuga Networks focusing on evaluations (and other things)

• Background in distributed systems, network protocols, security, digital forensics, etc.

• Assorted hands-on hacking at various levels (OS, library, application, UI, etc.)

• Been involved in digital forensics R&D for 10+ years including organizer for Digital Forensics Research Workshop (DFRWS) since 2005

Background: Me
Background: What is it?

**Digital Forensics:**

"Tools and techniques to recover, preserve, and examine digital evidence on or transmitted by digital devices;" "But also a lot more potential sources... Computers, disks, and cell phones."

**Digital evidence:**

BackGround: What is it?
Where is Evidence

- Media/disks
  - File system metadata
  - File data
  - Unallocated space (deleted files)

- Network
  - Flow logs, full packet
  - Attacker traceback
  - Passwords

- Memory
  - Process structures, RAM
  - Process structures, RAM
- Swap files
- Temp files
  - Browser caches
  - Databases, Registry, binary

- Files
  - Configuration
  - Logs
  - Temporary files
  - Structured
    - Media (video, audio)

  - Files
  - Structured
  - Temporary files
  - Media (video, audio)
May be used in a court of law, depending on context

Must be factual and clear

Present the results (report, deposition, court testimony, ...)

Impartial: looking for answers, not just cherry-picking results

How do we fix it, what did they do, what did they get? (IR)

Operate under rules (legal, policy, etc.)

What happened, who did it, when, and why (Internet)

Find out details

- Use cryptographically secure hashes to preserve evidence
- Chain of custody
- Make sure "nothing changed" (at least as much as possible)

Preserve information

- Background: Goals
Background: Hashes

- Cryptographically secure hash function
  - Map arbitrary size input into fixed size output (hash)
  - Used to preserve evidence
    - Show that evidence has not changed between time hash was taken
    - Examples: md5, SHA-1, SHA-3
      - (e.g., half the bits)
    - “Pre-image resistant”
      - Small change in input produces large change in output
      - Given output, cannot create input that produces output that matches
      - Given original input, cannot reproduce original input (one-way)
      - Map arbitrary size input into fixed size output (hash)

- “Used to preserve evidence”
  - Show that evidence has not changed between time hash was taken
  - Examples: md5, SHA-1, SHA-3
    - (e.g., half the bits)
  - “Pre-image resistant”
    - Small change in input produces large change in output
    - Given output, cannot create input that produces output that matches
    - Given original input, cannot reproduce original input (one-way)
    - Map arbitrary size input into fixed size output (hash)
Law Enforcement

Government Agencies

Companies (I.R.)

Archivists

Try not to get killed.
Analysts in the field.
Very time limited.

Get as much as you can.

Fix it NOW! Find out what you can.
Have a standard process to follow.

Try to preserve what you can.

Background: DF Users and Goals

"What's a cassette tape?"
"The program's on a cassette tape?"
"The end of the republic."
Preserve as much as possible "until"

Archivists

Preserve as much as possible "until"

Civil: "Clear and convincing evidence"
Civil: "A preponderance of evidence"
Civil: "Beyond a reasonable doubt"

Criminal: "Beyond a reasonable doubt"

Standard of Proof:
Preserve to a court of law.
Preserve evidence for years.
Follow the rules.

Law Enforcement
Examine Everything on the Disk

- Extract files by extension
- Extract files by types
- Extract deleted files (unallocated space)
- Slack space (steganography)
- Can analyze a disk with a hex editor or even vi!
Quick Delve Into...

• Slack space
• Basic file structures
• Deleting (and undeleting files)
• MAC Times and Other Metadata
Slack Space

File

Disk blocks (512 bytes each)

File size: 4000 bytes

8 blocks = 4096 bytes on disk

416 bytes used

by the file

96 bytes of

“slack space” of

File
If there was not enough data in the buffer to fill a block of a file, it would not be overwritten and could be retrieved.

Old data in the last block of a file could be secret data that could be used to hide secret data.

Modern OSs zero the buffers before writing, leaving no artifacts in the slack space.

• If there was not enough data in the buffer to fill a block in the slack space, just use what was already in the buffer.

Slack Space...
### Brief File Systems Overview

- Directory entry 1
  - Wee File
  - Long File Name
  - i = inode number
  - s = file name size

- Directory entry 2
  - Block Group
  - Super-block
  - Data blocks
  - Inode Table
  - Inode Bitmap
  - Block Bitmap
  - Inode Bitmap
  - Descript Block
  - Group Block Group

---

**Block Group**
Unix File Systems

• Directories contain files, each entry has:
  – the file name
  – the file size
  – permissions
  – owner and group

• Inodes (index nodes) contain information about:
  – pointer to an inode
  – the file name
  – the file size

• Directories contain files, each entry has:

Unix File Systems
File Systems...

Attributes, ownership, permissions

Data blocks

Direct block

Indirect block

Double indirect block

Triple indirect block

Data blocks
Recovering Deleted Files: FAT, ext2

- FAT: OS deletes files by setting first byte in directory entry in the first byte of the file itself.
- ext2: Entry removed from directory, inode and data blocks marked as free.
- Directory and marking data sectors free.
- Undelete by reading all directory entries or undeleting by re-reading all directory entries on disk (if they haven't been reused).

23
• Timestamps on files:
  – Modify, Access, Inode Change, Delete, Create (last two)
  – Modify, Access, Inode Change (optional)

- Where is investigator?
- Is or was it daylight saving's time but not anymore?
- System clock on local time or UTC?

- Synchronization with multiple computers?

- Time resolution
  – FAT: 10 msec (creation), 2 sec (modification), 1 day (access)
  – NTFS: updated 1 hour (access), 100 nsec precision
  – Linux: ext3: 1 sec, ext4: 1 nsec

- Time zone
  – High precision, not necessarily high accuracy

- Timestamps on files:
  – timestamps
File Metadata...

• When did an event occur
  – Editing a file
  – Copying files
  – Sitting in front of the computer
  – Downloading pictures

• Permissions
  – Who else could get at or alter the data?
  – Is this a potential possession or distribution crime?
In addition to when and who, it may provide provenance data—

- Browser history
- Document history
- Email (to, from, subject, time, message-id, IP address)

Examples

- Where did that phrase you cut and pasted in Word come from?
- Where did that document come from?
- Where did that printing come from?

Application Metadata
2000s Era: Automation

• Still manually intensive
• Build primitive timelines
• Find files, deleted files
  - The Coroner’s ToolKit (TCT) and later The Sleuth
  - The Coroner’s ToolKit (TCT) plus commercial tools
• Parse file systems without using the OS
• Search sophistication: less than grep
• Simple keyword searches
Modern File Systems

- Basic file structures
- Deleting (and undeleting files)

- Similar but a bit more complex

- NTFS deletes entry from dir but leaves MFT entry (inode) intact with pointer back to parent, can reconstruct file path, name, and contents.

- Ext3 wipes inodes, harder to undelete without carving

MAC Times and Other Metadata

- Privacy and OS efficiency modifications can make some information more difficult to get
File Carving

• Deleted files still exist as data runs of blocks on the disk in unallocated space.

• Challenge: Can these blocks be reassembled?

• Answer: File Carving (e.g., Scalpel)

• Carvers have deep knowledge of various file formats (jpeg, gif, png, mp3, pdf, etc.)

• Look for start and end markers

• Look for ways to link the end of one span to the beginning of another one
Modern Era Problems

- Disk capacity (base system is >=1TB)
  - 3TB @ $100

- SSIDs
  - Trim command permanently deletes data (boo)
  - Wear leveling creates many copies of data (yay)
  - Much faster than mechanical disks (yay)

- Encryption (file, whole disk, and network)
  - Without a password, it’s all opaque
  - With a password, it’s all opaque

- SSDs

- Disk capacity (base system is >=1TB)

- Modern Era Problems
Disks Keep Getting Bigger

Source: Wikipedia.org
Big Data, Big Problems

- Imaging a 1TB disk can take a few hours (depending on too many factors)

- Indexing the data can take a long time (depending on too many factors)

- Most forensic investigators (short of large government agencies) don’t have those kind of resources to allocate per case.

- Distributed computing, GPUs, and storage (many spindles) can help, but...
Solid State Drives
SSDs: Pages and Blocks

Page: 4KB or 16KB

Block: 128 - 512 pages (not shown)
A 16GB disk might have 3200 blocks.

Example: 512KB total (128 4KB blocks in this example)
SSDs: Writing

SSD can write a page at a time, if it's an empty page.

Zerod page

Non-zero page
SSDs: Erasing

SSD can only erase an entire block at a time, which is slow.
SSD Trim Command

Previously, an OS updates its own free block bitmap when it deletes files (does not tell the disk). SSD will not indicate a block is no longer in use. The disk doesn’t know if pages are free, so it must preserve all pages. When the SSD erases a block, it must copy all trimmed blocks. (SSDs have more space than they advertise to support this.)

Now, the OS sends Trim command to SSD to preserve (copy) trimmed blocks. The OS sends Trim command to SSD to indicate a block is no longer in use. SSD will preserve trimmed blocks.

SSD Trim Command
Trimmed blocks have a lifetime typically measured in seconds and cannot be recovered. Trimmed blocks have a lifespan of 100K or 1M writes. Blocks are constantly being erased in an SSD. Blocks have a limited lifetime (100K or 1M writes). SSDs perform wear leveling anytime they have power, i.e., whenever plugged into a USB even if OS isn’t running. Wear leveling used for writes to prevent a block from being written too few times (static) or too many times (dynamic). Wear leveling used for writes to prevent a block from being erased constantly being erased in an SSD. Trimmed blocks have a lifespan typically measured in seconds and cannot be recovered. Trimmed blocks have a limited lifetime (100K or 1M writes). Good for privacy, bad for forensics. Limits effectiveness of file carving. Trimmed blocks have a limited lifetime (100K or 1M writes). Blocks have a limited lifetime (100K or 1M writes).
• Analysis II: The Next Generation

• Tools provide a deep understanding of specific data formats

• Tools help with:
  – File carving, web history, timeline analysis, email (text messages for phones), registry, and more
  – File carving, web history, timeline analysis, peer-to-peer/file sharing analysis, email (text messages for phones), registry, and more

• Tools provide a deep understanding of specific data formats

• Tools provide a deep understanding of specific data formats
Registry

- Tree-like key/value store for Windows, represented as a linked list on disk
  - Hives (files)
  - Keys (path in a tree)
  - Values (path to leaf)
  - Data type (string, binary, 2 byte, 4 byte, multi-string, etc.)
  - Data (value of leaf) as a linked list on disk

For forensic investigators, registry can either be...

- Programs store configuration and status data
Registry

- Recently used files
- Typed URLs
- Installed programs
- Previously installed programs
- Programs automatically run at log on
- Devices that had been plugged in
- Configuration settings, data locations
- Recently used files
- Typed URls
Recently Used Files

HKLM\SOFTWARE\Microsoft\Office\12.0\PowerPoint\File MRU
Runs When A Specific User Logs In
Runs when any user logs in.
USB Devices the Computer Has Seen
Peer-to-Peer
Current Trends: Peer-to-Peer

- Many clients (programs) and servers (protocols) and servers (protocols)
- Each program stores files in different locations, has different defaults, and different configuration files
- Investigators need to determine where the shared files are
- Where the metadata and logs are
- What programs are present
- What data has been shared/distributed
- Then they must get it all in a forensically sound way
Trends: Live Forensics

- Get context of what's happening on system NOW
  - Running processes
  - Active network connections
  - Memory dump (memory-only programs, encryption keys, etc.)

- Can pre-deploy agent (plan to get hacked)
  - Push an agent on the machine
  - Reboot into a forensic DVD-only OS (for some machines rebooting erases all memory) and contaminating/overwrite some evidence

- Or, break out a can of compressed air and...

- Or, break out a can of compressed air, and...
Data Remanence
Remanence at Room Temp

Bye, bye Bitlocker.

-50°C Brrrrrrrrrr

Liquid Nitrogen: AN HOUR
Canned air: MINUTES

Source: http://citp.princeton.edu/memory

5 mins (Gone)
Trends

• Memory analysis
  – Best Practices 10 years ago was “pull the plug”
  – DFRWS 2005 Memory Challenge
    – DFRWS 2005 Memory Challenge

• Memory analyses
  – Identified “key schedules” in memory, reduce brute force search space and crack AES fast!
  – New tools when given a memory dump can

– Motivated a lot of work in the area of memory forensics
  – Provided scenario, memory dumps, kernel, network capture, etc., and 7 questions to answer

– Provided scenario, memory dumps, kernel, network

– Provided scenario, memory dumps, kernel, network

• Trends
Trends

- Link Analysis – graph who talks to whom

- New devices, new evidence sources
  - Cell phones, tablets, watches
  - Smart TVs store a lot of information

- Reverse Engineering
  - Defeat hiding mechanisms
  - Understand what’s going on

In Soviet Russia, TV watches YOU!

Trends

- Link Analysis – graph who talks to whom
Trends: Reverse Engineering for...

- Analysis of large installation break-ins
  - Break-ins at Target, Home Depot, JP Morgan, Chase, Walmart, (IRS?) just this year!
  - Deeply compromised systems
  - Quietly compromised
- Bot-nets (more sophisticated C&C
- Anti-forensics (detection)
- Line between IR and LE practitioners are fuzzy at best

"Advanced Persistent Threat"

Analyses of large installation break-ins...
Archival Forensics

• Donated collections, retiring professors, …
• OLD systems ('90s era and beyond)
• Privacy is a huge issue—what can be released to the public?

Is it “real” forensics? Yes!

• Drafts of famous papers and books
• Contacts
• Email

Archival Forensics
Archival Forensics

Goals:
- Preserve information
- Find out what happened, how it worked, catalog data and media
- Limited information/cooperation from owner
- Don’t change sources if possible (archival principles)
- Don’t present before a court of law

Main differences:
- Create new tools as needed
- Use standard tools (Encase, FTK, The Sleuth Kit, …)

Methods:
- Media, internet
- Preserve information

Constraints:
- Don’t change sources if possible (archival principles)
- Limited information/cooperation from owner
- Don’t present before a court of law
Archival Principles

• Provenance/Authenticity
  – Respect des fonds – Keep records emanating from the same source together
  – Group without mixing them with others
• Preserve order
  – Don’t organize things in a way that distorts the original
• Access
  – Least restrictive access, reasonable redaction
  – Donor agreement
• Don’t speak for them
  – Donor agreement
• Access context

Archival Principles
Forensics benefits from information leaks from applications, OS, and more.

Forensics and security are at odds (privacy too).

Correlation of multiple data sources can provide a much more detailed picture of events.

Generally most sources were not intended to be used for forensics and security are at odds (privacy too). Applications, OS, and more.

Privacy mechanisms can foil some analyses.

Often these sources change or are removed.

Forensics benefits from software written without regard to deep state inspection was never considered.

Analysis requires deep knowledge and meticulous detail.

Summary
Challenges and Research Problems

- Handling big data (need to move beyond floppy disk mentality)
  - Using partial images/samples (legal and technical)

- Integrating more sources of evidence
  - Using partial images/samples (legal and technical)

- Automating event reconstruction

- Reverse engineering
  - Anti-forensics (tools and techniques to defeat forensic analyses)

- Privacy concerns

- Different groups have different goals, requirements, and restrictions
  - "If it's good enough for law enforcement... not true for everyone"
  - "If it's good enough for law enforcement... not true for everyone"

- Automation while avoiding "black box" syndrome ("because the program said so" is not a valid response to "why did you claim my client is responsible?")

- "If it's good enough for law enforcement... not true for everyone"

- Little research money and limited commercial market

- "If it's good enough for law enforcement... not true for everyone"
or later to: frank@notfrank.com
Questions, comments now,

Thank you
A Few References/Sources

Digital Forensics Research Workshop, www.DFRWS.org (all papers and presentations)

Journal

Digital Investigation


Fun Book

• The Cuckoo’s Egg by Cliff Stoll

Tools

• The Sleuth Kit (www.sleuthkit.org)
• Volatility (memory forensics) – http://www.volatility.org/
• Scalpel – https://github.com/sleuthkit/scalpel

Conference

• GRF’08, L.R. For Remote Live Forensics

Volatility (memory forensics) – http://code.google.com/p/volatility/

• The Cuckoo’s Egg by Cliff Stoll

• Network Forensics and Good Storytelling, 1989

“Cool” Research

• Digital Investigation/
digital-investigation/

Journal

• http://www.journals.elsevier.com/digital-investigation/

• all papers and presentations (all workshops, www.DFRWS.org)