



Overview

- What is Digital Forensics?
- Basic Approach and Process
- Basis Products



What is Digital Forensics?

- Process to answer questions about digital states and events.
- Digital Investigation is a more accurate term
- Typical forensics deals with:
 - Identification
 - Individualization
 - Comparisons
- Digital forensics is a process of searching and analyzing



Examples

- An employee is suspected of violating a company's Internet-usage policy.
- A hard disk is found in the house of a suspected terrorist.
- Abnormal logs are observed on a server - a security breach is suspected.
- A person is suspected of a murder or kidnapping.



Basic Approach / Process

1. Preserve the state of the computer (the crime scene)
2. Survey the data for “obvious” evidence
3. Based on survey results, search for more detailed evidence
4. Use evidence to reconstruct events



Step 1: Preservation

- Goal is to prevent the computer state from changing
- Typical approaches:
 - Unplug / power down the computer (if it is running)
 - Make a bit-wise copy of the hard disk
 - Copy only specific files
 - Copy memory before computer is powered off
- Approaches to copying hard disks:
 - Boot computer using trusted media:
 - *Media Exploitation Kit (Basis), Linux Live CD, DOS Floppy*
 - Remove hard disk and place in trusted system
- Legal Requirements:
 - May need to show that the original drive was not modified
 - May need to show that the copy has not been modified



Image File Formats

- Data are typically saved to a file in a special format
- Raw format
 - Same size as original hard disk
 - Does not store metadata (serial number, MD5 hash, etc.)
 - Is not compressed
 - Widely supported
- EnCase format
 - Compresses data
 - Stores metadata
 - Not open or extensible
- AFF (Basis)
 - Compresses data
 - Stores metadata
 - Open and extensible
 - Can store metadata separate from image data



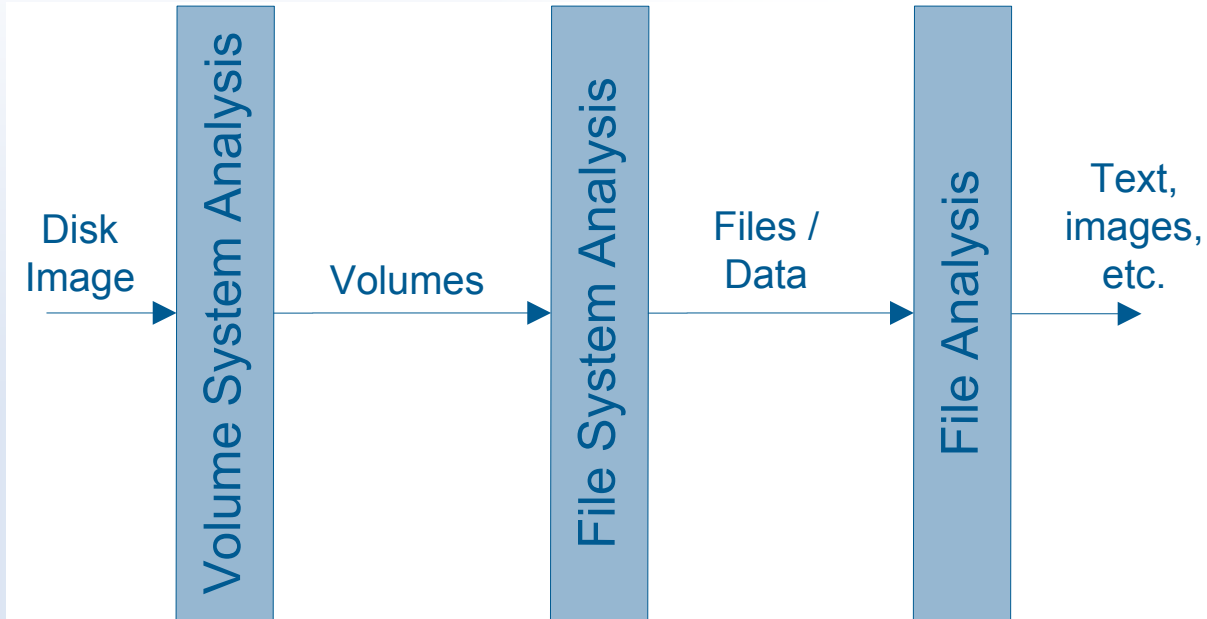
Step 2: Survey for Obvious Evidence

- From experience, investigators initially look for general types of evidence based on the scenario.
 - Intrusions: Logs, rootkits, hidden files and directories
 - Contraband graphic images: Image files, web history
 - Intelligence: Documents, e-mails
- Specialized tools are used for this.
- The tools serve four major functions:
 - Process data structures
 - “Data analysis”
 - Present the data to the investigator
 - Perform searches of the data



Data Structure Processing

- Process data structures and extract data:
 - File systems
 - E-mail boxes
 - Office documents
- Recover unallocated data





Data Analysis

- Examine general types of data for information
 - Text analysis
 - *Unicode normalization*
 - *Language Identification*
 - *Named entity extraction*
 - *Transliteration*
 - Image analysis
 - *Steganography detection*
 - *Computer-generated vs. real image*
 - Video analysis
 - Executable analysis
 - File clustering / classification
 - Password cracking



Data Presentation

- File system data typically presented in “File Manager”-type interface
- The investigator looks for relevant data
- Non-automated

The screenshot shows a software interface with a menu bar at the top containing: FILE ANALYSIS, KEYWORD SEARCH, FILE TYPE, IMAGE DETAILS, META DATA, DATA UNIT, HELP, and CLOSE. On the left, there is a 'View Directory:' section for 'E:\' with buttons for 'OK', 'ALL DELETED FILES', and 'EXPAND DIRECTORIES'. The main area displays a table of files with columns for permissions, name, creation/modification times, size, and attributes.

Permissions	File Name	Creation Time	Modification Time	Size	Attributes	Other
r/r	label.exe	1996.10.14 01:38:00 (EDT)	2002.06.13 17:08:40 (EDT)	32016	48 0	182-128-4
r/r	legacy.inf	1996.10.14 01:38:00 (EDT)	2002.06.13 17:08:40 (EDT)	4654	48 0	183-128-4
r/r	lights.exe	1996.10.14 01:38:00 (EDT)	2002.06.13 17:08:40 (EDT)	35600	48 0	184-128-4
r/-	LMREPL.EXE	0000.00.00 00:00:00 (GMT)	0000.00.00 00:00:00 (GMT)	0	0 0	0
r/r	LMREPL.EXE	1996.10.14 01:38:00 (EDT)	2002.06.13 17:08:40 (EDT)	86800	48 0	185-128-4
r/r	loadfix.com	1996.10.14 01:38:00 (EDT)	2002.06.13 17:08:40 (EDT)	1131	48 0	186-128-4 (realloc)
r/r	loadfix.com	1996.10.14 01:38:00 (EDT)	2002.06.13 17:08:40 (EDT)	1131	48 0	186-128-4

Below the table, there are menu options: ASCII (display - report) * Strings (display - report) * Export * Add Note. The file type is identified as 'MS Windows PE 32-bit Intel 80386 GUI executable'. The bottom section shows the 'String Contents Of File: E:\system32\inetins.exe' with the following text:

```
!This program cannot be run in DOS mode.  
.text  
.rdata  
.data  
.rsrc  
.reloc  
MSVCRT.dll  
KERNEL32.dll  
USER32.dll  
OSVW
```



Data Searching

- If the investigator knows what he is looking for, the tool can search for it:
 - Keywords
 - *e-mail addresses*
 - *names*
 - File types
 - File names
 - File times
 - Specific files (hashes)



Keyword Searching

- Type in keyword
- Searches for ASCII, Unicode, and/or code page variants
- Logical or physical search



- Indexing is used to make searches faster
- Most tools support regular expressions
- Result shows the location of the keyword
- Current limitations:
 - Little assistance with keyword expansion for different encodings or normalization
 - Some tools do not properly extract text from files (i.e. PDF)



File Type Searching

- Most file types have a unique signature
- All JPEG files start with 0xffd8 and end with 0xffd9, PDF starts with “PDF”
- One technique is to search for all files of a given type (i.e. JPEG)
- Can detect extension mismatches (rename bad.jpg to bad.exe)
- Can “carve” files from unallocated space in a file system



File Hash Searching

- Goal is to easily identify “Known” files
- Hashes of known files are calculated and stored
 - NIST NSRL
- Search to identify “Known Bad” files
 - Hacking tools
 - Training manuals
 - Contraband photographs
- Ignore “Known Good” files
 - Microsoft Windows files
 - Standard application files
 - Standard build files (corporate server deployments)



Cross-drive Searching

- Goal is to find relationships between computers
- Correlate computers based on:
 - E-mails and chat logs
 - Unique files
 - Names
 - Locations of files
 - Network shares and domains



Step 4: Event Reconstruction

- The evidence is the “effect” of an event.
- How did it get there? What program created it? What user caused the event?
- Common practice in:
 - Intrusion cases
 - Child pornography: Intentional download?



Basis and Digital Forensics

- Focus is on large scale data and foreign language support
- Large scale data:
 - Increasing hard disk sizes
 - Increasing number of connected computers
 - Provide ability for better collaboration among investigators
 - Provide more automation, correlation, and data analysis techniques
 - Provide open, plug-in frameworks for analysis modules
- Foreign languages:
 - Integrate text analysis functionality
 - Provide assistance with code pages and keyword expansion
 - Provide triage support to include language types and named entities



Basis Products

- **Media Exploitation Kit:**
 - Bootable CD and external USB / Firewire Drive
 - Automatically copies contents of hard disk to external drive
 - Requires little training
- **Advanced Forensic Format (AFF)**
 - Open and extensible disk image storage format
- **Open Digital Forensic Framework (ODFF)**
 - Open plug-in framework for analysis techniques
- **Automated Data Reduction System**
 - Identifies unknown files and clusters them
- **Multiple-drive Analysis System**
 - Allows for interactive analysis and collaboration
 - Can correlate computers together



Brian Carrier

brianc@basistech.com

617-386-7132