

Foundational Concepts of Digital Evidence

MSCS630: Cybercrime Investigation and Digital Evidence

Dr. Mouatez Karbab



أكاديمية جوعان بن جاسم للدراسات الدفاعية
Joaan Bin Jassim Academy for Defence Studies

October 2, 2026



Outline

- 1 The Order of Volatility
- 2 The Three States of Digital Data
- 3 File Structure: Data, Metadata, and Signatures
- 4 Modern and Emerging Evidence Sources

Introduction to Foundational Concepts of Digital Evidence I

Objectives

- Order evidence by volatility; collect the most volatile first.
- Tell data in transit, in use and at rest apart, and how to collect each.
- Separate content from metadata; read MACB times; identify files by signature.
- List where smart-device, cloud and disappearing-message evidence survives.

Why This Matters

- Knowing the nature of digital evidence helps you find it.
- It also stops you from accidentally destroying it.



Outline

- 1 The Order of Volatility
- 2 The Three States of Digital Data
- 3 File Structure: Data, Metadata, and Signatures
- 4 Modern and Emerging Evidence Sources



Volatility and the Collection Rule I

Definition: Volatility

- How easily data can be lost or changed.
- Volatile data vanishes at power-off or on a state change [4].

The Rule

- **Collect the most volatile evidence first**, especially in a *live response*.
- Collect first whatever would be lost first [6].
- Pulling the plug erases Random Access Memory (RAM).
- RAM holds running programs, network connections and decrypted data.

Real-World Scenario: The Hasty Shutdown I

The Hasty Shutdown

- Responder cuts power to a leaking server to take it to the lab.
- Lost: *fileless malware*, running only in memory.
- Lost: the attacker connection and the list of other victims.
- The disk later shows only a normal operating system.
- A solvable case became unsolvable.
- With disk encryption, the only unlocked key copy is lost too.

Terms Used in the Volatility Tiers I

- **Address Resolution Protocol (ARP) cache:** maps Internet Protocol (IP) to physical addresses.
- **Process table / routing table:** running programs; how traffic is sent.
- **Swap / pagefile:** on disk, but overwritten so often it counts as volatile.
- **Dead-box forensics:** analysing a powered-off computer; main source is the disk.
- **Unallocated space:** disk area no active file uses; may hold deleted file parts.
- Central processing unit (CPU) registers change too fast to collect directly.
- Network-attached storage (NAS) sits in tier 7 with backups.



The Order of Volatility I

	Tier	Examples	Typical lifetime	How it is collected
1	CPU registers and cache	Values the processor is working on now	Nanoseconds	Not collected directly in practice
2	Network and process state	Routing table, ARP cache, process table, kernel statistics	Seconds to minutes	Trusted live-response commands run from external media
3	System memory (RAM)	Running programs, open connections, decrypted data, encryption keys	Until power-off	Memory acquisition tool writing a memory image to external media
4	Swap space and temporary files	Pagefile, swap partition, program temporary files	Minutes to days, overwritten in use	From the disk image, or copied live
5	Disk (HDD, SSD)	User files, system files, logs, unallocated space	Until overwritten or trimmed	Forensic imaging through a write-blocker
6	Remotely logged data	Central log server, firewall logs, cloud activity logs	Set by retention policy	Export from the log server, or legal process to the provider
7	Archived media and backups	Backup tapes, external drives, NAS, cloud backups	Months to years	Imaging or copying the backup media

Table: The order of volatility. From most to least volatile, adapted from Request for Comments (RFC) 3227

[4]. The RFC places memory in the same tier as network and process state, and lists physical configuration and

The Order of Volatility II

network topology between remote logs and archival media. Lifetimes are typical, not guaranteed: a busy system overwrites swap within minutes, and a retention policy can delete remote logs within days.



Section 1: Assessment I

Question 1

Why did the hasty shutdown make the case unsolvable?

- A. The disk was damaged by the power cut.
- B. The write-blocker was missing.
- C. The fileless malware existed only in RAM.
- D. The remote logs were deleted.



Section 1: Assessment - Answer I

Question 1

Why did the hasty shutdown make the case unsolvable?

- A. The disk was damaged by the power cut.
- B. The write-blocker was missing.
- C. The fileless malware existed only in RAM. (✓)
- D. The remote logs were deleted.

Explanation:

- Collect first whatever would be lost first [6].



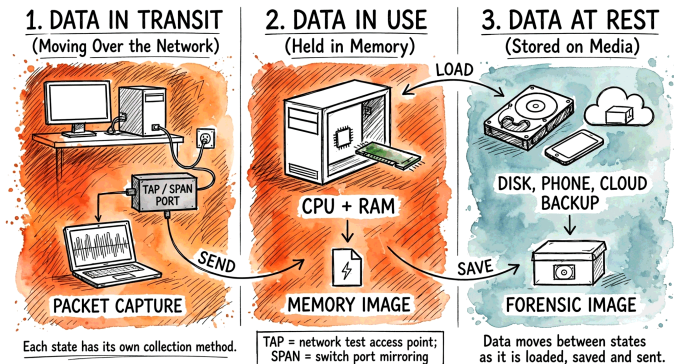
Outline

- 1 The Order of Volatility
- 2 The Three States of Digital Data
- 3 File Structure: Data, Metadata, and Signatures
- 4 Modern and Emerging Evidence Sources



The Three States of Digital Data I

- Volatility says *when*; the state of the data says *how*.
- In transit: network tap or SPAN port; in use: memory image; at rest: forensic image.
- Data moves between states as it is loaded, saved and sent.



Data in Transit and Packet Capture I

- Data moving over a local area network (LAN) or the internet; lost if not captured live.
- Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) packets, Hypertext Transfer Protocol (HTTP) and HTTPS requests.
- Voice over IP (VoIP) calls, application programming interface (API) calls, file transfers.
- **Packet capture:** network card in *promiscuous mode*.
- On a switch, use a Switched Port Analyzer (SPAN) port or tap; promiscuous mode alone sees mostly broadcast [18].
- Tools: **Wireshark** (graphical), **tcpdump** (command line) [7].

Practical Demo: Wireshark Filter I

- Show only HTTP POST requests, the method browsers use to send forms:

```
http.request.method == "POST"
```

- Works for plain HTTP only.
- HTTPS shows connection metadata, not form contents, without session keys.

Data in Use: The Memory Image I

- Data the CPU is processing or held in RAM; goal of a *live acquisition*.
- Running processes, unsaved documents, chats, clipboard.
- Most important: *cryptographic keys* for an encrypted disk.
- **Memory image:** bit-for-bit copy of physical memory, machine left on.
- Trusted tool from external media: WinPmem, Magnet RAM Capture, AVML or LiME [15].
- Hash the image file as soon as it is written.

Data at Rest: Forensic Imaging I

- Saved on storage; survives without power; the most stable state.
- Documents, photos, databases, emails, system and program files.
- Hard disk drives (HDDs), solid-state drives (SSDs), USB drives, memory cards, backups.
- Collected carefully and repeatedly, not urgently.
- **Forensic imaging:** bit-for-bit copy through a write-blocker [5].

Section 2: Assessment I

Question 1

On a switched network, what does promiscuous mode alone capture?

- A. All traffic on every switch port.
- B. Little more than broadcast traffic.
- C. Only encrypted traffic.
- D. Only traffic from the default gateway.

Section 2: Assessment - Answer I

Question 1

On a switched network, what does promiscuous mode alone capture?

- A. All traffic on every switch port.
- B. Little more than broadcast traffic. (✓)
- C. Only encrypted traffic.
- D. Only traffic from the default gateway.

Explanation:

- A SPAN port or network tap is needed [18].

Outline

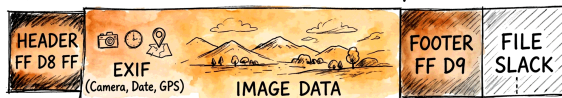
- 1 The Order of Volatility
- 2 The Three States of Digital Data
- 3 File Structure: Data, Metadata, and Signatures**
- 4 Modern and Emerging Evidence Sources



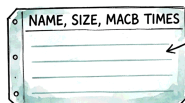
Metadata Inside and Outside a JPEG File I

- **Content:** what the file holds. **Metadata:** data about data, written automatically.
- Joint Photographic Experts Group (JPEG) file: signature, Exchangeable Image File Format (EXIF) block, footer, slack.
- File system record: name, size, MACB times, pointer to clusters.

1. INSIDE THE FILE (The Bytes on Disk)



2. OUTSIDE THE FILE (The File System Record)



MACB = Modified, Accessed, Changed, Born

The signature and EXIF metadata travel inside the file.

Name and MACB times are kept by the file system, outside the file.



MACB Timestamps Across Three File Systems I

- New Technology File System (NTFS), ext4, Apple File System (APFS): MAC or MACB times [5].

	Meaning	NTFS (\$SI)	ext4	APFS	Normally updated by
M	Content last modified	Modified	mtime	Modified	A write to the file's content
A	Content last read	Accessed	atime	Accessed	A read, subject to the last-access policy
C	Metadata record last changed	\$MFT entry modified	ctime (inode change)	Changed	A rename, permission change or content write
B	Created on this volume	Created	ctime, shown by stat	Created	File creation; a copy gets a new B time

Table: MACB timestamps across three file systems. NTFS times are those of the \$STANDARD_INFORMATION (\$SI) attribute, whose record lives in the Master File Table (MFT); APFS field names follow the Apple File System Reference [1]. The same letter can name a different event on each file system; on ext4 the C time is the inode change time, not a creation time. Last-access updates are often delayed or switched off for performance (`relatime` on Linux [17], `NtfsDisableLastAccessUpdate` on Windows [12]), so check the system's setting before relying on an A time.



Application Metadata and ExifTool I

- EXIF: camera model, settings, date and time, Global Positioning System (GPS) coordinates.
- Office documents: author and editing history; Portable Document Format (PDF): creation software.
- **ExifTool** reads GPS coordinates [9]:

```
exiftool -gpslatitude -gpslongitude image.jpg
```

Limits

- Places the device, not necessarily the suspect.
- EXIF is editable; corroborate before relying on it.



Common File Signatures I

- Extensions are part of the name and easy to change.
- Identify the type by its **file signature** (magic number) [11].
- Extension and signature disagree: a warning sign of hiding.

File type	Header (hex)	As text	Footer (hex)
JPEG image	FF D8 FF	(not printable)	FF D9
PDF document	25 50 44 46	%PDF	ends with %%EOF (25 25 45 4F 46)
Windows executable	4D 5A	MZ	none

Table: Common file signatures. A JPEG header continues with a fourth byte that names the variant, commonly E0 for the JPEG File Interchange Format (JFIF) or E1 for EXIF, so a carver matches on the first three bytes only. The text column shows the header bytes as American Standard Code for Information Interchange (ASCII) characters where they are printable.



Data Carving I

- Deletion usually removes only the pointer; space becomes unallocated.
- On a hard disk the data stays until overwritten.
- **Data carving**: scan raw data for a header, copy to footer or size limit [8].
- Works even when all file-system metadata is gone.
- Tools: **Foremost**, **Scalpel**, **PhotoRec**.
- SSDs: TRIM erases freed blocks; carving often recovers little [3].

Illustrative Case: A Recovered Planning Document

- Deleted spreadsheet with an attack plan, tied to the suspect's account and dated before the crime.
- One dated, attributable document can outweigh any volume of undated material.



Section 3: Assessment I

Question 1

On ext4, what does the C time record?

- A. File creation.
- B. Inode (metadata) change.
- C. Last content read.
- D. Last copy to another volume.



Section 3: Assessment - Answer I

Question 1

On ext4, what does the C time record?

- A. File creation.
- B. Inode (metadata) change. (✓)
- C. Last content read.
- D. Last copy to another volume.

Explanation:

- The same letter names different events per file system.



Outline

- 1 The Order of Volatility
- 2 The Three States of Digital Data
- 3 File Structure: Data, Metadata, and Signatures
- 4 Modern and Emerging Evidence Sources

Internet of Things and Smart Devices I

- **Internet of Things (IoT):** connected devices with sensors and software.
- Speakers, watches, thermostats, vehicle infotainment, security cameras.
- Non-standard formats, little storage, key data sent to the cloud [14].
- Combined data builds a detailed timeline of movements and routines.
- Can support or contradict an alibi, once sensor accuracy and user are established.

Cloud Storage and Ephemeral Messaging I

Cloud Storage

- Evidence may persist on provider servers after local deletion [10].
- Needs legal process on the provider; may involve several countries.

Ephemeral Messaging

- Snapchat deletes 24 hours after all viewed [13]; Signal uses a timer.
- Remnants: database free pages and write-ahead log [16].
- Notification history, caches, cloud backups, screenshots, the other party's device.

File-Based Encryption

- iPhone: per-file keys wrapped in metadata [2].
- Metadata gone means remnants cannot be decrypted.

Section 4: Assessment I

Question 1

Why is carving rarely productive on an iPhone?

- A. iPhones use magnetic storage.
- B. Each file has its own key; remnants cannot be decrypted once metadata is gone.
- C. iPhones have no unallocated space.
- D. Apple forbids forensic tools.

Section 4: Assessment - Answer I

Question 1

Why is carving rarely productive on an iPhone?

- A. iPhones use magnetic storage.
- B. Each file has its own key; remnants cannot be decrypted once metadata is gone. (✓)
- C. iPhones have no unallocated space.
- D. Apple forbids forensic tools.

Explanation:

- File-based encryption wraps each key in the file's metadata.

Conclusion: Key Takeaways I

- **Volatility:** decides what to collect first.
- **Three states:** in transit, in use, at rest decide how to collect it.
- **Files:** metadata and signature often say more than content.
- **Emerging sources:** copies and remnants often survive somewhere else.

Next Step

Apply these concepts to acquiring evidence from live and powered-off systems.

References I

- [1] Apple Inc.
Apple file system reference, 2020.
- [2] Apple Inc.
Apple Platform Security, 9 2022.
Published periodically.
- [3] Graeme B. Bell and Richard Boddington.
Solid state drives: The beginning of the end for current practice in digital forensic recovery?
Journal of Digital Forensics, Security and Law, 5(3), 2010.
- [4] D. Brezinski and T. Killalea.
Guidelines for evidence collection and archiving.
RFC 3227, IETF, 2 2002.
- [5] Brian Carrier.
File System Forensic Analysis.
Addison-Wesley Professional, Boston, MA, 2005.



References II

- [6] Eoghan Casey.
Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet.
Academic Press, San Diego, CA, 3rd edition, 2011.
- [7] Laura Chappell.
Wireshark Network Analysis: The Official Wireshark Certified Network Analyst Study Guide.
Chappell University, 2nd edition, 2012.
- [8] Simson L. Garfinkel.
Carving contiguous and fragmented files with fast object validation.
Digital Investigation, 4:2–12, 2007.
Proceedings of the Seventh Annual DFRWS Conference.
- [9] Phil Harvey.
Exiftool by phil harvey.
Software Website, 2024.
Accessed on August 2, 2025.



References III

- [10] Martin Herman, Michaela Iorga, Ahsen Michael Salim, Robert Jackson, Mark Hurst, Ross Leo, Richard Lee, Nancy Landreville, Anand Kumar Mishra, Yien Wang, and Rodrigo Sardinas.
NIST cloud computing forensic science challenges.
Technical Report NIST IR 8006, National Institute of Standards and Technology, 2020.
- [11] Gary C. Kessler.
GCK's file signatures table, 2025.
Accessed on October 2, 2026.
- [12] Microsoft.
fsutil behavior.
Microsoft Learn, Windows Commands reference, 2026.
Accessed on October 2, 2026.
- [13] Snap Inc.
When does Snapchat delete Snaps and Chats?
Snapchat Support, 2026.
Accessed on October 2, 2026.

References IV

- [14] Maria Stoyanova, Yannis Nikoloudakis, Spyridon Panagiotakis, Evangelos Pallis, and Evangelos K. Markakis.
A survey on the internet of things (IoT) forensics: Challenges, approaches, and open issues.
IEEE Communications Surveys & Tutorials, 22(2):1191–1221, 2020.
- [15] Joe Sylve.
Lime - linux memory extractor.
GitHub repository, 2012.
The canonical source code repository for the LiME tool.
- [16] The SQLite Project.
Write-ahead logging, 2026.
Accessed on October 2, 2026.
- [17] The util-linux project.
mount(8) — linux manual page.
man7.org, 2026.
Accessed on October 2, 2026.

References V

- [18] The Wireshark Foundation.
CaptureSetup/Ethernet.
Wireshark Wiki, 2026.
Accessed on October 2, 2026.

Acronyms and Terminology I

- APFS** Apple File System: The default file system on current Apple devices. It records created, modified, changed and accessed times for each file.
- API** Application Programming Interface: A defined set of calls that lets one program request data or services from another. API calls between services are one form of data in transit.
- ARP** Address Resolution Protocol: A network process that links a logical network address (like an IP address) to a physical device address (MAC address). The ARP cache stores this information temporarily and is a source of volatile evidence.
- ASCII** American Standard Code for Information Interchange: A character encoding that maps letters, digits and symbols to byte values. Some file signatures, such as %PDF and MZ, are readable ASCII text.
- Cache** A temporary storage area used by a program or operating system to hold data for quick access. Caches can contain remnants of data, even after it has been supposedly deleted.
- CPU** Central Processing Unit: The primary processor of a computer. Its internal memory (registers and cache) contains the most volatile form of data.
- Data at Rest** Data that is saved in a non-volatile state on a storage device, such as a hard drive or SSD. This data persists even when the device is turned off.



Acronyms and Terminology II

- Data Carving** The process of recovering files from a storage medium by searching for known file signatures in the raw data, often used on unallocated space where deleted files may reside.
- Data in Transit** Data that is actively moving across a network, from one point to another. It is highly volatile and must be captured as it moves or it will be lost.
- Data in Use** Data that is actively being processed by a computer's CPU or that resides in its volatile memory (RAM).
- Dead-box Forensics** The analysis of a computer or digital device that has been powered off. This process primarily focuses on examining data at rest.
- Digital Remnants** Traces or fragments of data that remain on a device even after a user or application has attempted to delete them.
- Ephemeral Messaging** A type of communication where messages are designed to be automatically deleted after being viewed, such as in apps like Snapchat or Signal.
- EXIF** Exchangeable Image File Format: A standard for storing metadata within image files. This data can include the camera model, date, time, and GPS coordinates of where the photo was taken.
- ExifTool** A powerful command-line tool for reading, writing, and editing metadata (like EXIF data) within a wide variety of file types.

Acronyms and Terminology III

File Signature (Magic Numbers) A unique sequence of bytes at the beginning (and sometimes end) of a file that identifies the true file format, regardless of the file's extension.

File System The method and data structure an operating system uses to control how data is stored and retrieved on a storage device like a hard drive.

Fileless Malware Malicious software that exists and runs only in a computer's active memory (RAM) without writing itself to the hard drive, making it undetectable in a powered-off system.

Foremost A command-line tool used to perform data carving to recover deleted files based on their headers and footers.

GPS Global Positioning System: A satellite navigation system. Phones and cameras can write GPS coordinates into a photo's EXIF metadata.

HDD Hard Disk Drive: A traditional data storage device that stores data magnetically on rotating platters.

HTTP Hypertext Transfer Protocol: The protocol web browsers and servers use to exchange pages and form data. Plain HTTP traffic can be read in a packet capture.

HTTPS Hypertext Transfer Protocol Secure: HTTP carried over an encrypted connection. A packet capture shows its connection metadata but not its content unless the session keys are available.



Acronyms and Terminology IV

IoT Internet of Things: The vast network of interconnected physical devices, such as smart watches, security cameras, and home assistants, that collect and exchange data and can serve as sources of digital evidence.

IP Internet Protocol: The protocol that addresses and routes packets between networks. The ARP cache links IP addresses to physical device addresses.

JFIF JPEG File Interchange Format: A common JPEG variant, identified by the byte E0 after the FF D8 FF header.

JPEG Joint Photographic Experts Group: The common compressed image format used by cameras and phones. Its files begin with the bytes FF D8 FF and end with FF D9.

LAN Local Area Network: A network that connects computers within one building or site, such as a company office network.

Live Acquisition The process of collecting data, particularly from volatile memory (RAM), from a computer while it is still running.

Live Response The set of actions an investigator performs on a computer system while it is still powered on, with the primary goal of collecting volatile data before it is lost.

 **MAC Times** A set of critical file system metadata timestamps associated with a file, including its Modified, Accessed, Changed (metadata) and Born (created) times, often written MACB.



Acronyms and Terminology V

- Memory Image** A bit-for-bit copy of the contents of a computer's physical memory (RAM) at a specific point in time.
- Metadata** Data about data. It is information that describes a file's properties, such as its creation date, author, file size, and timestamps, rather than the content of the file itself.
- MFT** Master File Table: The NTFS table holding one record for every file and folder, including the \$SI timestamps.
- NAS** Network-Attached Storage: A dedicated file storage device connected to a network that allows multiple users and devices to store and retrieve data from a central location.
- NTFS** New Technology File System: The standard file system of Microsoft Windows. It keeps MACB timestamps in each file's MFT record.
- Order of Volatility** The fundamental principle in digital forensics that dictates evidence should be collected in order from most volatile (most likely to be lost) to least volatile.
- Packet** A small segment of a larger message that is sent over a computer network.
- Packet Capture (Packet Sniffing)** The act of intercepting and logging data packets as they travel across a digital network.



Acronyms and Terminology VI

PDF Portable Document Format: A document format whose files begin with the ASCII characters %PDF.

PhotoRec A file data recovery software tool designed to recover lost files, including video, documents and archives from hard disks, and lost pictures from digital camera memory.

Promiscuous Mode A setting for a network interface card (NIC) that allows it to capture all network traffic it can see, not just the traffic addressed to it.

RAM Random Access Memory: A computer's volatile, short-term memory where applications, the operating system, and current data are kept for fast access by the processor. Its contents are lost when power is turned off.

RFC Request for Comments: A document series published by the Internet Engineering Task Force. RFC 3227 gives guidelines for evidence collection, including the order of volatility.

Scalpel An open-source data carving tool that is an overhaul and improvement of the Foremost tool.

\$SI \$STANDARD_INFORMATION: The NTFS attribute that holds a file's main set of MACB timestamps.

SPAN Switched Port Analyzer: A switch feature that copies traffic from other ports to one mirror port, so a capture device can see it.



Acronyms and Terminology VII

SSD Solid-State Drive: A modern data storage device that uses integrated circuit assemblies to store data persistently, typically using flash memory.

TCP Transmission Control Protocol: A connection-oriented transport protocol that delivers data reliably and in order, used by the web and email.

tcpdump A common command-line packet analyzer that allows a user to intercept and display network traffic.

UDP User Datagram Protocol: A connectionless transport protocol used where speed matters more than guaranteed delivery, such as voice calls and name lookups.

Unallocated Space The portion of a storage drive that is not currently allocated to any active file by the file system. It may contain data from previously deleted files.

VoIP Voice over IP: Telephone calls carried as packets over an IP network.

Volatility A term describing how long data is expected to last. Volatile data is temporary and is lost when a device loses power.

Wireshark A popular and powerful network protocol analyzer with a graphical user interface, used for packet capture and deep inspection of network traffic.

Write-blocker A device, either hardware or software, that prevents any writes or modifications from being made to a storage device, ensuring the integrity of the original evidence during a forensic acquisition.



Thank You!

