Lecture 37 – Computer Forensics

Ryan Cunningham
University of Illinois
ECE 422/CS 461 – Fall 2017

Security News

- Apple OS X login vuln patched
- Krebs: Leakbase went dark
- 3rd NSA TAO contractor pleads guilty for taking classified data home (Shadow Brokers leak)

INTRODUCTION

Digital Forensics

- Forensics –*the use of science or technology to discover evidence for a court of law*
- Used in both criminal and civil law
- Digital forensics – forensics relating to digital devices
- Usually trying to recreate a chain of events

Sub-disciplines

- Computer forensics
- Electronic discovery
- Network forensics
- Mobile forensics
- Database forensics
- Forensic data analysis

- Related fields
  - incident response
  - data recovery

LAW AND PROCEDURES

US Legal System

- Adversarial system – contest between injured parties and those responsible
- Civil law – settles disputes between private citizens (plaintiffs vs. defendants)
- Criminal law – punishes parties for illegal activities (prosecutors vs. defendants)
  - injured party is the state
- Judges and/or juries weigh evidence and arguments to make a decision

4th Amendment

- When can the state access data?
- *The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.*
- Warrant required when person has reasonable expectation of privacy
- Exceptions: plain view, consent, exigent
- "Fruit of the poisonous tree"

Digital Evidence

- Admissibility – must be relevant, not prejudicial
- Digital evidence is circumstantial
  - Links computers or devices to events, not people
- Chain of custody
  - Document every person who handles the evidence (hash values)
- Considered *scientific evidence*

Daubert Standard

- Rules for scientific evidence
- Judge is the ultimate “gatekeeper”
- Evidence must be relevant and reliable
  - Is the method subject to testing?
  - Are there established error rates?
  - Is it generally accepted by the community?
  - Has the technique been peer reviewed?

Digital Forensics Reporting

- Should be written for a lay audience
- Document how evidence was gathered in detail
- Present and explain evidence
- Four sections (assessment, acquisition, examination/analysis, and conclusions)

Electronic Discovery

- Discovery – in a civil case, both sides must share evidence
- E-discovery – digitally processing document requests for discovery process
- Indexing and searching large amounts files distributed across many computers
  - imagine Oracle vs. Google
- Lots of $$$ in this field

COMPUTER CRIMES

Computer Fraud and Abuse Act

- Title 18 U.S. Code Section 1030
- Enacted in 1986
  - Protected “federal computers” (government interest)
- In 1996, expanded to “protected computers”
  - Used by federal government or financial institution
  - Involved in interstate commerce
  - Also expanded to include using malicious code

18 U.S. Code § 1030 e

(1) the term “computer” means an electronic, magnetic, optical, electrochemical, or other high speed data processing device performing logical, arithmetic, or storage functions, and includes any data storage facility or communications facility directly related to or operating in conjunction with such device, but such term does not include an automated typewriter or typesetter, a portable hand held calculator, or other similar device;

(2) the term “protected computer” means a computer—

> **(A)** exclusively for the use of a financial institution or the United States Government, or, in the case of a computer not exclusively for such use, used by or for a financial institution or the United States Government and the conduct constituting the offense affects that use by or for the financial institution or the Government; or
>
> **(B)** which is used in or affecting interstate or foreign commerce or communication, including a computer located outside the United States that is used in a manner that affects interstate or foreign commerce or communication of the United States;

COMPUTER FORENSICS

Investigation Procedure

1. Preparation – get authorization, equipment, and personnel
2. Survey – consider the scene and make a plan
3. Preservation – seize physical devices and data
4. Examination – extract evidence (artifacts)
5. Analysis – combine evidence and draw conclusions
6. Reporting – summarize findings

Preservation Decision

- Live analysis – look at evidence on live computer
  - prevents file encryption/passwords
  - risks damaging evidence
- Dead analysis – image drives
  - bit-for-bit forensic duplicate using a write blocker
  - analysis can be conducted later
  - identify hidden or deleted files

Digital Artifacts

- Operating system: event logs, registry data
- File system: access times, modification times
- Disk: deleted files, hidden partitions
- Internet: browser history, email
- Media: photos, videos, audio
- Documents: Office, PDFs, RTF, XML
- Databases: MySQL, Oracle
- Application data: instant messaging

FILE SYSTEM FORENSICS

Common File Systems

- UFS: Unix File System
- Ext[2,3,4]: Extended file system
- FAT[12,16,32]: File Allocation Table
- NTFS: New Technology File System
- ReFS: Resilient File System
- HFS[+]: Hierarchical File System

Master Boot Record (MBR)

- Data structure stored on the first sector of the drive.
- Cross-platform industry standard for locating and booting disk partitions
- Used by BIOS/UEFI to boot (start) the OS
- Contains:
  - The partition table describing each partition
  - Boot loader code to fetch and execute a partition

Boot Sector

- Jump instruction to bootstrap code
- Data structures describing partition
  - Magic number (for NTFS)
  - Bytes per sector
  - Sectors per cluster
  - Sectors per track
  - Location of master file table ($MFT) and mirror
  - Serial number, checksum
- Bootstrap code (loads operating system)

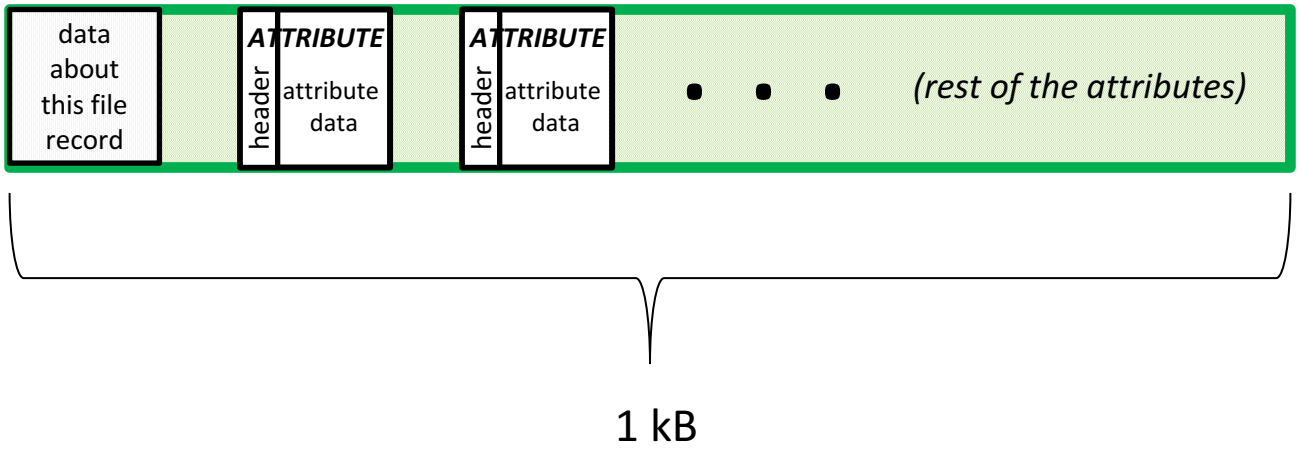
Master File Table

- File/folder records are stored in $MFT
- Records are 1 kb
- Records contain ***attributes*** that define the characteristics of a file, including the data itself

MFT File Record Attributes

- In NTFS, files are a collection of attributes
- Attributes define the characteristics of a file, including the data
- Attribute header specifies name, size, and type
  - Multiple attributes of the same type
- Attribute content

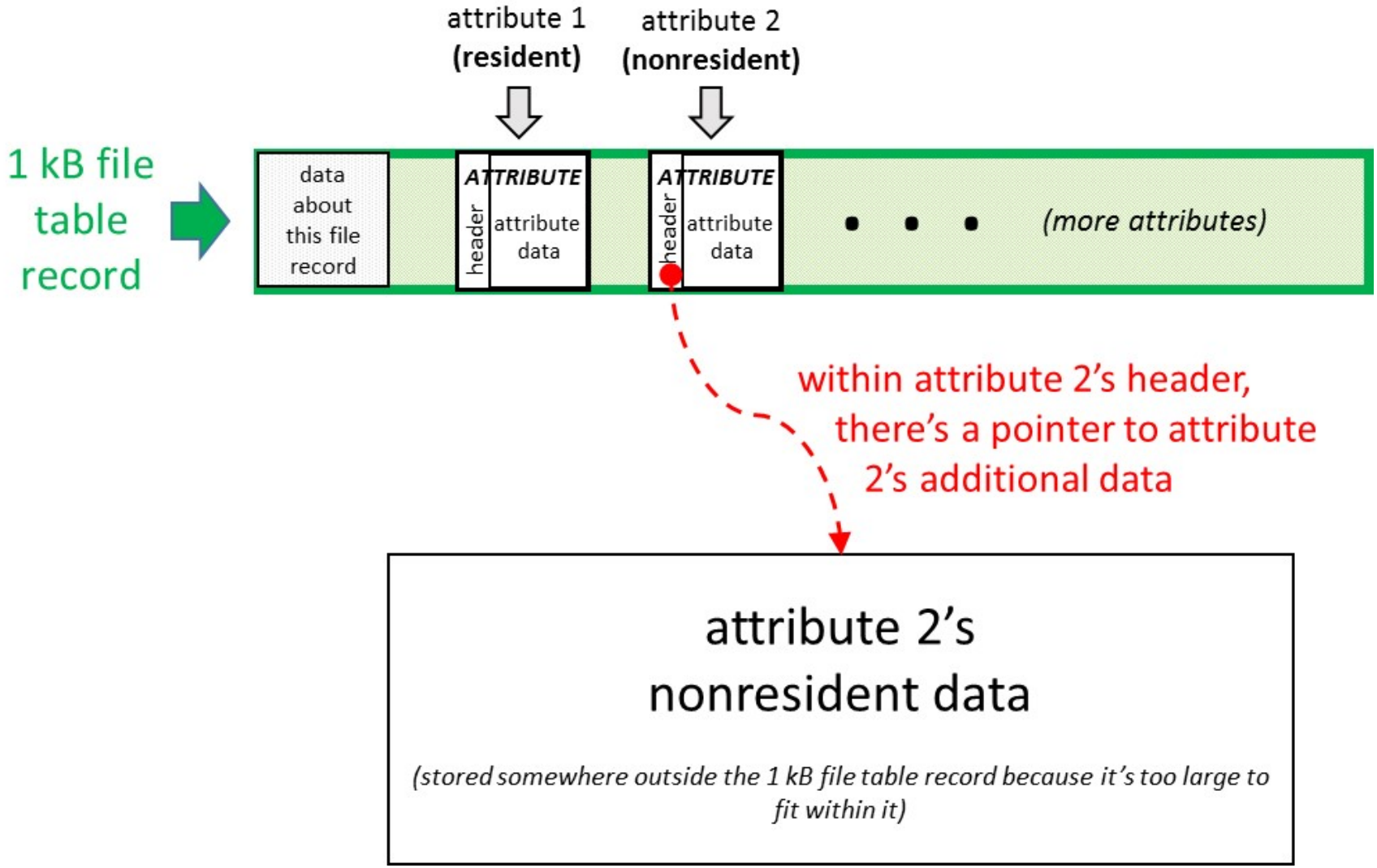
a file table record:



Nonresident Attributes

- Attributes can be any size and can often be ***larger than a single MFT entry***
- Resident attributes live in the MFT itself
- Nonresident attributes have some data located outside the MFT
- Pointers to disk locations outside the MFT
- Still considered part of the file record

FTR with Nonresident Data



Standard Attribute Types

- Most file record attributes are of predefined, standard types.
- $STANDARD_INFORMATION
  - Owner, creation time, link count, etc.
- $INDEX_ROOT, $INDEX_ALLOCATION, and $BITMAP
  - Implement folders and B+ tree.

Standard Attribute Types, cont.

- $FILE_NAME
  - Name of file, parent directory, size on disk, etc.
- $SECURITY_DESCRIPTOR
  - The access control list (ACL).

$STANDARD_INFORMATION

- Every file and directory has one.
- Contains primary timestamps.
- Also contains general properties flags.
  - E.g., read-only or compressed.

$FILE_NAME

- Every file record has one.
- Contains a pointer to its parent directory's file record.
- Can be used to recover path of deleted file.
- Contains timestamps, but Windows does not update them reliably.

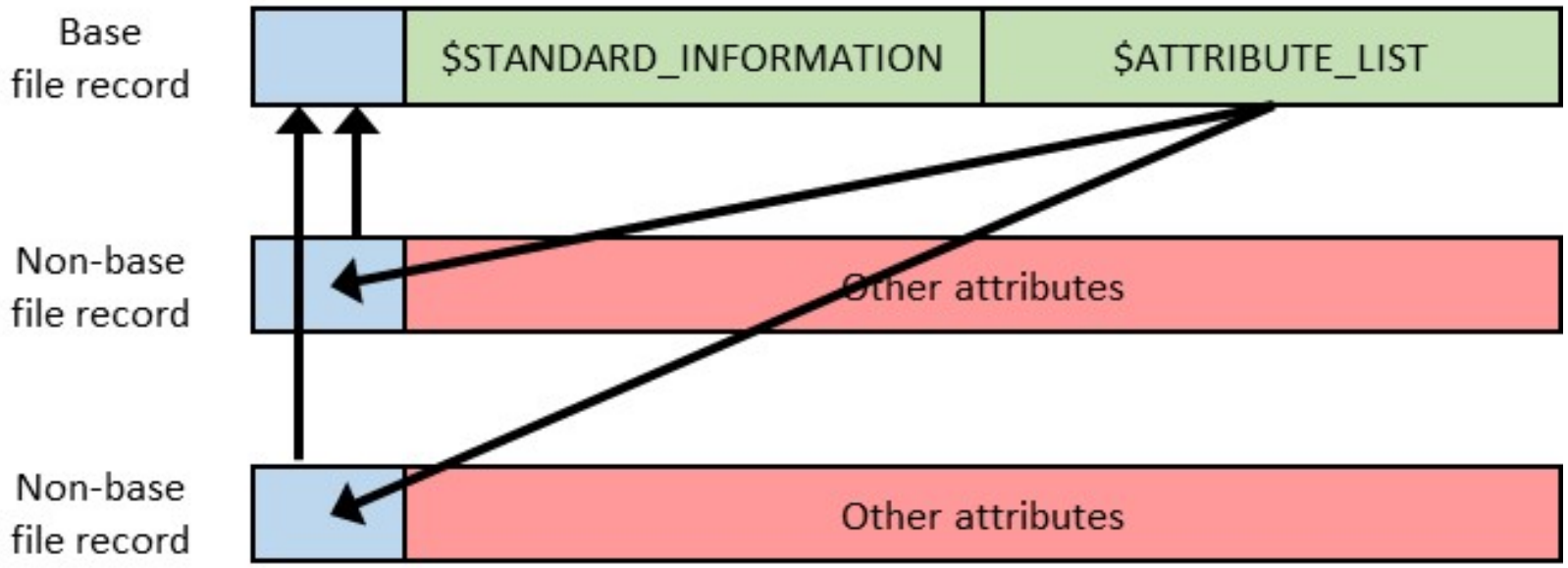
$DATA

- Stores file ***content***.
- Every file has one.
- Directories do not normally have them.
- If there is more than one $DATA attribute, the subsequent instances are called *Alternate Data Streams* (*ADS*).

Non-Base File Records

- File records have fixed size.
- If attribute headers are too large, non-base file records are allocated.
- Non-base file records have pointer to base.
- Base file record has $ATTRIBUTE_LIST.

$ATTRIBUTE_LIST



- $ATTRIBUTE_LIST contains location of all other attributes.
- Non-base entries contain location of base entry.

Metadata Files

- Files that contain data about the file system itself
- $MFT: master file table (file records)
- $MFTMirr: mirror of $MFT
- $LogFile: transaction log to recover from failure
- $BadClus: bad clusters
- . Root directory
- $Secure: special access control list database

$MFT

- First file record in the master file table.
- $DATA attribute contains the rest of the master file table.
- $BITMAP attribute stores allocation status of ***file records***.
- $STANDARD_INFORMATION attribute has file system creation date.

$MFTMirr

- Backup of most important file records
- $DATA attribute contains $MFT, $MFTMirr, $LogFile, and $Volume
- Can be used by a recovery tool or forensic analyst to restore a damaged file system

$Boot

- Contains boot sector (yes, it's also a file).
- $DATA attribute content must be at sector 0.
- $DATA attribute will contain the location of the first master file table file record and the boot code.
- Boot code used to load operating system.

$BitMap

- ***Careful!*** We're talking about the $BitMap ***metadata file***, *not an attribute*.
- Defines allocation status of ***clusters***.
- $DATA attribute contains actual bitmap.
- 0 indicates an unallocated cluster; 1 indicates an allocated cluster.

Recycle Bin

- Just another folder from the file system's point of view.
- Contains a folder for each user account.
- Clicking "Delete" in Windows just updates the file name and parent directory.

Files in the Recycle Bin

- For each file shown in the Recycle Bin, two files exist on disk.
- A file with name starting with “$R” followed by a pseudorandom number contains the original contents of the file.
- A file with name starting with “$I” followed by the same pseudorandom number contains the original filename and path.

Example: Delete \Examples\file.dat

1. Process the $INDEX_ROOT and $INDEX_ALLOCATION attributes of the root directory's file record to find the file record for the "Examples" directory. Update the last accessed time of the root directory.
2. Process the $INDEX_ROOT attribute of the "Examples" directory's file record to find the "file.dat" index entry, which contains the file record number for "file.dat".

Example: Delete \Examples\file.dat

3. Remove the index entry for “file.dat” from the “Examples” directory index and rebalance the tree if needed. Update last written, modified, and accessed times for “Examples” directory.
4. Deallocate the file record for “file.dat” by clearing the “in use” flag in its file record header and clearing the corresponding bit in the $BITMAP attribute of the $MFT file record.

Example: Delete \Examples\file.dat

5. Process the nonresident attribute headers in the file record for “file1.dat” to find the clusters storing the nonresident content. Set the corresponding bits to 0 in the $Bitmap metadata file.

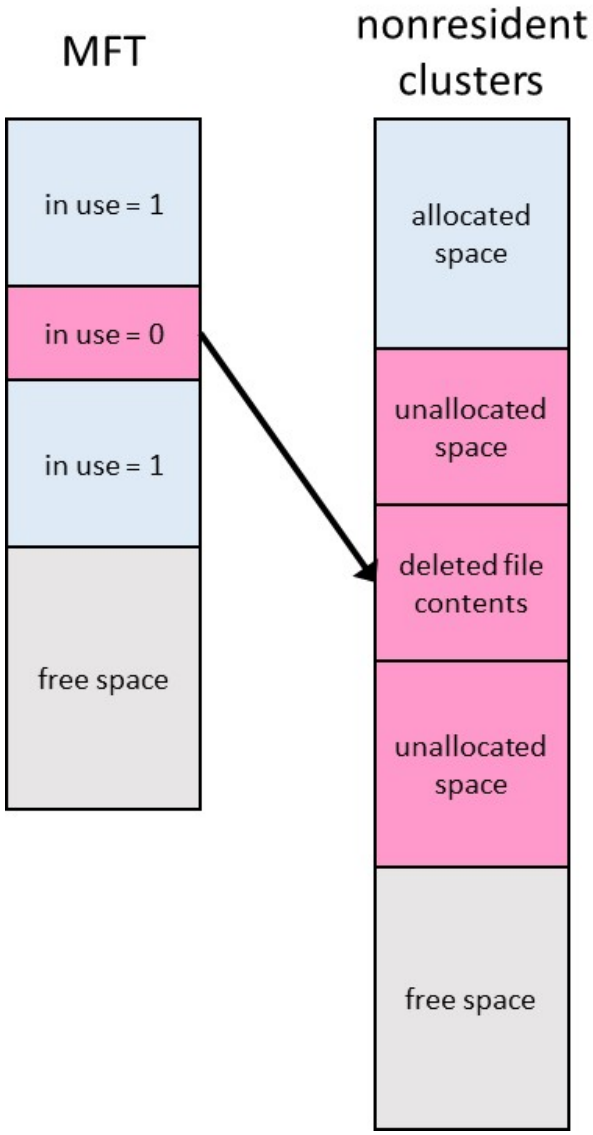
- Note: Nonresident content clusters were NOT overwritten and the file record “file.dat” was NOT overwritten; they were just marked as unallocated.

Deleted File Recovery

- Resident attribute: just read the content.
- Nonresident attribute: read attribute headers to find location of content.
- Easy to automate.
- Sometimes called *undelete*.

Recovery with File Record

- File record intact.
- File contents intact.



FILE CARVING

File Carving

- Modern file systems tend to overwrite metadata for deleted files
- Recovery of files ***without*** file system metadata
- Can be done without ***any*** metadata
- Can carve many different kinds of media

Recovery vs. Carving

- Recovery: file system metadata are intact; use them to find file (“undelete”)
- Carving: pulls the ***raw*** bytes from the media

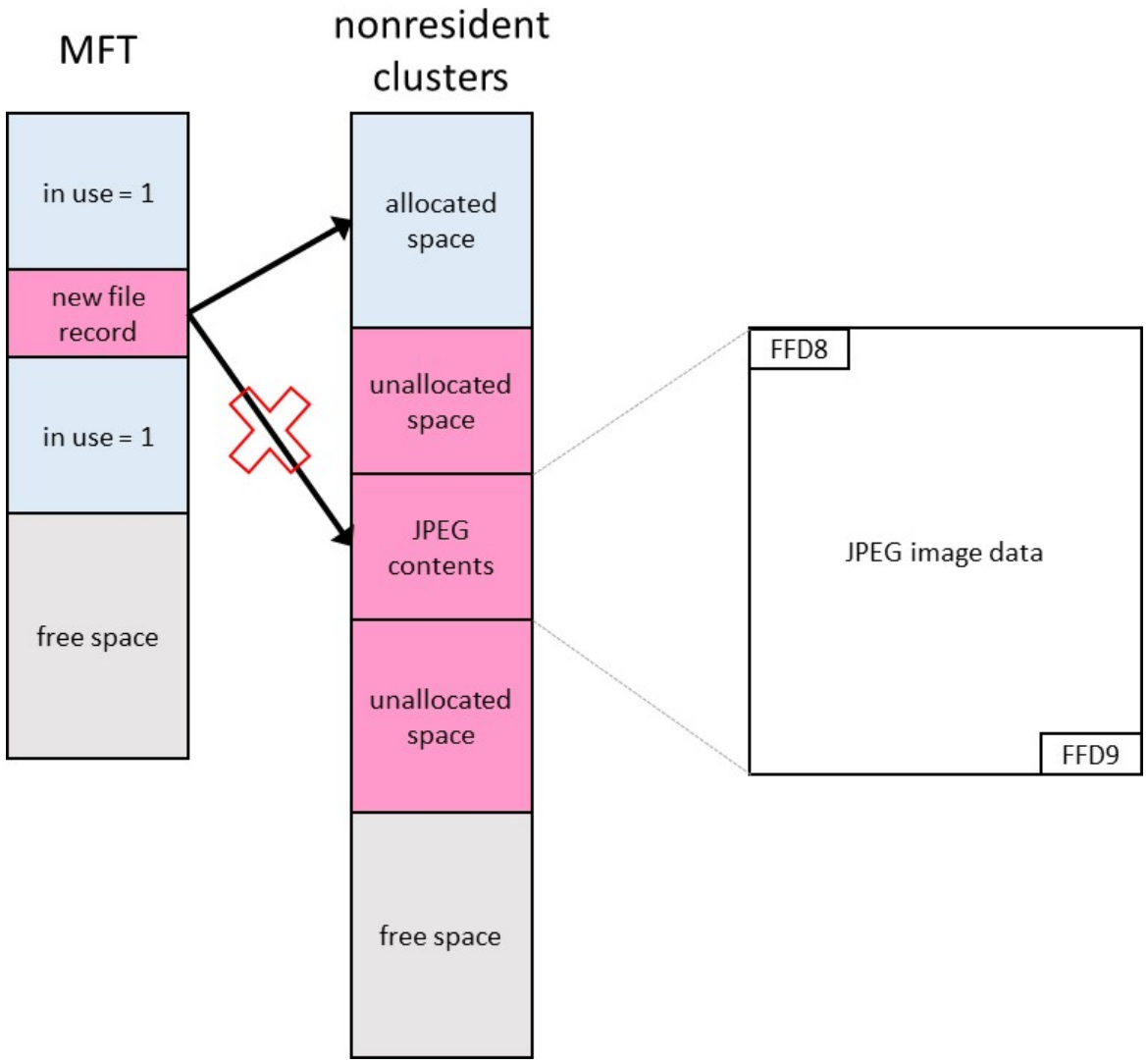
Basic File Carving Techniques

- Header-footer carving
- File-structure-based carving
- Content-based carving

Header-Footer Carving

- Many file types have standard headers and footers stored inside
- Example: A JPEG starts with “Start of image” header 0xFFD8 and ends with “End of image” footer 0xFFD9
- Carve out everything between JPEG header and footer $\rightarrow$ image file

Carving a JPEG



Header-Footer Carving Problems

- Header or footer might be a common string.
  - E.g., the header of an MP3 is “mp”
  - Produces false positives
- The beginning of the file might be missing
  - If file was deleted, might get squashed by a newer file
- The footer might be missing
  - Big output!
- The disk could be fragmented

File Structure Carving

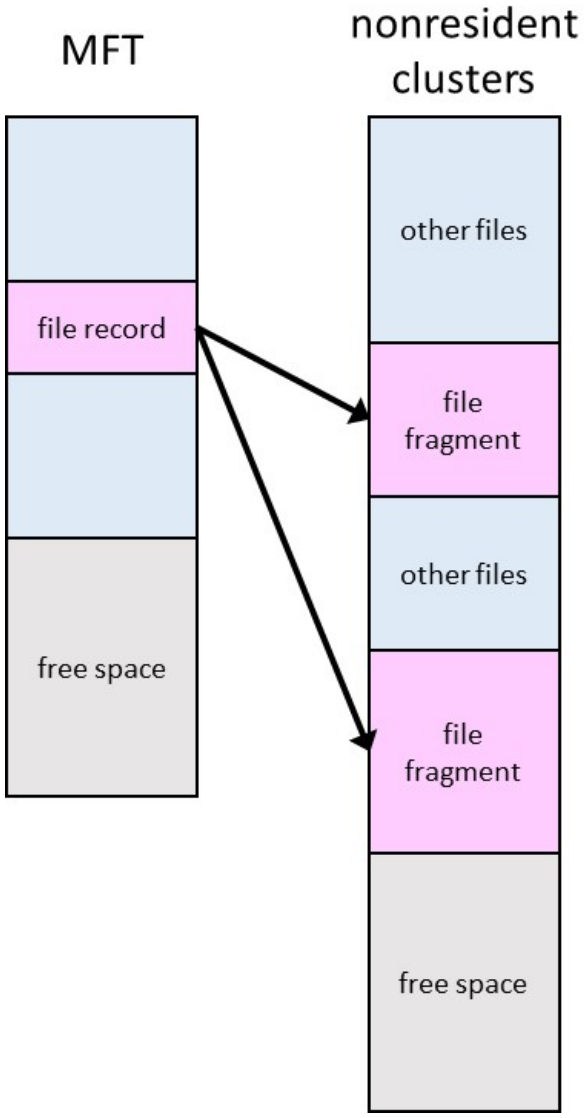
- Use internal file metadata if available
- Find cluster size
- Read entire cluster and hunt for internal signatures (in addition to header/footer)
- Example: Foremost, PhotoRec

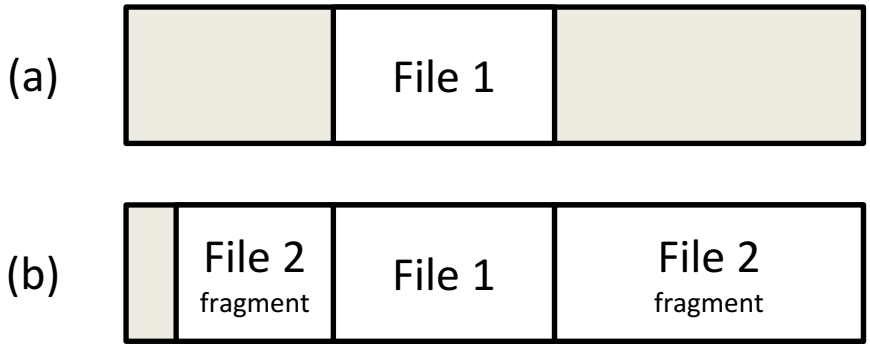
Content-Based Carvers

- Look for statistical signatures indicating language or file content
- Machine-learning/statistic-based
- Semantic carvers

What is Fragmentation?

- Nonresident content is split into noncontiguous chunks
- Multiple pointers in attribute header
- Causes: partition close to full, file size changes
- Becoming less common

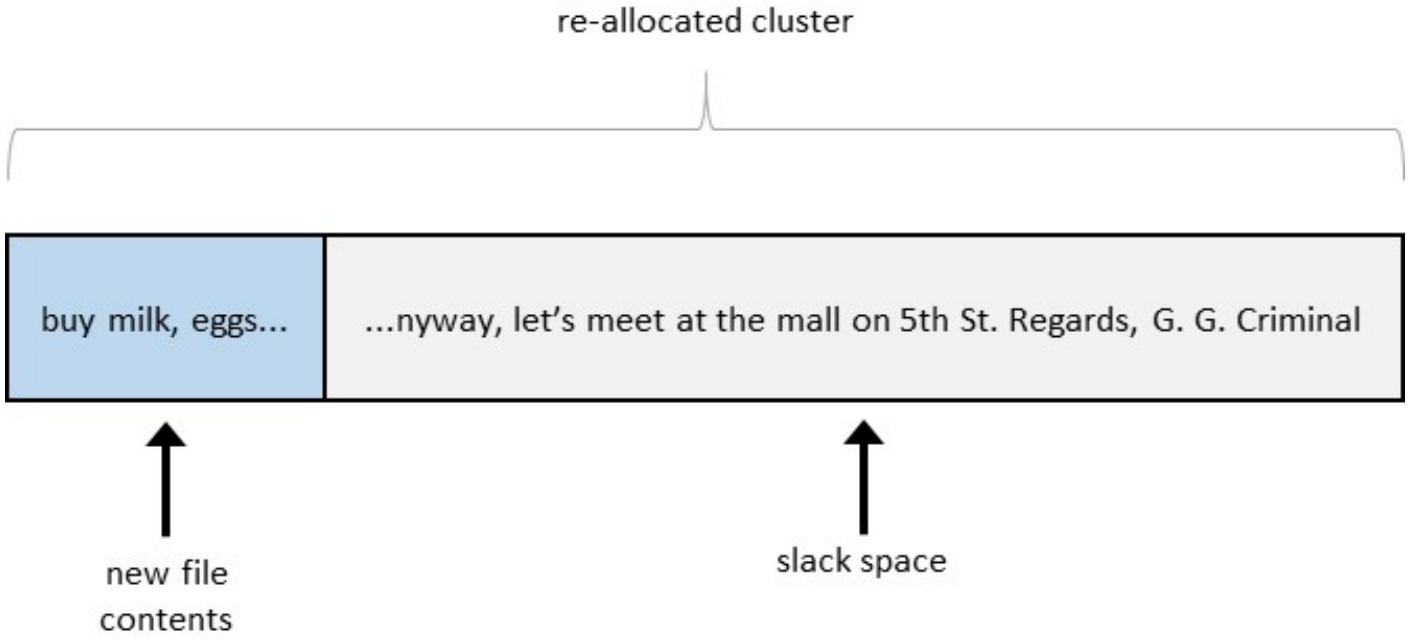




With Fragmented Clusters

- Much harder
- Many theoretical solutions, few practical
- Two production-quality solutions:
  - Bifragment Gap Carving
  - SmartCarving

Slack Space



- *Slack space*: leftover space after file contents in a cluster.
- Old file's data may remain in slack space.

WINDOWS ANALYSIS

Registry Keys and Values

- Each hive is a single file
- Hives define a hierarchical structure of keys and values
- Keys are like directories; values are like files
- Registry keys have timestamps, but many Registry editors (such as regedit) cannot view these timestamps

Traces of User Log On/Off

- Last user to log in:
  - HKLM\SOFTWARE\Microsoft\Windows\CurrentVersion\Authentication\LogonUI\LastLoggedOnUser
- Last time the computer was shut down:
  - HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Control\Windows\ShutdownTime
- Last write time on the ntuser.dat file indicates last logout time.

Connection of USB Devices

- Recently connected USB devices are shown in HKLM\SYSTEM\CurrentControlSet\Enum\USBSTOR
- Contains two levels of subkeys:
  - The device type (e.g., Cruzer micro 1 GB flash storage)
  - The instance ID (usually a serial number, but will be pseudorandom if no serial number is found)

Event Logs

- Windows tracks events logged by applications and the system in separate log files in C:\Windows\system32\winevt\logs
- Contains information about some of the same events as the Registry
- Discrepancy between event logs and Registry may indicate tampering

Viewing Event Logs

- Since Vista, stored in a documented XML format containing date and time, user account, event ID, and a description of the event
- Can be opened outside original system, but event description may not be correct
- The Registry links the log file to the application DLLs to find the message text

Reconstructing Event Logs

- To view event logs accurately, need original Registry entries and DLLs
- Options:
  - Collect logs on original device (requires live analysis)
  - View logs on same version of Windows (need DLLs for any application logs)
  - Open disk image in virtual environment

TIMELINE ANALYSIS

Timeline Analysis

- Goal of forensics is to reconstruct events that happened in the past
- Useful to integrate digital evidence into one coherent timeline
  - System logs, Registry, and file system timestamps
  - Internal file timestamp data
  - Combine multiple media
- Tools exist to create tabular timeline (Excel .csv file)

Timeline Creation

- mactime (from TSK)
- log2timeline (Perl) $\rightarrow$ plaso
  - Super-timeline of Apache Web logs, Windows event logs, pcap files, Exif data, Linux syslogs, link file metadata, Windows firewall data, McAfee antivirus logs, mactime data, Web browser data, and Registry data
- Autopsy, EnCase

Timeline Investigation

- Establish a baseline/pattern of life for users
  - When do they typically log on/off?
  - Anomalous behavior could be important
- Build histogram of events based on time of day or day of week
- Two approaches: targeted vs. bulk timeline
  - Gather everything, then do SQL/Excel analysis?
  - Find relevant artifacts, then build focused timeline?

Timeline Difficulties

- Timestamps depend on time zone. Be very careful in integrating evidence.
- Unconnected/synced devices can get clock skew.
- Lots of data, difficult to visualize/analyze.
- Be careful when data mining and anomaly hunting in someone's life.
  - Easy to profile.