Android Forensics

Simson L. Garfinkel, Ph.D

Associate Professor, Naval Postgraduate School

http://www.simson.net/

NPS is the Navy's Research University.

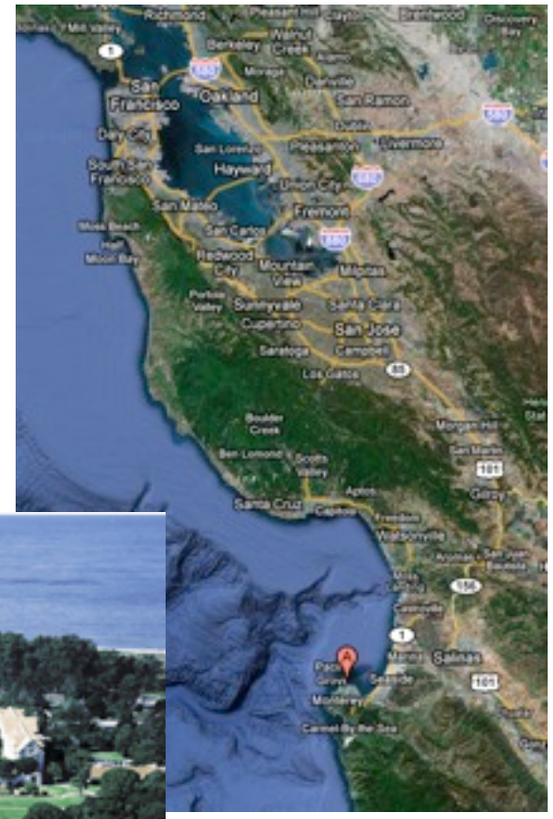
Location: Monterey, CA

Students: 1500

- US Military (All 5 services)
- US Civilian (Scholarship for Service & SMART)
- Foreign Military (30 countries)
- All students are fully funded

Schools:

- Business & Public Policy
- Engineering & Applied Sciences
- Operational & Information Sciences
- International Graduate Studies



NCR Initiative:

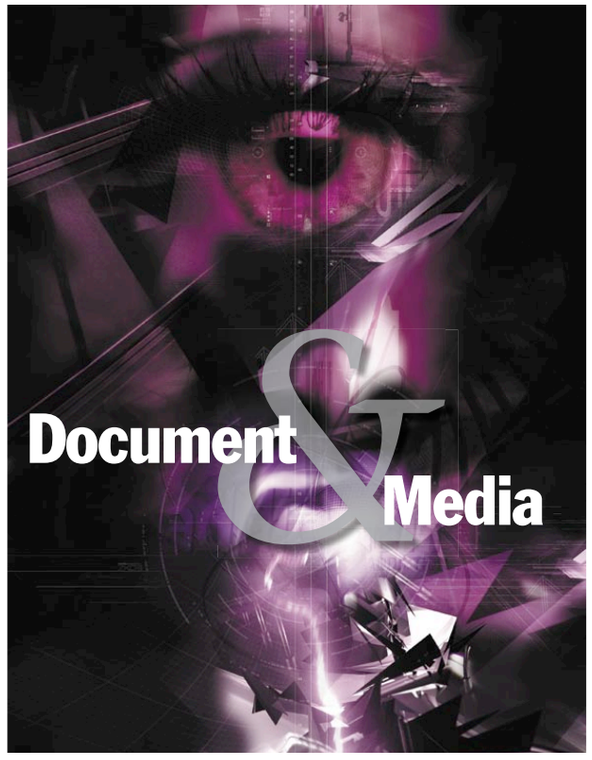
- 8 offices on 5th floor, 900N Glebe Road, Arlington
- FY12 plans: 4 professors, 2 postdocs, 2 researchers
- **Immediate slots for .gov PhDs!**

My current research: Automated Document & Media Exploitation

The DOMEX challenge is to turn digital bits into actionable intelligence.

Recent publications:

- DoD Risks from Facebook
- Ascription of Carved Data on multi-user systems
- Forensic Carving of Network Packets and Associated Data Structures
- Finding and Archiving the Internet Footprint



A computer used by Al Qaeda ends up in the hands of a *Wall Street Journal* reporter. A laptop from Iran is discovered that contains details of that country's nuclear weapons program. Photographs and videos are downloaded from terrorist Web sites.

As evidenced by these and countless other cases, digital documents and storage devices hold the key to many ongoing military and criminal investigations. The most straightforward approach to using these media and documents is to explore them with ordinary tools—open the word files with Microsoft Word, view the Web pages with Internet Explorer, and so on.

Although this straightforward approach is easy to understand, it can miss a lot. Deleted and invisible files can be made visible using basic forensic tools. Programs called *carvers* can locate information that isn't even a complete file and turn it into a form that can be readily processed. Detailed examination of e-mail headers and log files can reveal where a computer was used and other computers with which it came into contact. Linguistic tools can discover multiple documents that refer to the same individuals, even though names in the different documents have different spellings and are in different human languages. Data-mining techniques such as cross-drive analysis can reconstruct social networks—automatically determining, for example, if the computer's previous user was in contact with known terrorists. This sort of advanced analysis is the stuff of DOMEX, the little-known intelligence practice of document and media exploitation.

The U.S. intelligence community defines DOMEX as "the processing, translation, analysis, and dissemination

Exploitation

SIMSON L. GARFINKEL, PH.D.

The DOMEX challenge is to turn digital bits into actionable intelligence.

http://www.simson.net/clips/academic/2007.ACM.Domex.pdf

I also maintain the Forensics Wiki: http://www.forensicswiki.org/

http://www.forensicswiki.org/

Log in / create account

page | discussion | view source | history

Main Page

This is the **Forensics Wiki**, a Creative Commons -licensed wiki devoted to information about digital forensics (also known as computer forensics). We currently list a total of 498 pages.

Much of computer forensics is focused on the tools and techniques used by investigators, but there are also a number of important papers, people, and organizations involved. Many of those organizations sponsor conferences throughout the year and around the world. You may also wish to examine the popular journals and some special reports.

navigation:
- Main Page
- Categories

about forensicswiki.org:
- Recent changes
- Random page
- Donations

search
Go Search

toolbox
- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

Selected Forensics Research

2008-Aug-13

Lest We Remember: Cold Boot Attacks on Encryption Keys

J. Alex Halderman, Princeton University; Seth D. Schoen, Electronic Frontier Foundation; Nadia Heninger and William Clarkson, Princeton University; William Paul, Wind River Systems; Joseph A. Calandrino and Ariel J. Feldman, Princeton University; Jacob Appelbaum; Edward W. Felten, Princeton University

USENIX Security '08 Refereed Paper

Awarded Best Student Paper

Increasingly memory analysis is of interest in forensic research---both because new malware only resides in memory, and because memory analysis is frequently the only way for analysts to get the keys that are used to protect cryptographic file systems. In this paper the authors show that cryptographic keys in memory are vulnerable to exploitation *after the computer is turned off.* The authors show that the contents of dynamic RAM are retained seconds, and sometimes

This is an introductory tutorial!

1:30 - 3:45 Forensics Overview & Android Forensics

- Forensics and digital Investigations (overview)
- Challenges facing computer forensics today
- Cell phone forensics (overview)
- Forensics Targets in Android Phones
  - —*Allocated vs. Residual Data*
  - —*Logical vs. Physical*
- FAT32 & YAFFS analysis
- Clearing, Sanitization and Anti-forensics
- Feature Extraction, Cross-Drive Analysis, and Research Opportunities

3:45 - 4:00 Break

4:00 - 5:00 Android System Analysis

- Extracting the data
- Open Source and Commercial Tools
- Q&A

Forensics & digital investigations

Forensic definitions
The “magic camera”
Hypothesis-based investigation

“Forensics” has two meanings.

fo·ren·sics n. (used with a sing. verb)

The art or study of formal debate; argumentation.

The use of science and technology to investigate and establish facts in criminal or civil courts of law.

(American Heritage Dictionary, 4th Edition)

Courts settle disputes, redress grievances, and mete out punishment

Deciding some disputes requires the use of physical evidence:

- Fingerprints
- DNA
- Handwriting
- Polygraph

Judges and juries can't examine physical evidence and make a determination

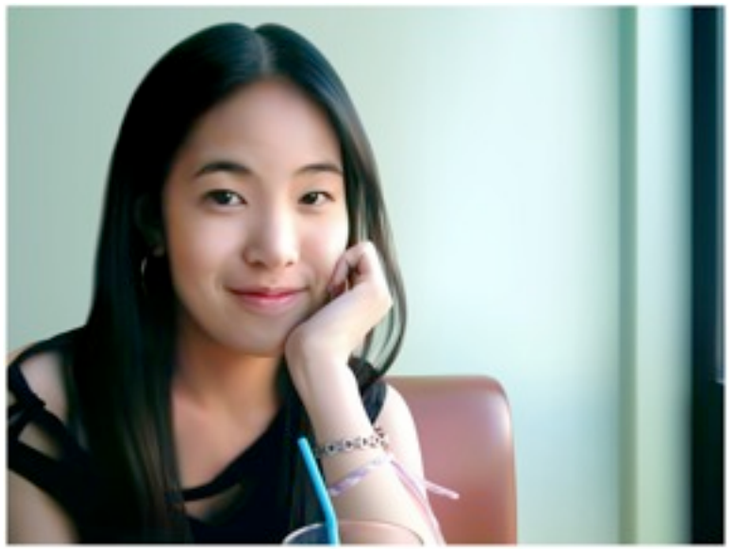
They don't have the expertise

- Evidence may be open to interpretation.

Even photographs may require interpretation

When were these photographs taken?

Were they faked?



The Commissar Vanishes documents how Stalin's Soviet Union tampered with the past.

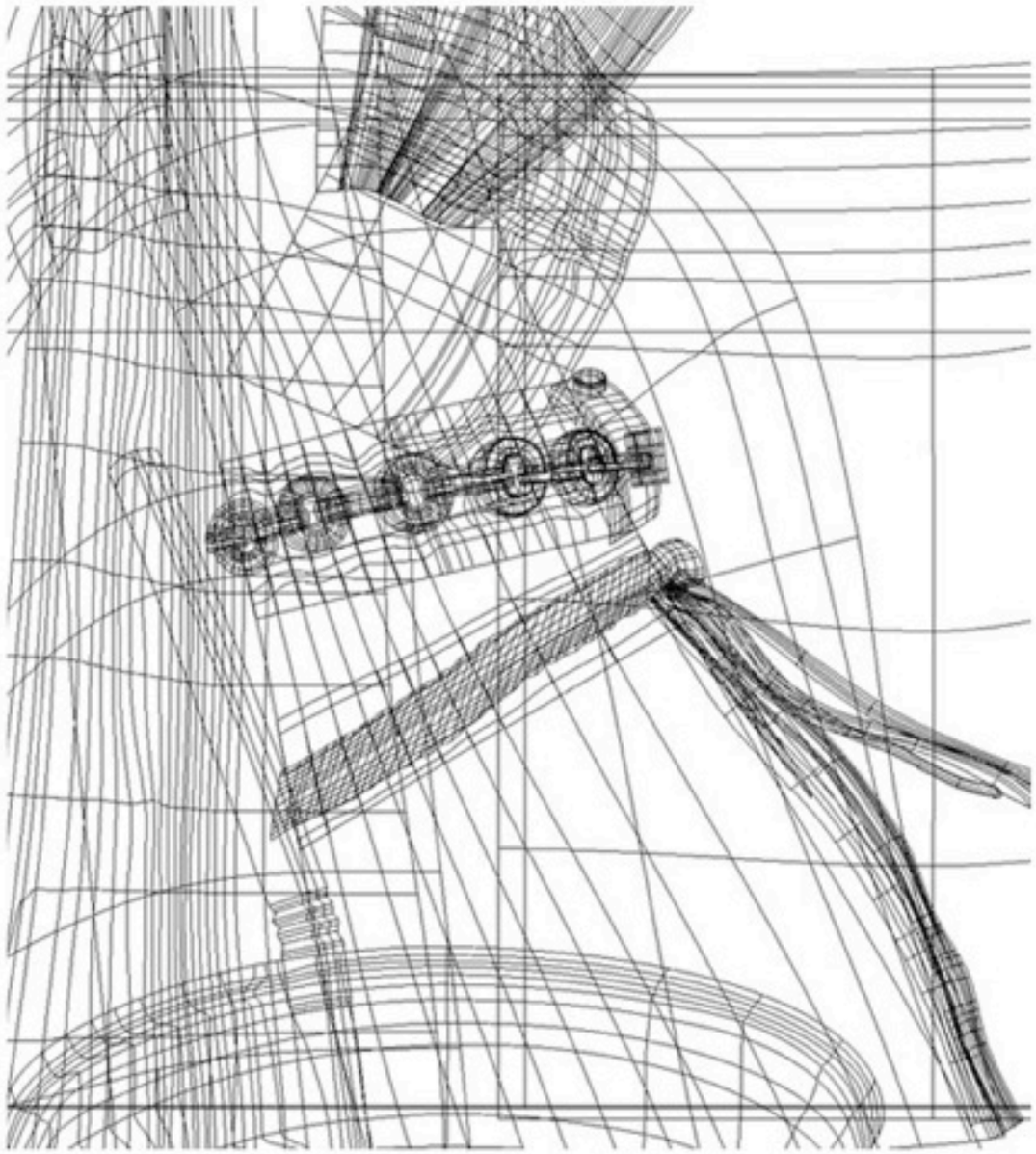
Abel Yenukidze:

- Shot during the purges of 1936-1938
- Photo removed from official photographs by Stalin’s darkroom

—http://www.hoover.org/publications/digest/3531641.html

—http://www.newseum.org/berlinwall/commissar_vanishes/

Computer graphics are so realistic...
... it is easy to mistake a simulated photo for reality.



Pisan Kaewma 2006

—*Can Digital Photos Be Trusted, Steve Casimiro, 9/11/2005, popsci.com*

—*Seeing is Not Believing, Steve Casimiro, Popular Science, Oct. 2005,*

Digital media means it's easier to create forgeries.

Most photos are not "doctored."

- But most photographs are not taken into court

If the interpretation of a photo is high-stakes...

- ... then *someone* has an interest in the photo being doctored

This is true of all evidence

But it's easier to doctor digital evidence

- "Digital Doctoring: can we trust photographs?"
  - —*Hany Farid,* In *Deception: Methods, Motives, Contexts and Consequences, 2007*
  - —*http://www.cs.dartmouth.edu/farid/*

Published

Source #1

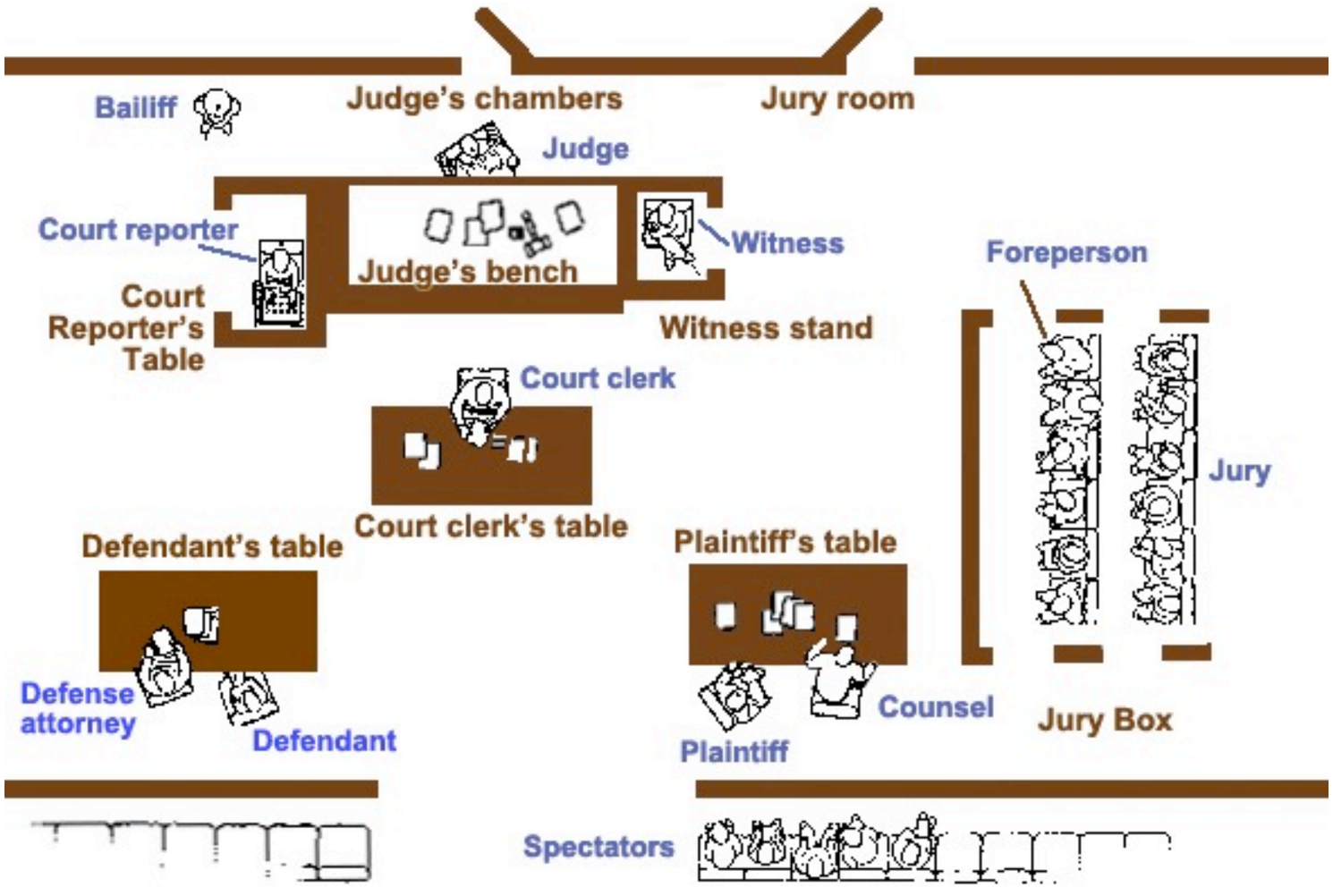
Source #2

Figure 3. The published (top) and original LA Times photographs showing a British soldier and Iraqi civilians.

Forensic experts interpret scientific evidence.

US courts employ an adversarial process.

- Prosecution (or plaintiff) & Defense hires its own experts.
- In some cases, the court may hire a third expert for the judge.



Investigators for the prosecution: conduct the investigation and build the case.

Criminal Digital Investigators:

- Sworn Law Enforcement Officer (usually)
- Writes search warrants
- Receives computers, cameras, and other evidence
- Acquires & analyzes data
- Presents findings
- Prepares report
- Testifies in court

—True for most English-Speaking countries

Investigators for the defense: rebut the evidence and create doubt.

Defense Experts:

- Employed by the Defense
- Works with defense attorney
- Receives evidence from law enforcement
- May conduct independent investigation, but usually funds do not permit
- May work with other experts
- May testify in court

A defense expert can win a case by showing:

- Prosecution experts did not follow their own procedures
- Prosecution experts improperly trained or incompetent
- Chain-of-custody problems
- No scientific basis for employed techniques

Traditional forensics is dominated by the Locard Exchange Principle

Dr. Edmund Locard (1877-1966) - "Every contact leaves a trace."

- "**Wherever he steps**, whatever he touches, whatever he leaves, even unconsciously, **will serve as a silent witness against him**.

Not only his fingerprints or his footprints, but his hair, the fibers from his clothes, the glass he breaks, the tool mark he leaves, the paint he scratches, the blood or semen he deposits or collects.

All of these and more, bear mute witness against him. This is evidence that does not forget. It is not confused by the excitement of the moment.
It is not absent because human witnesses are. It is factual evidence.

Physical evidence cannot be wrong, it cannot perjure itself, it cannot be wholly absent. Only human failure to find it, study and understand it, can diminish its value.

Every contact leaves a trace...

Every contact leaves a trace...

Every contact leaves a trace...

Every contact leaves a trace...

Every contact leaves a trace...

Digital forensics applies these principles to computers.

Some definitions for computer forensics/digital forensics:

- "Involves the preservation, identification, extraction, documentation, and interpretation of computer data."

  —*(Computer Forensics: Incident Response Essentials, Warren Kruse and Jay Heiser.)*

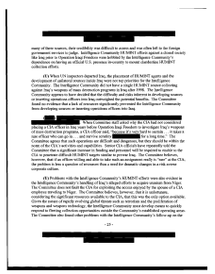
- "The scientific examination, analysis, and/or evaluation of digital evidence in legal matters."

  —*(Scientific Working Group on Digital Evidence, http://www.swgde.org)*

Digital evidence is:

- "Information stored or transmitted in binary form ... relied upon in court." [Int02]
- "Information of probative value ... stored or transmitted in binary form." [Sci05]
- "[D]ata of investigative value ... stored ... or transmitted by a computer." [Ass05]
- "[D]ata ... that support or refute a theory of how an offense occurred or that address critical elements ... such as intent or alibi." [Cas04]

If it involves computers, it's probably digital evidence



Like mater, *data* can be *evidence* or *the crime itself.*

Evidence of a crime:

- Financial records
- Emails documenting a conspiracy
- Photographs of a murder

The crime itself:

- Possession or distribution of obscene photographs
- Child pornography
- Computer break-ins
- Denial-of-service attacks
- Threats sent by email

Like mater, *data* can be *evidence* or *the crime itself.*

Evidence of a crime:

- Financial records
- Emails documenting a conspiracy
- Photographs of a murder

The crime itself:

- Possession or distribution of obscene photographs
- Child pornography
- Computer break-ins
- Denial-of-service attacks
- Threats sent by email

The best digital evidence is proactively collected!

Systems can record and retain:

- Log files — Recording events (syslog aggregation)
- Disk images (Snapshots)
  - —*Guidance Software's EnCase Forensic*
  - —*Access Data's FTK*
- Network packets and packet flows (Network Forensics)
  - —*Network Flight Recorder (NFR)*
  - —*NetIntercept (Niksun)*

Storage is cheap!

- A 1TB drive holds more than a week's worth of a consumer broadband traffic (@ 100%)

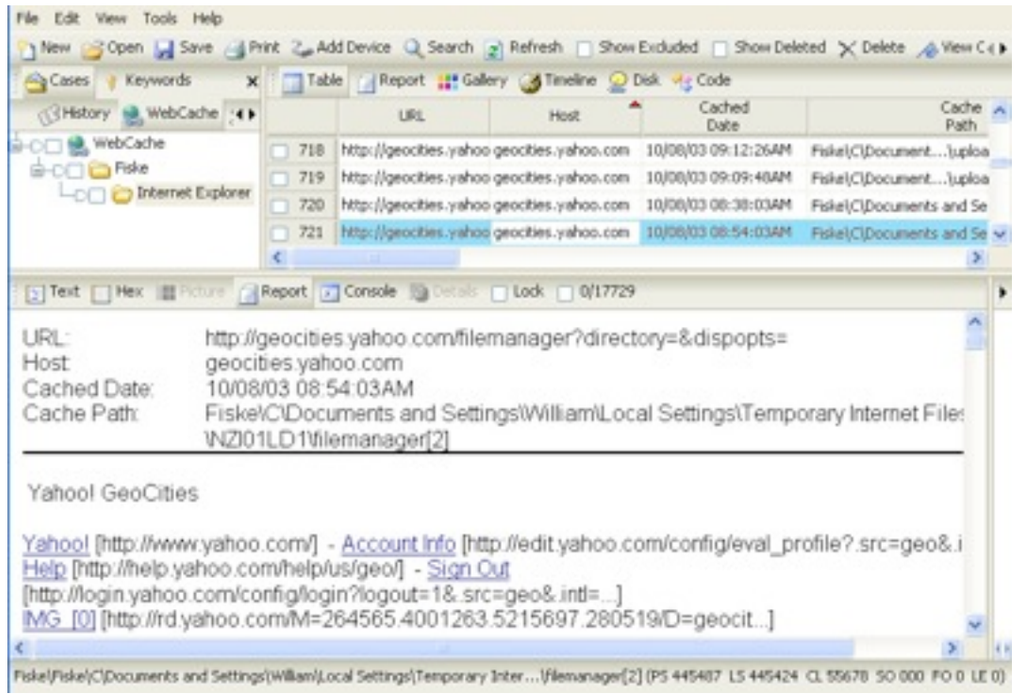
Proactive evidence allows investigators to discover:

- How a crime was committed
- Extent of damage / Presence of illegal activity
- Confirm/disprove an alibi

Digital Forensics lets investigators go back in time... (sort of.)

A magic camera that can:

- View previous versions of files
- Recover “deleted” files
- Find out what was typed
- Report websites visited in the past



Why does this work?

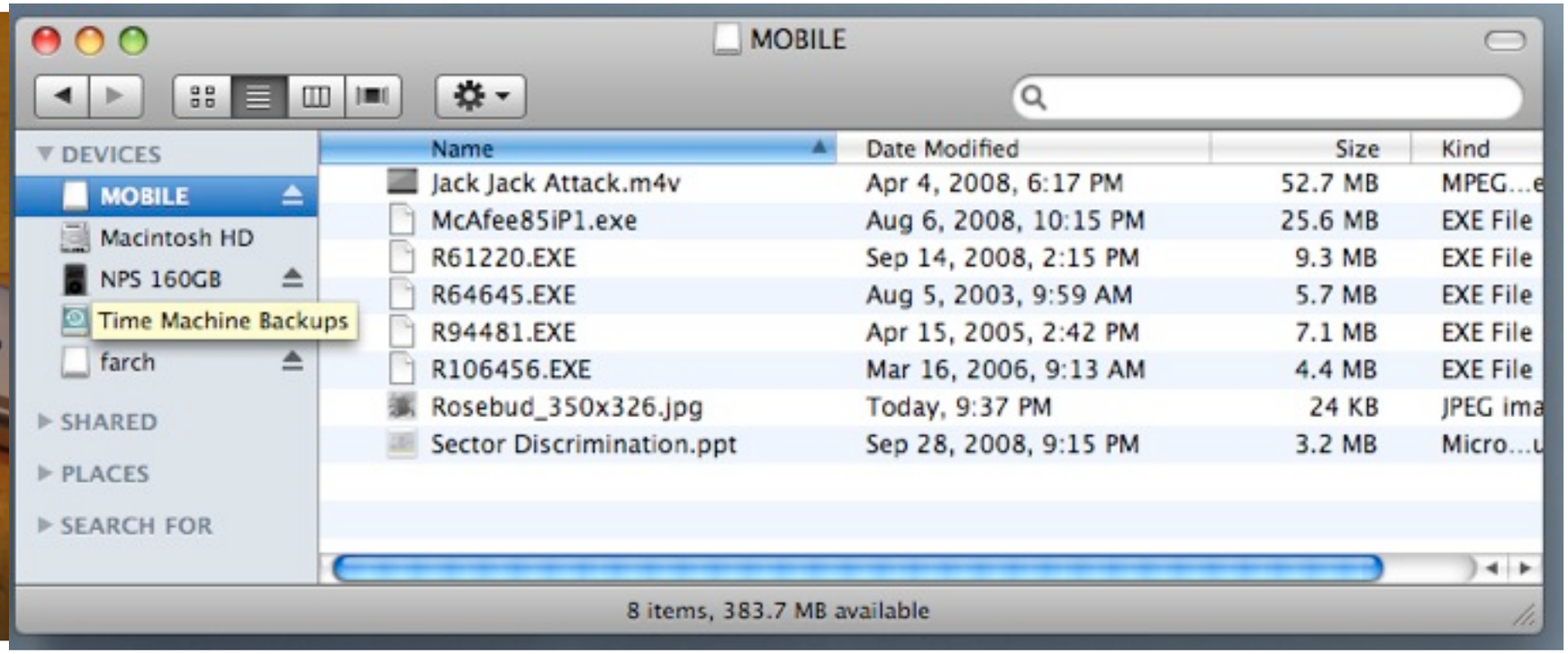
- Cache files and keep extensive logs
- Web pages; wi-fi router logs
- Residual Data — Programmers rarely clear memory when they are finished with it
- free() doesn’t erase memory
  - —*/bin/rm doesn’t overwrite sectors*
  - —*newfs and FORMAT* don't clear disks*
- Most data is not encrypted

But digital evidence is easily faked!

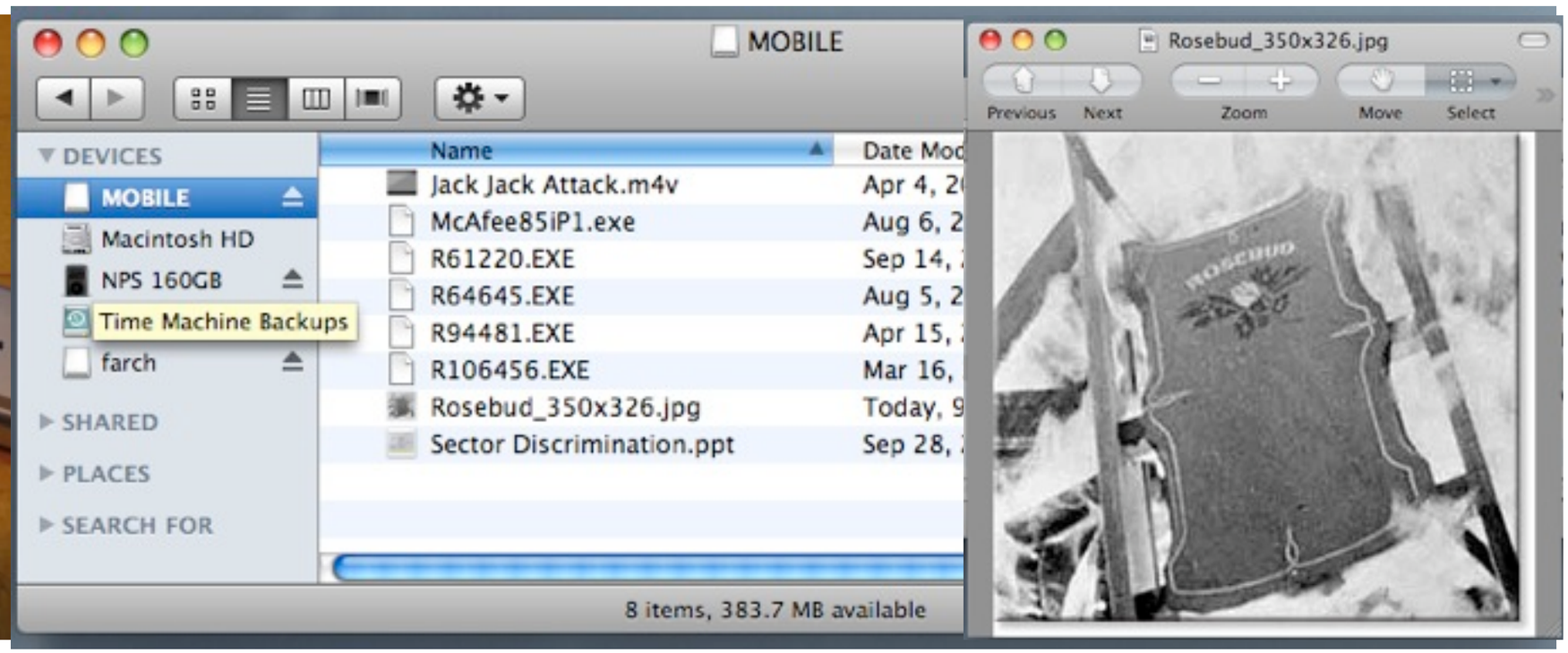
The Exchange Principle doesn't seem to apply to electronic media.

The Exchange Principle doesn't seem to apply to electronic media.

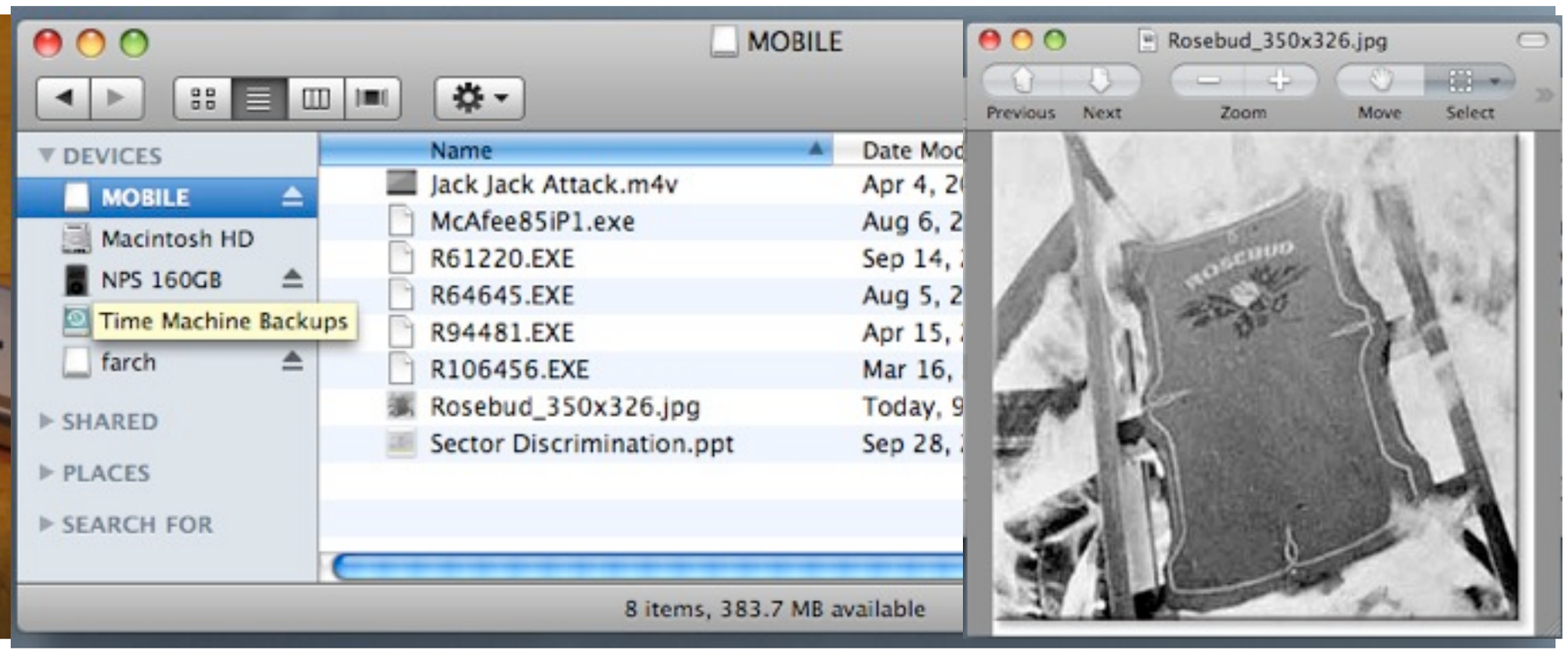
The Exchange Principle doesn't seem to apply to electronic media.

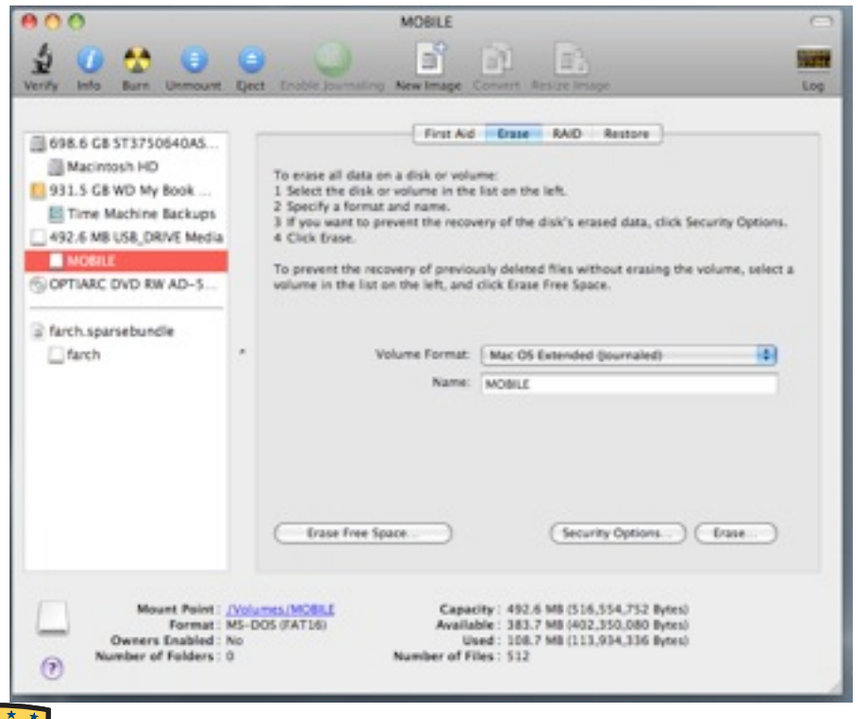


The Exchange Principle doesn't seem to apply to electronic media.

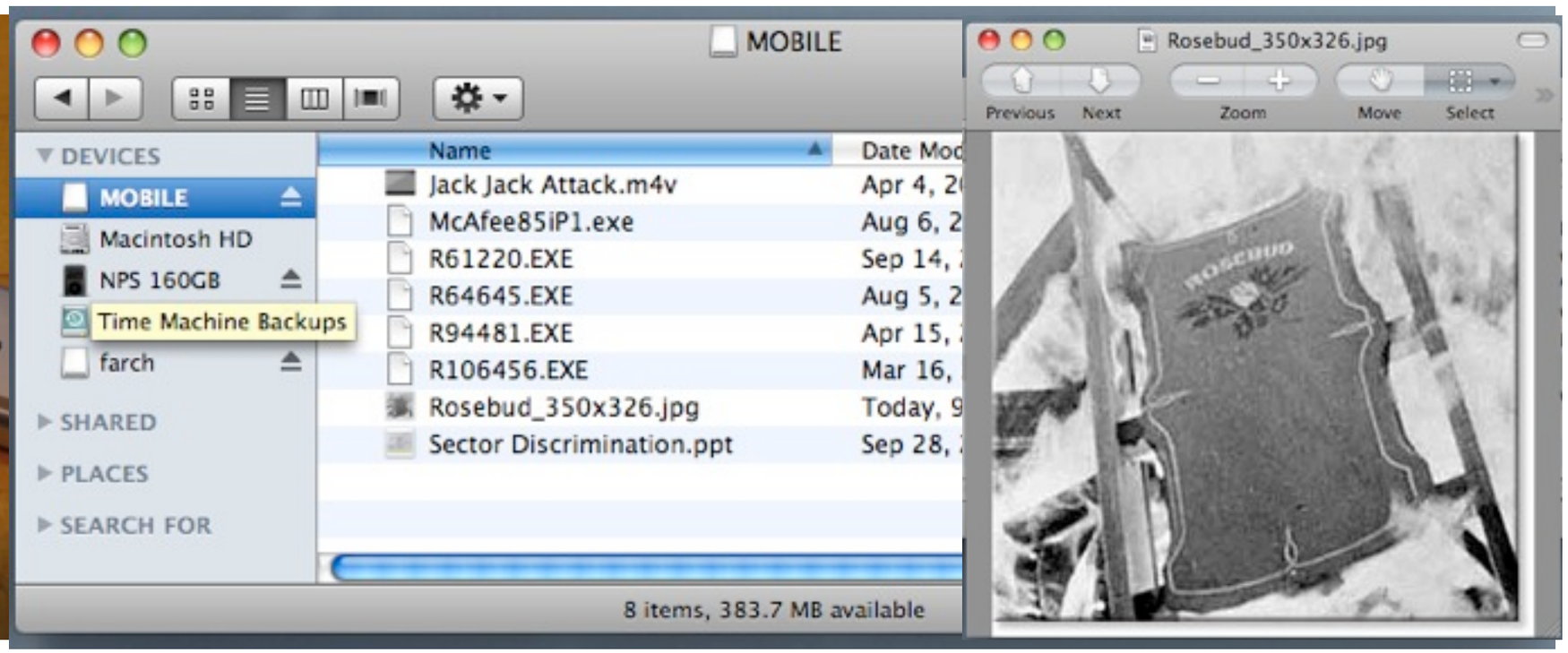


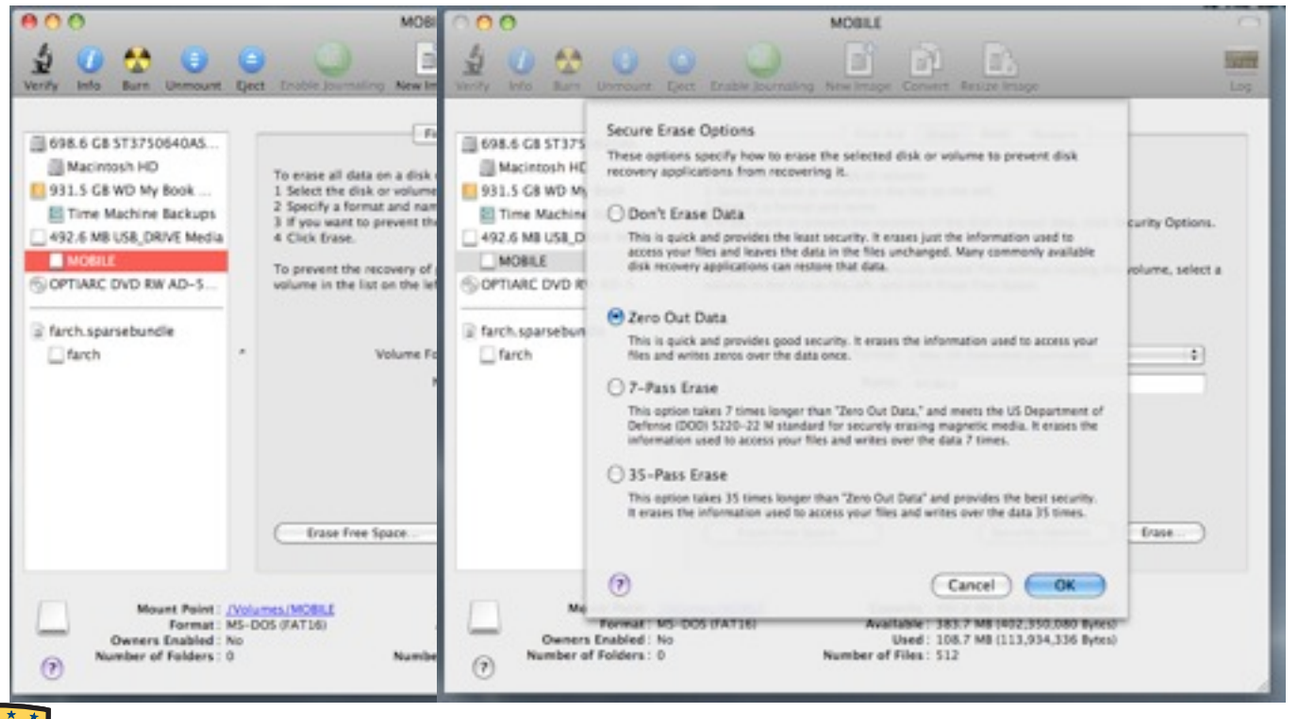
The Exchange Principle doesn't seem to apply to electronic media.



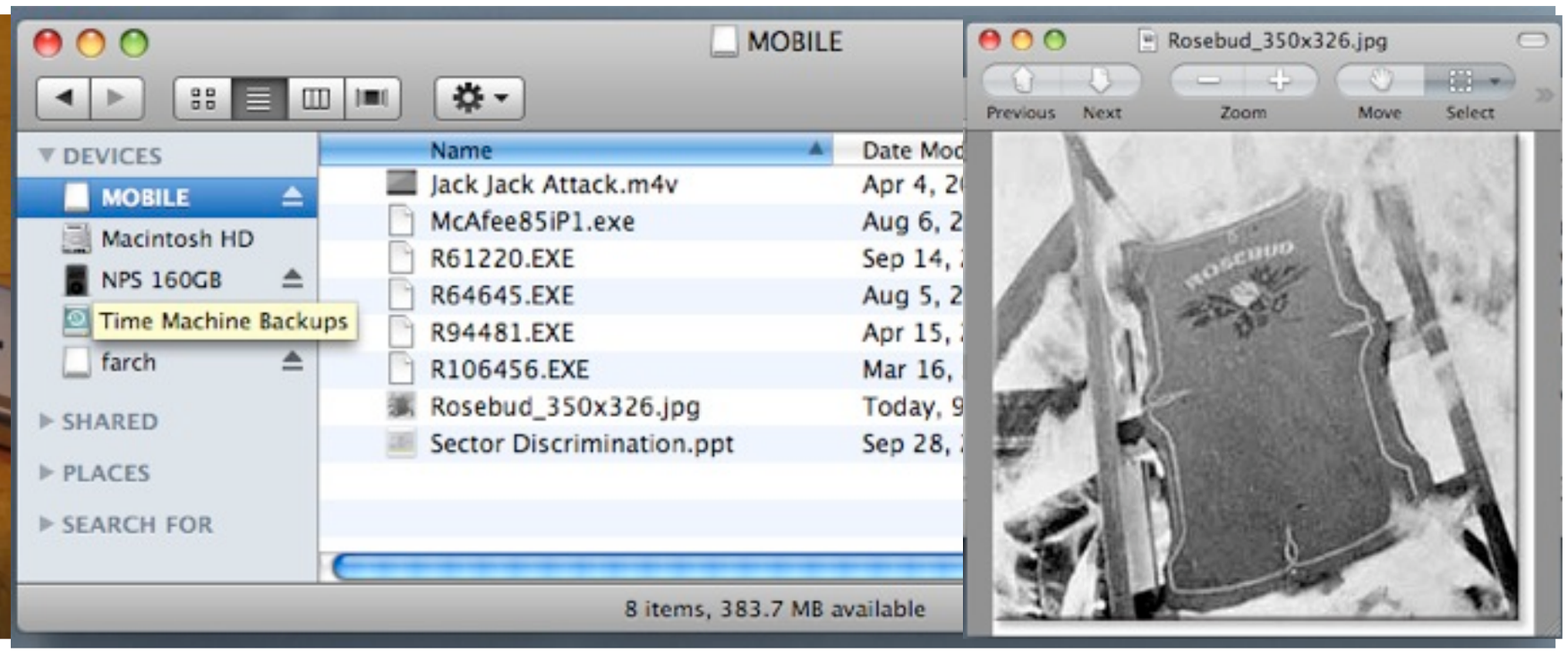


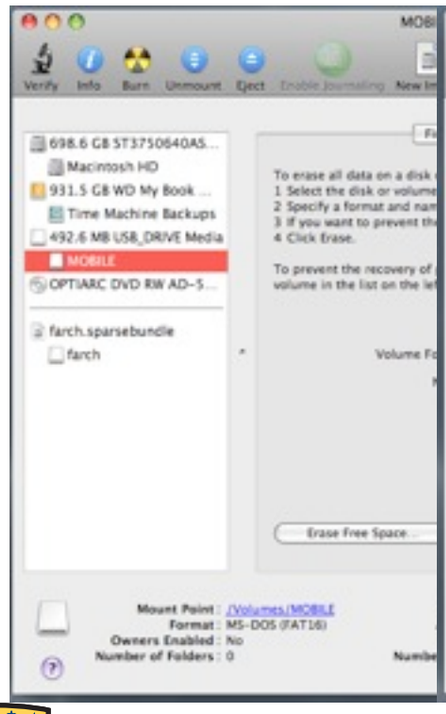
The Exchange Principle doesn't seem to apply to electronic media.

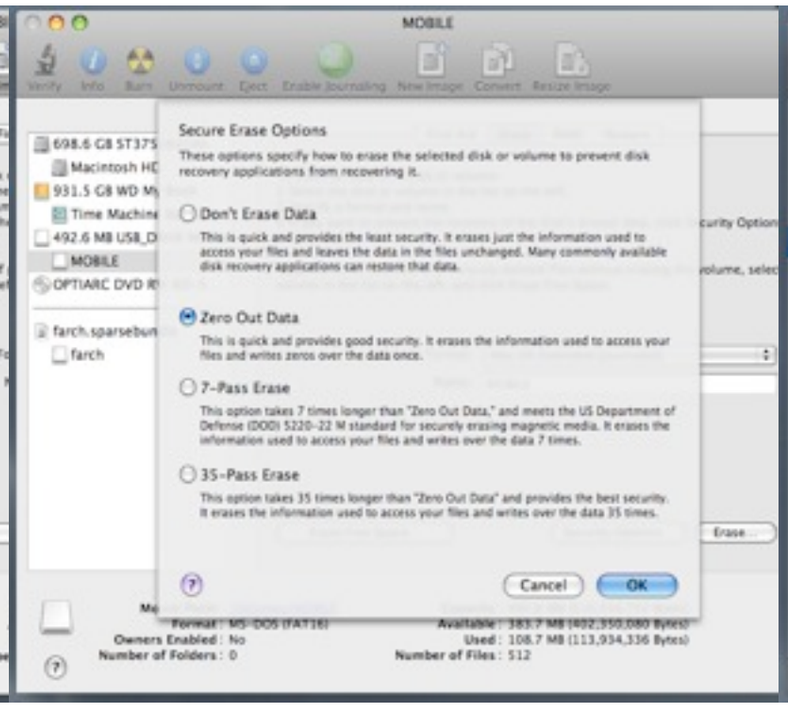


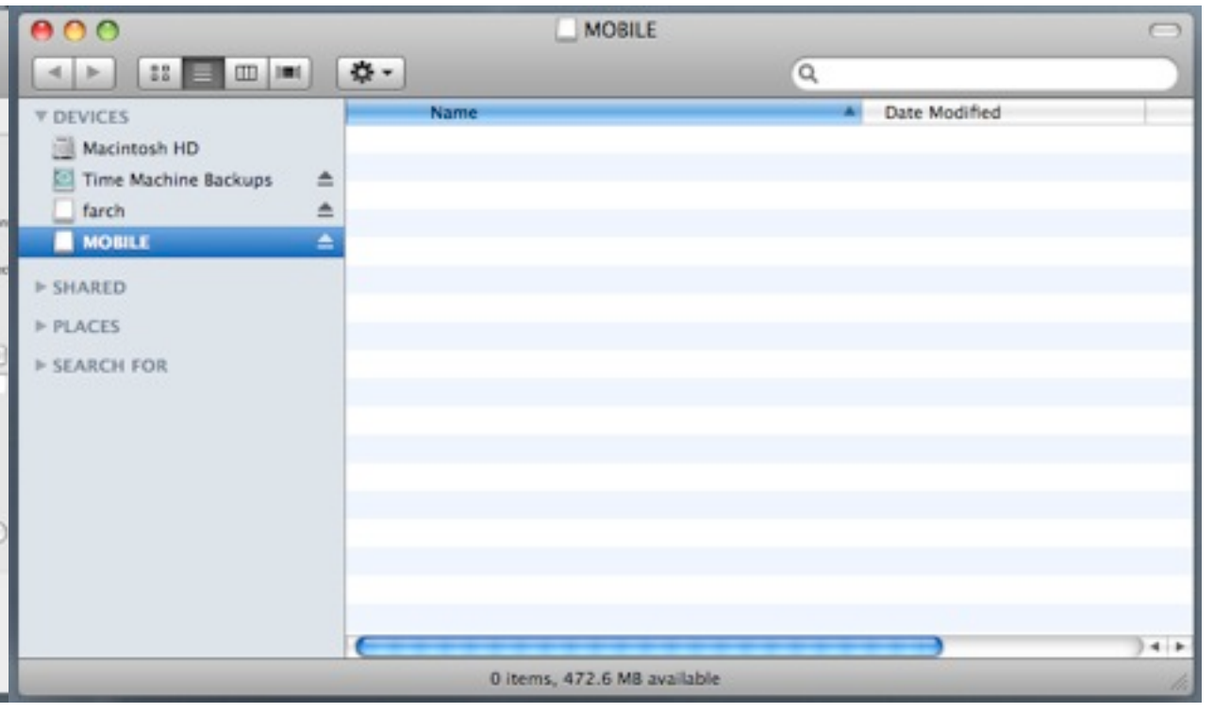


The Exchange Principle doesn't seem to apply to electronic media.





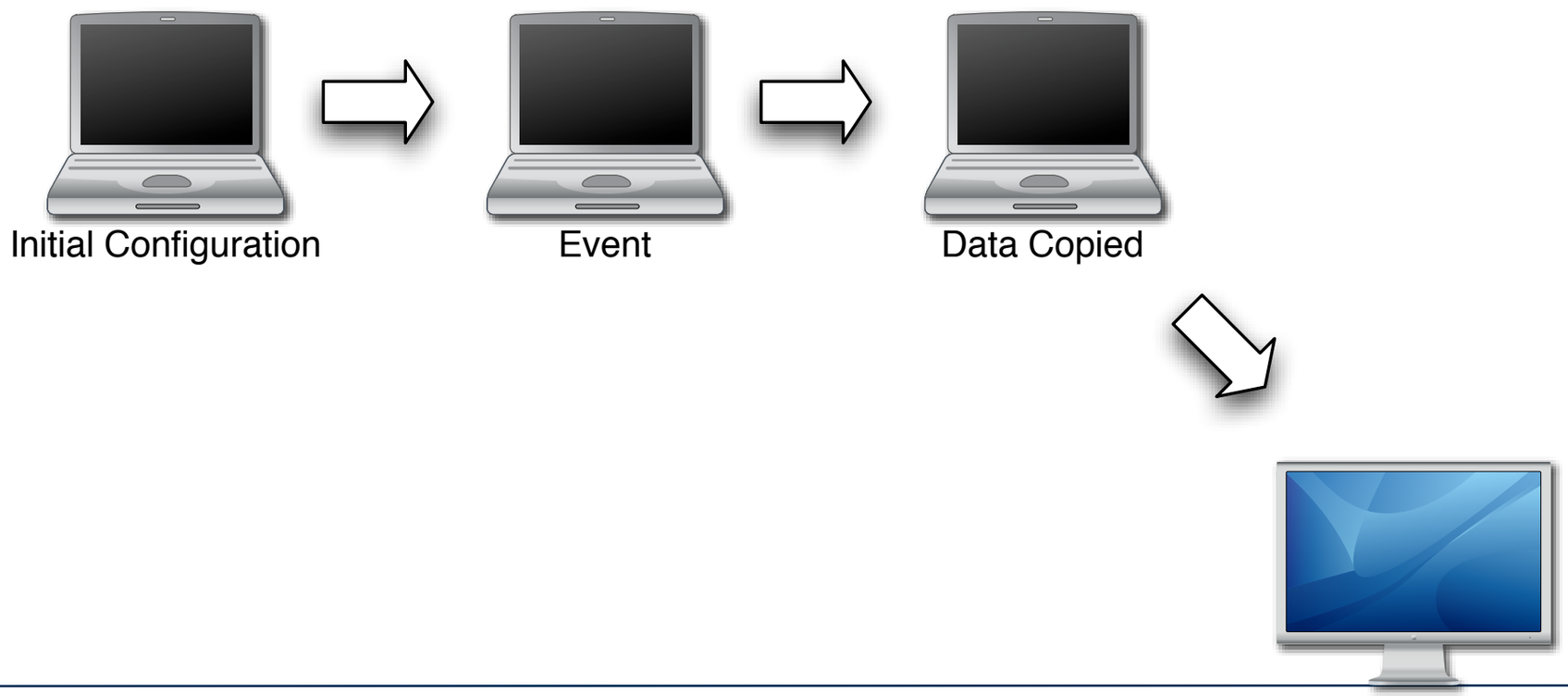




When we look at a computer system, we build a *hypothesis* about the computer's past.

The hypothesis makes assumptions about:

- The system under investigation:
  - —*hardware (stock hardware? modified? firmware?)*
  - —*software (stock? custom? patch level?)*
- The flow of time
- The movement of evidence
- The system being used to investigate the data



But any piece of digital evidence can be explained by *multiple explanations.*

Consider this printout:

```
07:16 AM Black:~/Downloads$ ls -l
07:17 AM Black:~/Downloads$ ls -l
total 74
-rw-r--r--   1 simsong  simsong  73625 Jun 16 06:30 afyi.pdf
07:18 AM Black:~/afyi$
```

When was afyi.pdf downloaded?

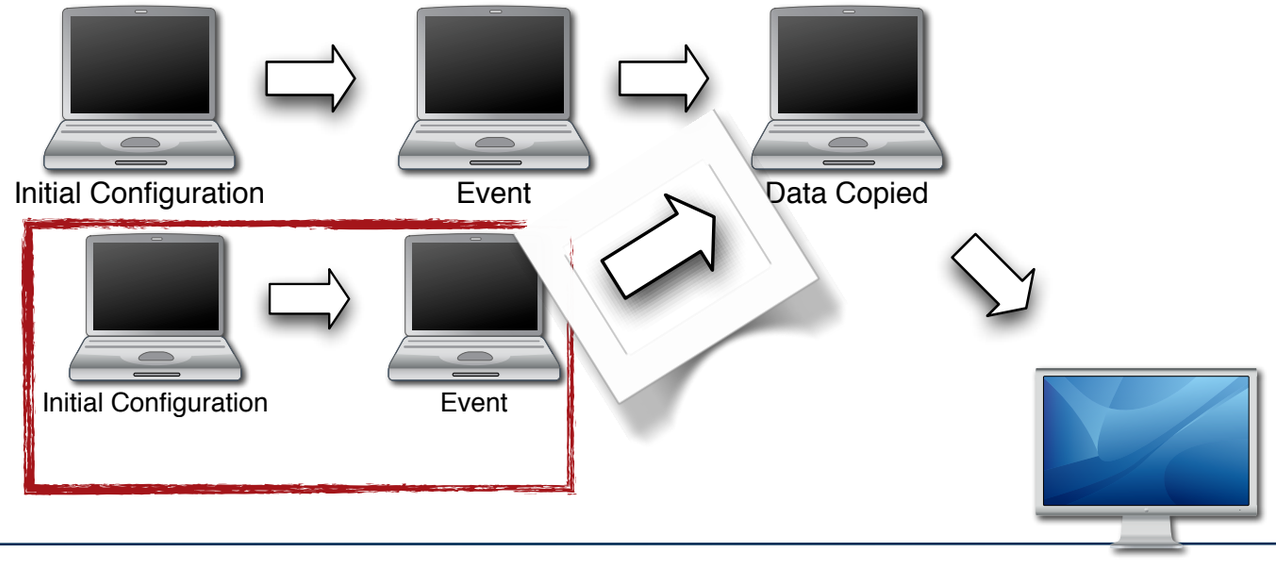
- Possible explanations:
  - —*At 7:17 AM, but Safari set the timestamp to be the time on the server*
  - —*At 6:30AM on a different day; the file was moved into the directory.*
  - —*The computer's clock was changed before the file was downloaded*
  - —*The whole example was faked.*

The most likely explanation may not be correct one.

Likely assumptions are usually correct... but sometimes they are not.

We assume:

- The event didn't fake the initial configuration
- Vulnerabilities we find were used by the attacker.
  - —*The attacker could have created a new vulnerability to hide what was actually used*
- We can can copy all of the computer's data
  - —*We can't get stuff out of L2 cache, some firmware, coprocessor, etc.*
- Our forensic tools are reliable
  - —*The attack might be invisible due to a bug in the forensic tool*
    - A Hypothesis-Based Approach to Digital Forensic Investigations, Brian D. Carrier, PhD. Thesis, Purdue University, 2006



In court, testimony is governed by the Federal Rules of Evidence

Article I. General Provisions

Article II. Judicial Notice

Article III. Presumptions In Civil Actions And Proceedings

Article IV. Relevancy And Its Limits

Article V. Privileges

Article VI. Witnesses

Article VII. Opinions and Expert Testimony

Article VIII. Hearsay

Article IX. Authentication and Identification

Article X. Contents of Writings, Records and

Article XI. Miscellaneous Rules

US Federal Rules of Evidence
Article VII regulates the testimony of “experts”

Rule 702. Testimony by Experts

Rule 703. Bases of Opinion Testimony by Experts

Rule 704. Opinion on Ultimate Issue

Rule 705. Disclosure of Facts or Data Underlying Expert Opinion

Rule 706. Court Appointed Experts

These rules apply in the Federal Court; many states follow the rules as well

- http://www.law.cornell.edu/rules/fre/

Rule 702. Testimony by Experts

"If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if

- the testimony is based upon sufficient facts or data,
- the testimony is the product of reliable principles and methods, and
- the witness has applied the principles and methods reliably to the facts of the case."

Rule 702. Testimony by Experts

"If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if

- the testimony is based upon sufficient facts or data,
- the testimony is the product of reliable principles and methods, and
- the witness has applied the principles and methods reliably to the facts of the case."

Note: specify your expert domain explicitly.
If it's too general, your expert status may be challenged.

Rule 703. Bases of Opinion Testimony by Experts

“The facts or data in the particular case upon which an expert bases an opinion or inference may be those perceived by or made known to the expert at or before the hearing.

If of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject, the facts or data need not be admissible in evidence in order for the opinion or inference to be admitted.

Facts or data that are otherwise inadmissible shall not be disclosed to the jury by the proponent of the opinion or inference unless the court determines that their probative value in assisting the jury to evaluate the expert's opinion substantially outweighs their prejudicial effect.”

Rule 703. Bases of Opinion Testimony by Experts

“The facts or data in the particular case upon which an expert bases an opinion or inference may be those perceived by or made known to the expert at or before the hearing.

If of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject, the facts or data need not be admissible in evidence in order for the opinion or inference to be admitted.

Facts or data that are otherwise inadmissible shall not be disclosed to the jury by the proponent of the opinion or inference unless the court determines that their probative value in assisting the jury to evaluate the expert's opinion substantially outweighs their prejudicial effect.”

This means that experts can rely on hearsay data, provided that it is supported by technical information.

Rule 704. Opinion on Ultimate Issue

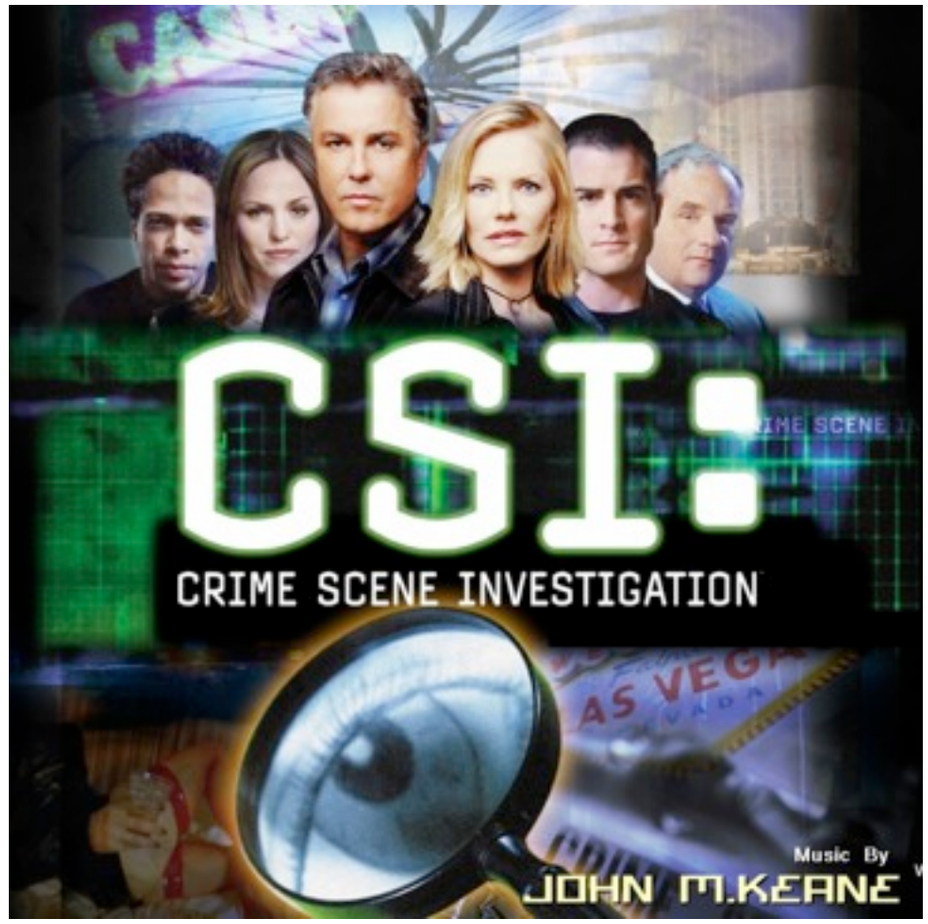
“(a) Except as provided in subdivision (b), testimony in the form of an opinion or inference otherwise admissible is not objectionable because it embraces an ultimate issue to be decided by the trier of fact

“(b) No expert witness testifying with respect to the mental state or condition of a defendant in a criminal case may state an opinion or inference as to whether the defendant did or did not have the mental state or condition constituting an element of the crime charged or of a defense thereto. Such ultimate issues are matters for the trier of fact alone.”

The “CSI Effect” causes victims and juries to have unrealistic expectations.

On TV:

- Forensics is swift
- Forensics is certain
- Human memory is reliable
- Presentations are highly produced



TV digital forensics:

- Every investigator is trained on every tool
- Correlation is easy and instantaneous
- There are no false positives
- Overwritten data can be recovered
- Encrypted data can usually be cracked
- It is impossible to delete anything

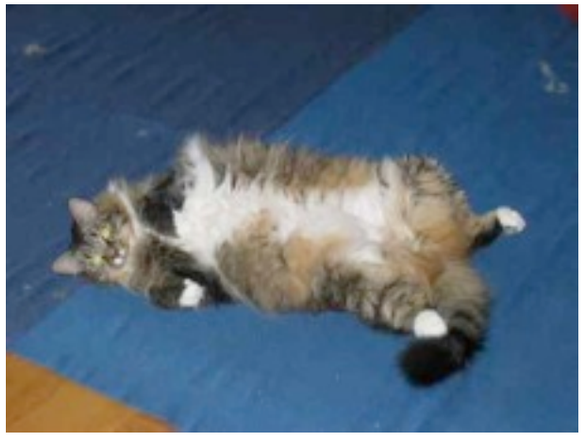
The reality of digital forensics is less exciting.

Every investigation is beset by problems:

- Data that is overwritten cannot be recovered
- Encrypted data usually can't be decrypted
- Forensics rarely answers questions or establishes guilt
- Forensics rarely provides specific information about a specific subject
- Tools crash a lot

Traditionally this hasn’t mattered, because:

- Most digital forensics was used to find child pornography
- When the pornography is found, most suspects plead guilty



Forensics has many uses beyond the courtroom.

Data Recovery:

- Recover deleted files
- Recover data from physically damaged media

Testing and Evaluating:

- System Performance
- Privacy Properties & Tools
- Security Policies

Spot-check regulatory compliance:

- Internal information flows
- Data flow across network boundaries
- Disposal policies

Performance Evaluation

Conclusion:
Forensics and Digital Investigations

Scientific evidence requires interpretation to get it into a court room:

- You give a disk image to a jury

Digital evidence is easy to fake

- You can completely wipe a computer or restore it's hard drive
- You just can't do that with a physical crime scene

Main uses today of digital forensics:

- Finding child pornography
- Recovering deleted files

Open Question:
Does Locard's Exchange Principle apply to flash media?

Magnetic media and flash media are *fundamentally different*

Magnetic Media

—*Sectors overwritten in place*

—*Each sector can be rewritten billions of times*

—*Designed around magnetic remanence*

- Flash

—*Sectors overwritten in different locations*

—*Each sector can be rewritten thousands of times*

—*Flash Translation Layer (FTL) maps logical-to-physical sectors*

Today most forensic tools treat flash as magnetic media

But perhaps we can do more by understanding the flash layer.

Today's Digital Forensics Crisis

Today there is a growing digital forensics crisis.

Much of the last decade's progress is quickly becoming irrelevant

Tools designed to let an analyst find a file and take it into court

… don't scale to today's problems

We have identified 5 key problems.

Problem 1 - Increased cost of extraction & analysis.

Data: too much and too complex!

- Increased size of storage systems

- Cases now require analyzing multiple devices

  —*2 desktops, 6 phones, 4 iPods, 2 digital cameras = 1 case*

- Non-removable flash

  —*It's hard to physically get to the data*

- Proliferation of operating systems, file formats and connectors

  —*XFAT, XFS, ZFS, YAFFS2, Symbian, Pre, iOS,*

Consider FBI Regional Computer Forensic Laboratories growth:

- Service Requests: 5,057 (FY08) ➔ 5,616 (FY09) (+11%)
- Terabytes Processed: 1,756 (FY08) ➔ 2,334 (FY09) (+32%)

Problem 2 — RAM and malware forensics is really hard.

RAM Forensics—in its infancy

- RAM structures change frequently (no reason for consistency)
- RAM is constantly changing

Malware is especially hard to analyze:

- Encryption; Conditional execution
- Proper behavior of most software is not specified

Malware can hide in many places:

- On disk (in programs, data, or scratch space)
- BIOS & Firmware
- RAID controllers
- GPU
- Ethernet controller
- Motherboard, South Bridge, etc
- FPGAs

Problem 3 — Mobile phones are really hard to examine.

Cell phones present special challenges

No standard connectors

- No standard way to copy data out
- Difficult to image & store cell phones without changing them.

How do we validate tools against thousands of phones?

- No standardized cables or extraction protocols

NIST's *Guidelines on Cell Phone Forensics* recommends:

- "searching Internet sites for developer, hacker, and security exploit information."

How do we forensically analyze 100,000 apps?

Problem 4 — Encryption and Cloud Computing make it hard to get to the data

Pervasive Encryption — Encryption is increasingly present

TrueCrypt

- BitLocker
- File Vault
- DRM Technology

Cloud Computing — End-user systems won't have the data

Google Apps

- Microsoft Office 2010
- Apple Mobile Me

—But they may have residual data!

Problem 5 — Time is of the essence.

Most tools were designed to perform a complete analysis

Find all the files

- Index all the terms
- Report on all the data
- Take as long as necessary!

Increasingly we are racing the clock:

- Police prioritize based on statute-of-limitations!
- Battlefield, Intelligence & Cyberspace operations require turnaround in days or hours.

My research focuses on three main areas:

Area #1: Data collection and manufacturing

- Large data sets of real data enable science. (+20TB)
- Small data sets of realistic data enable education, training and publishing. (<1TB)

Area #2: Bringing data mining and machine learning to forensics

- Breakthrough algorithms based on correlation and sampling
- Automated social network analysis (cross-drive analysis)
- Automated ascription of carved data

Area #3: Working above and below the files

- Most work to date is with files
- Digital Forensics XML (DFXML)
  - —*Connecting tools*
  - —*Representing applications, behaviors, users*
  - —*Forensics of bulk data*

Emphasis on *building tools* and *working with practitioners*.

Cell Phone and PDA Forensics (overview)

Who did you call?
Where have you been?
What did you do?

PDAs and Cellphones: Difficult times for computer forensics

Powerful computers

- 100—1.2 GHz processors
- 16MB — 1GB of RAM (or more)
- 500GB — 64GB of Flash
- Networks: Cellular, Bluetooth, WiFi, IR and near field
- Cameras; Sensors

Little standardization:

- Android, iOS, RIM, PalmOS, Windows Mobile, Symbian, RIM, Linux, & others
- Minor revs of a phone may have *radically different data layout*

Technical challenges:

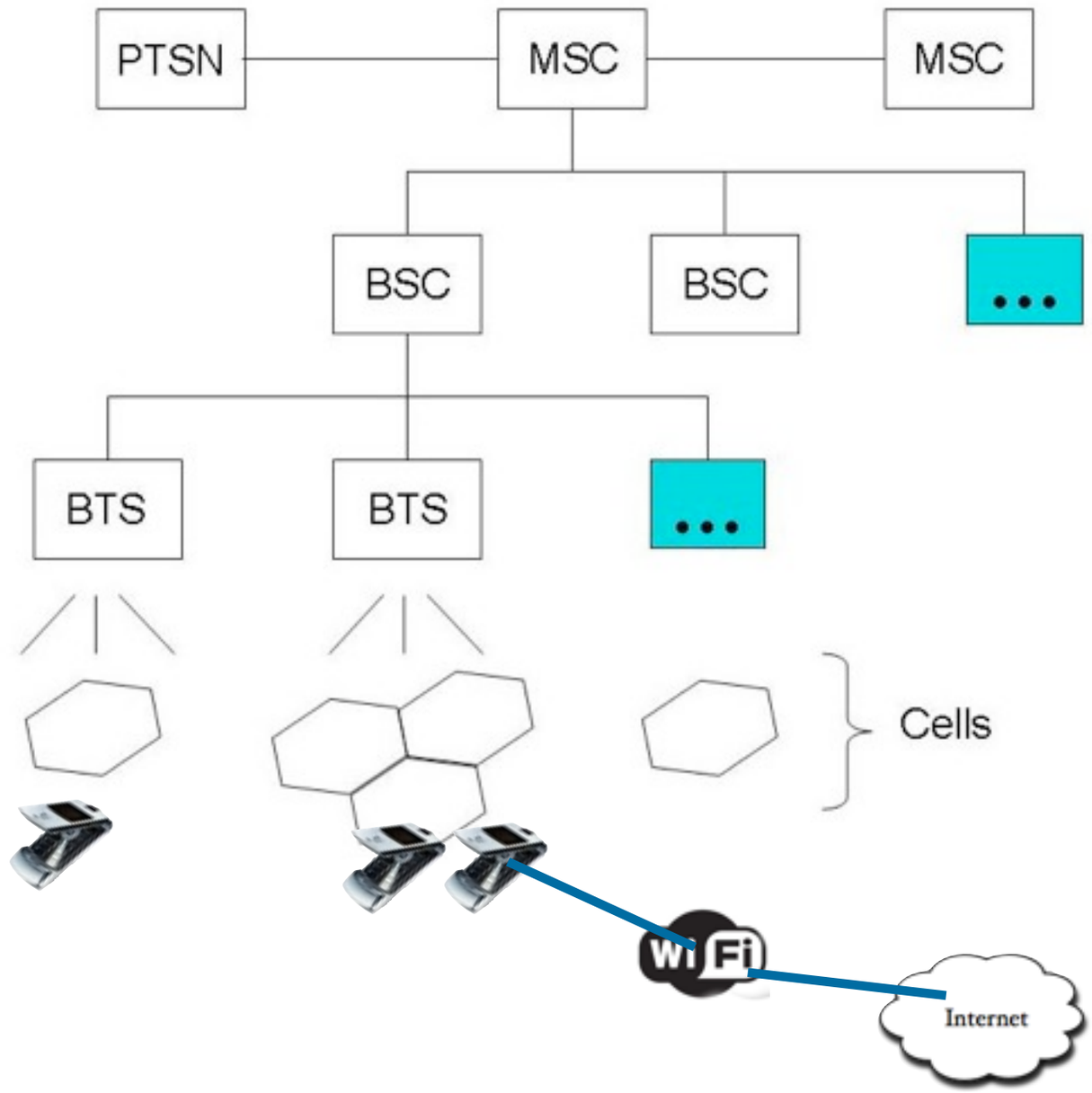
- Old phones are *widely used*
- Downloadable Apps;
- Some systems lose memory w/o power
- Smart phones have 2 processors & multiple memory banks
- Removable media — the phone may not have taken the picture on the SD card

Different phones typically require different tools.

Cell phones are part of a system

Infrastructure:

- Public Telephone Switched Network
- Mobile Switching Center
- Base Station Controller
- Base Transceiver Station
- Cell sites



Customer Provided:

- Cell phones
  - —*Hardware*
  - —*Software*
- Bluetooth
- Wi-Fi networks for data & voice

Administrative:

- Account management and billing

Cell phones are part of a system

Infrastructure:

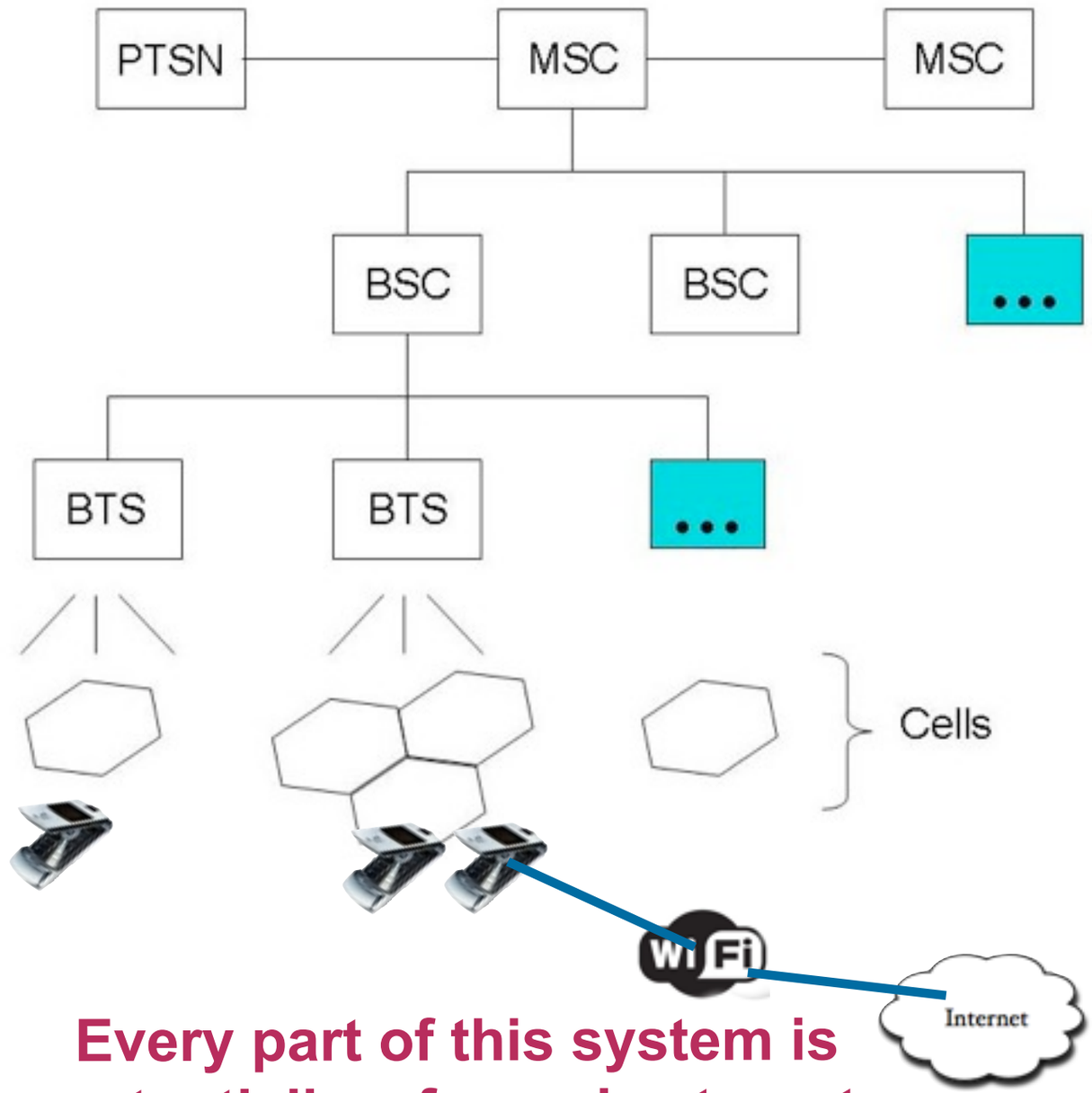
- Public Telephone Switched Network
- Mobile Switching Center
- Base Station Controller
- Base Transceiver Station
- Cell sites

Customer Provided:

- Cell phones
  - —*Hardware*
  - —*Software*
- Bluetooth
- Wi-Fi networks for data & voice

Administrative:

- Account management and billing



Every part of this system is potentially a forensics target

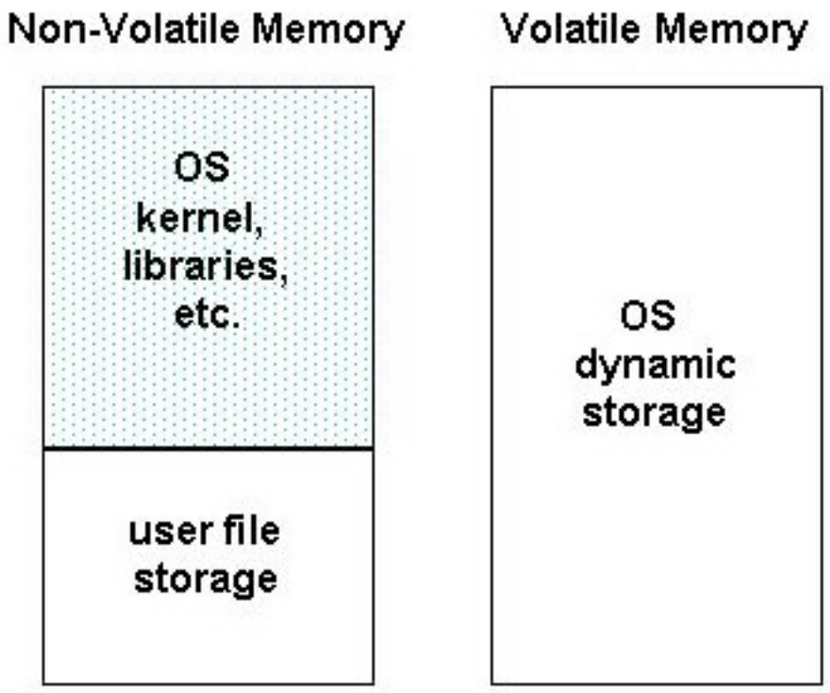
Software of a typical cell phone:

	Basic	Advanced	Smart
OS	Proprietary	Proprietary	Linux, Windows Mobile, Palm OS, Symbian, RIM
PIM	Simple Phonebook	Phonebook and Calendar	Reminder List, Enhanced Phonebook, Calendar
Applications	None	MP3 Player	MP3 Player, Office Document Viewing, &c
Messaging	Text Messaging	Text with Images	Text, Images, Movies
Chat	None	SMS Chat	SMS & Instant Messaging
Email	None	Via Network Operator's Service Gateway	Via POP or IMAP
Web	None	Via WAP Gateway	Direct HTTP
Wireless	IrDA	IrDA, Bluetooth	IrDA, Bluetooth, Wi-Fi

Source: NIST Guidelines on Cell Phone Forensics, p. 9

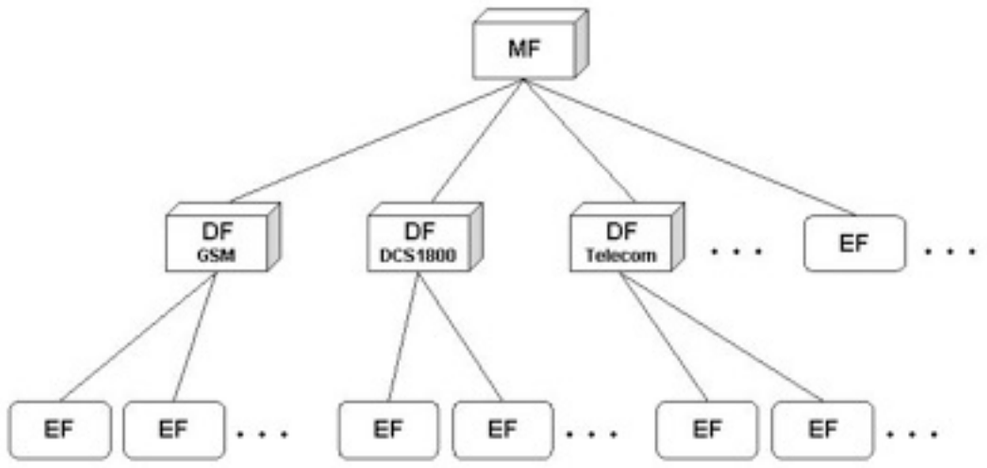
Memory Considerations: mobile phones have *multiple* memory systems.

Soldered-on Storage

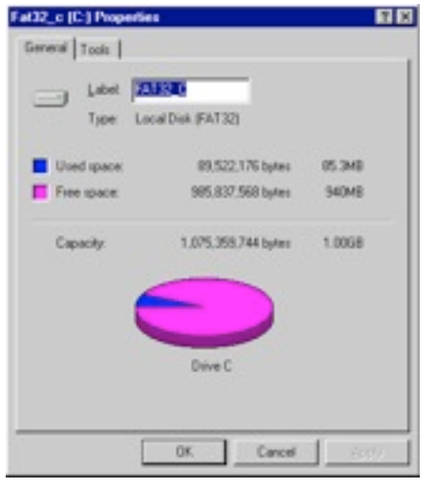


Telephone Service OS

SIM file system

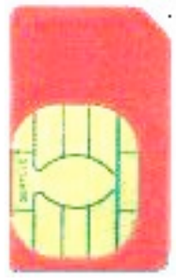


Removable Flash

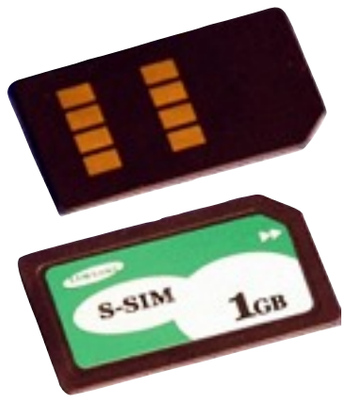


Identity Module Characteristics

GSM:



- Subscriber identity modules (SIMs) 16K-1GB (!)
  - —*Integrated circuit card identifier (ICCID)*
  - —*International mobile subscriber identity (IMSI)*
  - —*Authentication Key K_1*
  - —*Phone book storage*
  - —*Multimedia storage*
  - —*Encrypted storage*
- IMEI - International mobile equipment identifier (*#06#)
  - —*Unique to the GSM device*



CDMA:

- Electronic serial number (ESN)
- Mobile station ID (MSID) or Mobile identification number (MIN)
- May be stored in phone (US) or in a CDMA chip (China)

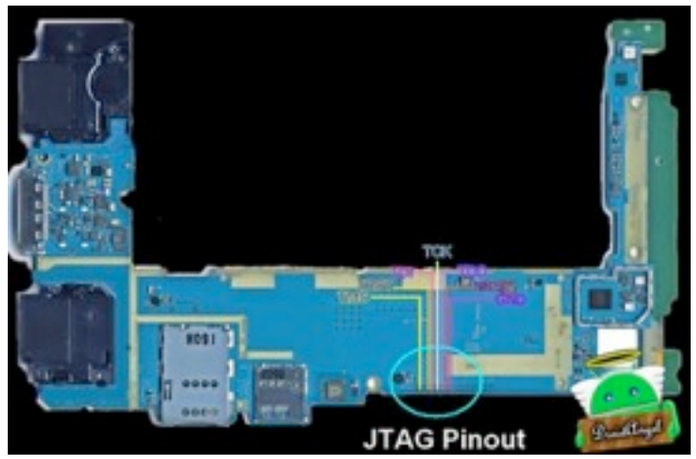
Cell phone forensic tools operate at different levels.

Tools use increasing amounts of sophistication to get data:

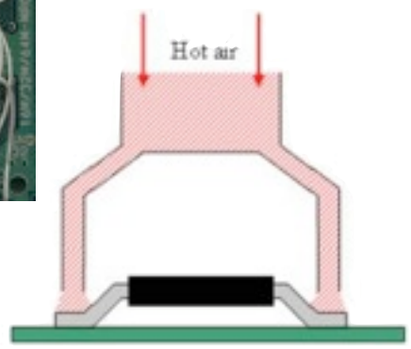
- Use the cell phone's own keyboard & display (project-a-phone)
- Download information through the cell phone's standard ports
  - —*Proprietary Phone; USB; Apple 30-pin*
  - —*SIM readers*

Project-a-Phone

- Access cell phone through programmer/proprietary ports
  - —*JTAG*



- Direct access to flash
  - —*Probes & Wires*
  - —*"Chip Off"*



http://www.data-recovery-news.com/data-recovery/how-to-remove-tsop-chips-and-micro-bga-chips/

Forensic Accessories improve “repeatability.”

“StrongHold Box” — RF shielding

- Prevents phone from calling home
- Stops remote wipe

“Device Seizure Toolbox” has lots of different cables

—http://www.paraben-forensics.com/

Acquisition

1. Identify the device: make, model, service provider

http://www.phonescoop.com/phones/finder.php

- http://www.gsmarena.com/search.php3
- http://mobile.softpedia.com/phoneFinder

2. Also note:

- Device Interface, labels, serial numbers, etc
- Synchronization software on associated computer
- Time displayed by phone

3. Select the appropriate tool

4. Extract the data (if possible).

Unobstructed Devices

Typically done with a forensically sound tool, if possible

Separately acquire the phone memory & SIM card

Usually requires phone to be turned on — which can cause problems

Unobstructed Devices

Typically done with a forensically sound tool, if possible

Separately acquire the phone memory & SIM card

Usually requires phone to be turned on — which can cause problems

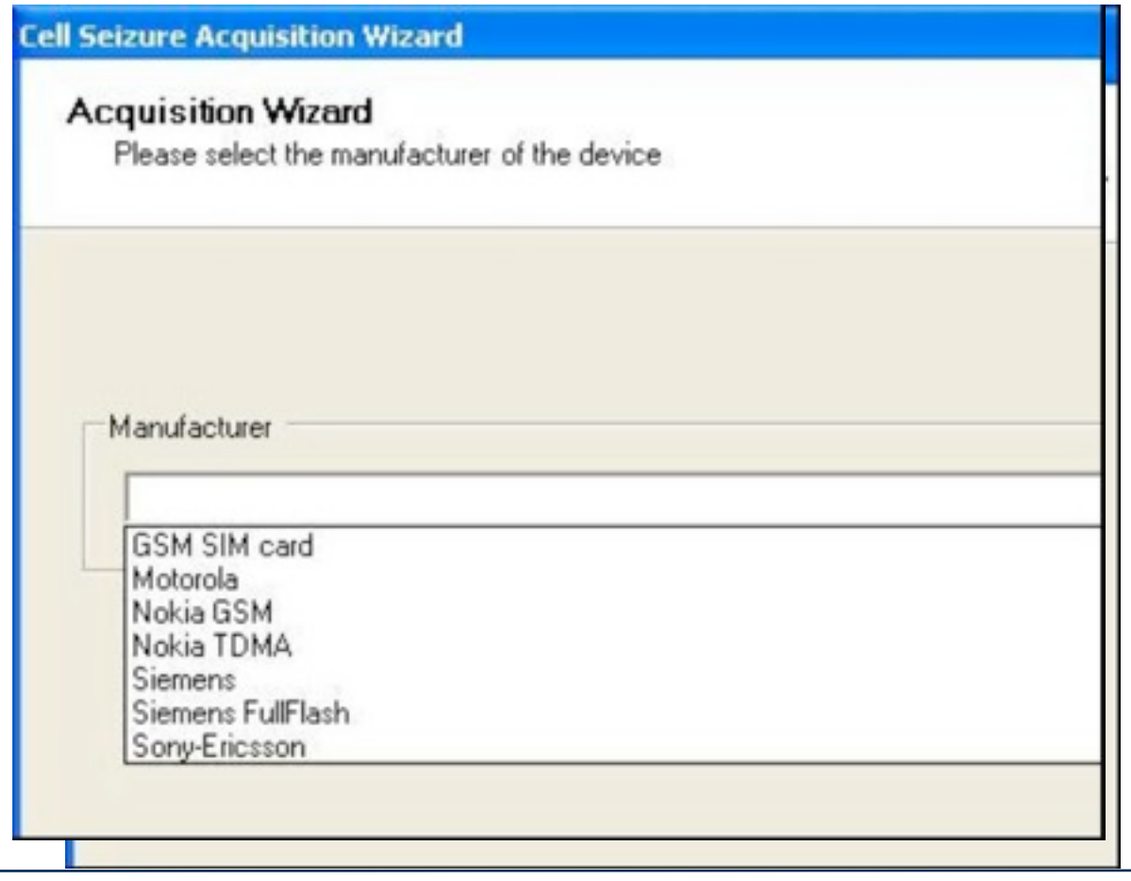
Source: NISTTIR 7250

Unobstructed Devices

Typically done with a forensically sound tool, if possible

Separately acquire the phone memory & SIM card

Usually requires phone to be turned on — which can cause problems



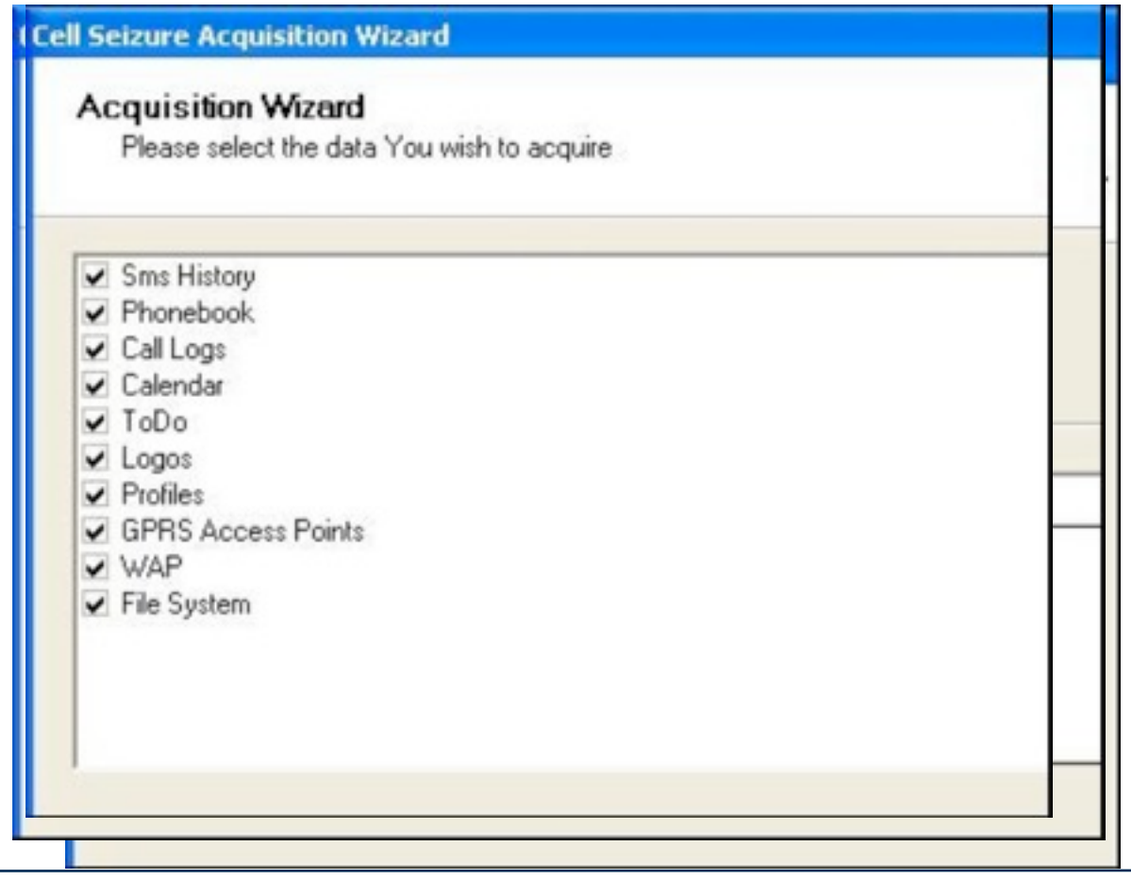
Source: NISTTIR 7250

Unobstructed Devices

Typically done with a forensically sound tool, if possible

Separately acquire the phone memory & SIM card

Usually requires phone to be turned on — which can cause problems



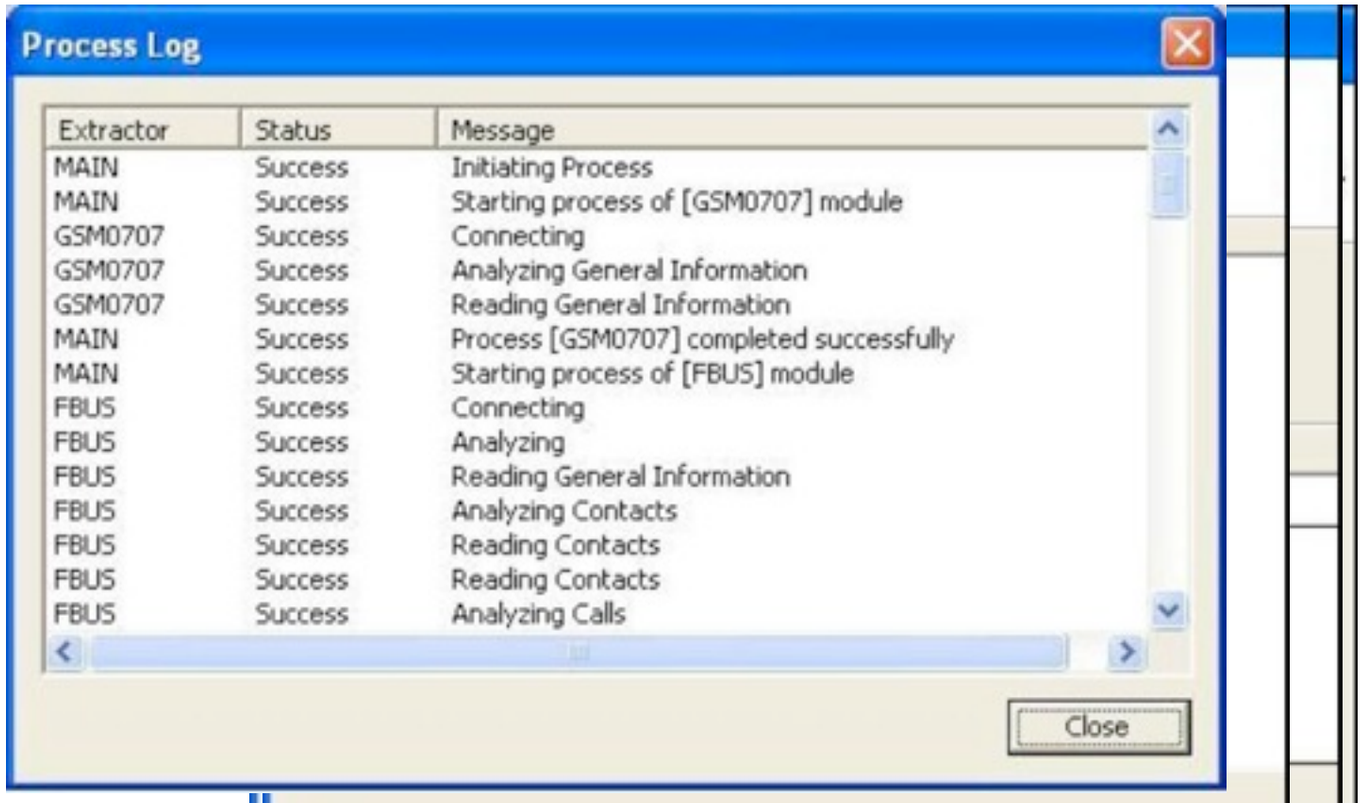
Source: NISTTIR 7250

Unobstructed Devices

Typically done with a forensically sound tool, if possible

Separately acquire the phone memory & SIM card

Usually requires phone to be turned on — which can cause problems



Source: NISTTIR 7250

“Obstructed Devices” — phones that resist extraction.

Password-protected phones or SIM cards

Blackberrys or other phones with PIN locks

- Locked SIM chips with unknown “PUK” (PIN Unlock Key) codes
- Beware remote wipe!

Options:

- Backdoor from manufacturer / Assistance from the provider
  —*Some providers will give LE the SIM’s PUK code (GSM)*
- Professionals who know how to attack the hardware
- Search Internet for developer information or hacker exploits
- Copies of the data on subject’s PC or in the cloud

Other options:

- Ask the suspect for the password, PIN, or other information
- Review seized material
- Guess (try 1234, etc.)

Obstructed Devices — Examples

PalmOS version 4.0 or earlier

- Password easily reversed after memory downloaded during sync
- Motorola DROID boot loader attack s
- A developer leaked the boot loader key
- Load custom ROMs to get around OS locks
- Nokia handsets:
- Master password that can be calculated from equipment identifier
- Netherlands Forensic Institute:
- Automated Brute Force — A machine “equipped with a robot arm and video camera the unit can systematically enter passwords until the correct entry is detected or, in the worst case, the keys become damaged.”
- General-purpose tool for examining memory chips.

Create a substitute SIM to take over the phone

Forensic SIM Toolkit

- GSM .XRY SIM Id Cloner
- TULP 2G SIMIC

Typical targets of examination and analysis: “Pattern of Life”

Investigators are looking for:

- Subscriber & equipment identifiers
- Date/time of calls, movements, etc
- Phonebook
- Appointment Calendar
- SMS, Text Messages, Instant Messages
- Dialed, incoming, & missed call log
- Electronic mail
- Photos
- Audio and video records
- Multi-media messages
- Instant messages
- Electronic Documents
- Location information

Paraben's PDA Seizure

Paraben's PDA Seizure

C:\Documents and Settings\Administrator\Desktop\cell acquisition data\Samsung_i700\Samsung_010504_1.PDA - PDA Seizure

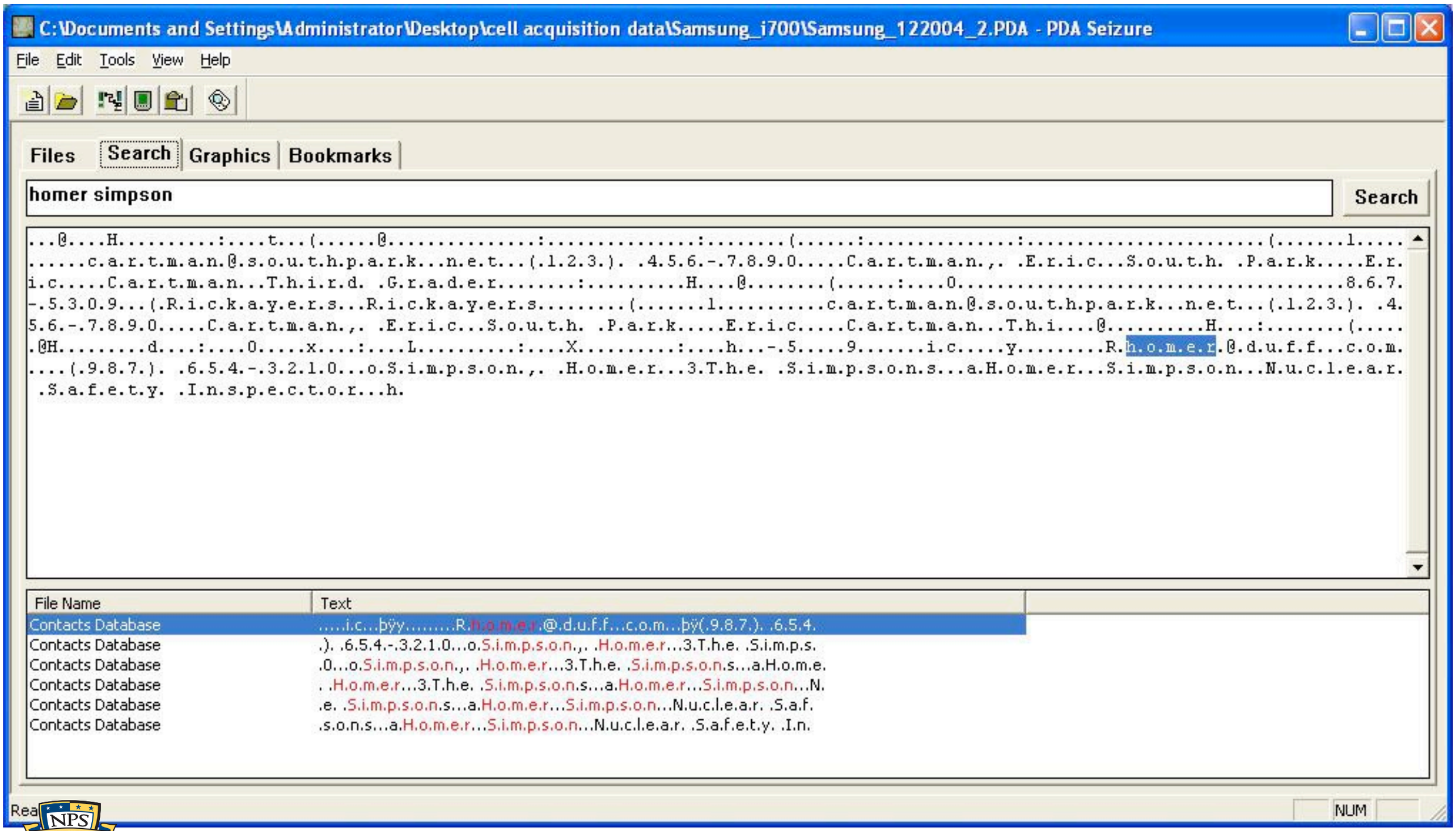
File Edit Tools View Help

Files | Search | Graphics | Bookmarks

File Path	File Name	Type	Create Date	Modify Date	Attri...	Size	Status	Location	MD5 Hash
\My Documents\	pic0001.jpg	.jpg	2005/01/05 12:41:52	2005/01/05 12:41:52	CA	27,495	Acquired	RAM	AA2B8B265DAAE3FF1D7DD36ACFA09283
\My Documents\	PIC0000.TMB	.TMB	2005/01/05 12:41:02	2005/01/05 12:41:02	CA	9,270	Acquired	RAM	03726217B9886CF17C6DD0C1A0709B10
\My Documents\	pic0000.jpg	.jpg	2005/01/05 12:41:02	2005/01/05 12:41:02	CA	41,231	Acquired	RAM	FFA24DFEFCD6EA6D28478175C134D036
\My Documents\My Pictures\	Sunset.jpg	.jpg	2005/01/05 12:46:48	2005/01/05 12:46:48	CA	71,189	Acquired	RAM	1BC5B77F3E50B7FBE12C792EE438DA45
\My Documents\My Pictures\	french.mp3	.mp3	2005/01/05 12:46:31	2005/01/05 12:46:31	CA	7,523	Acquired	RAM	A2B4FD7568F735463D11D9BAD0A29938
\My Documents\My Pictures\	chare.wav	.wav	2005/01/05 12:46:28	2005/01/05 12:46:28	CA	39,694	Acquired	RAM	FCB34DE0D5E433A2B64A45A8A265BFD3
\My Documents\My Pictures\	winter.bmp	.bmp	2005/01/05 12:46:23	2005/01/05 12:46:23	CA	353,478	Acquired	RAM	3632A33D141A8759065AD3588E40CA25
\My Documents\My Pictures\	Beer.png	.png	2005/01/05 12:46:18	2005/01/05 12:46:18	CA	2,548	Acquired	RAM	1B5BA7972C0F642A5E59D8A38DECF6A5
\My Documents\Templates\	Vehicle Mileage Log.pxt	.pxt	2003/03/21 08:20:56	2003/03/21 08:20:56	CHRA	7,498	Acquired	RAM	9C91BBEFB134B471A1330DB64FA87134
\My Documents\Templates\	To Do.psw	.psw	2003/03/21 08:20:56	2003/03/21 08:20:56	CHRA	2,616	Acquired	RAM	0F7982DEE180764A7ACB88757DA1001B
\My Documents\Templates\	Phone Memo.psw	.psw	2003/03/21 08:20:56	2003/03/21 08:20:56	CHRA	2,008	Acquired	RAM	9443F21C4AC49604D371BDCD848EB0F3
\My Documents\Templates\	Memo.psw	.psw	2003/03/21 08:20:56	2003/03/21 08:20:56	CHRA	2,112	Acquired	RAM	523694AF6762CF19DB802B8E2436DFE0
\My Documents\Templates\	Meeting Notes.psw	.psw	2003/03/21 08:20:55	2003/03/21 08:20:55	CHRA	1,908	Acquired	RAM	40FB8E424E340886885482228E45AB97
\My Documents\Templates\	Blank Document.psw	.psw	2003/03/21 08:20:55	2003/03/21 08:20:55	CHRA	0	Acquired	RAM	
\My Documents\Templates\	To Do.pwi	.pwi	2003/03/21 08:20:55	2003/03/21 08:20:55	CHRA	3,096	Acquired	RAM	B25EAC50156BC12E6FAD5ABCEACBFA29
\My Documents\Templates\	Phone Memo.pwi	.pwi	2003/03/21 08:20:55	2003/03/21 08:20:55	CHRA	2,008	Acquired	RAM	7F2CCAB0FE75072F7AB89DCD1D7136B3
\My Documents\Templates\	Memo.pwi	.pwi	2003/03/21 08:20:55	2003/03/21 08:20:55	CHRA	2,112	Acquired	RAM	CAC4C826FBA6F47AD9D088941343D5D4
\My Documents\Templates\	Meeting Notes.pwi	.pwi	2003/03/21 08:20:55	2003/03/21 08:20:55	CHRA	1,592	Acquired	RAM	B876D7DE671DE6DCCBAC8D5726DCDE82
\My Documents\Templates\	Blank Note.pwi	.pwi	2003/03/21 08:20:55	2003/03/21 08:20:55	CHRA	0	Acquired	RAM	
\Windows\	CESeizure.dll	.dll	2005/01/05 12:51:04	2005/01/05 12:51:04	CA	4,608	Acquired	RAM	148E9FEDDEB1F90CD42DAA78DDA0E58E
\Windows\	System.mky	.mky	2003/03/21 16:24:04	2003/03/21 16:24:04	CHSA	52	Acquired	RAM	D02937A4B0BB164D2AD71C0676B5A7E6
\Windows\	MsgQueueMapFileMicroso		2005/01/05 12:45:54	2005/01/05 12:45:54	CA	268,292	Acquired	RAM	1ED7E3BF7DFF04456B7002C946F37E0E
\Windows\	MsgQueueDataFileMicrosc		2005/01/05 12:45:54	2005/01/05 12:45:54	CA	2,850,820	Acquired	RAM	987F1B368E0A33EDACEF80D783EC4834
\Windows\Messaging\	0000192d1000001f.mpb	.mpb	2005/01/05 12:44:01	2005/01/05 12:44:01	CA	2	Acquired	RAM	C4103F122D27677C9DB144CAE1394A66
\Windows\Messaging\	0000192d81030102.mpb	.mpb	2005/01/05 12:43:05	2005/01/05 12:43:05	CA	0	Acquired	RAM	
\Windows\Messaging\	010017b81000001f.mpb	.mpb	2003/03/23 16:14:11	2003/03/23 16:14:11	CA	2	Acquired	RAM	C4103F122D27677C9DB144CAE1394A66
\Windows\Messaging\Attachme	192d-1931.att	.att	2005/01/05 12:43:04	2005/01/05 12:43:04	CA	13,320	Acquired	RAM	933F8BC6F80311C84B6EE11527975580

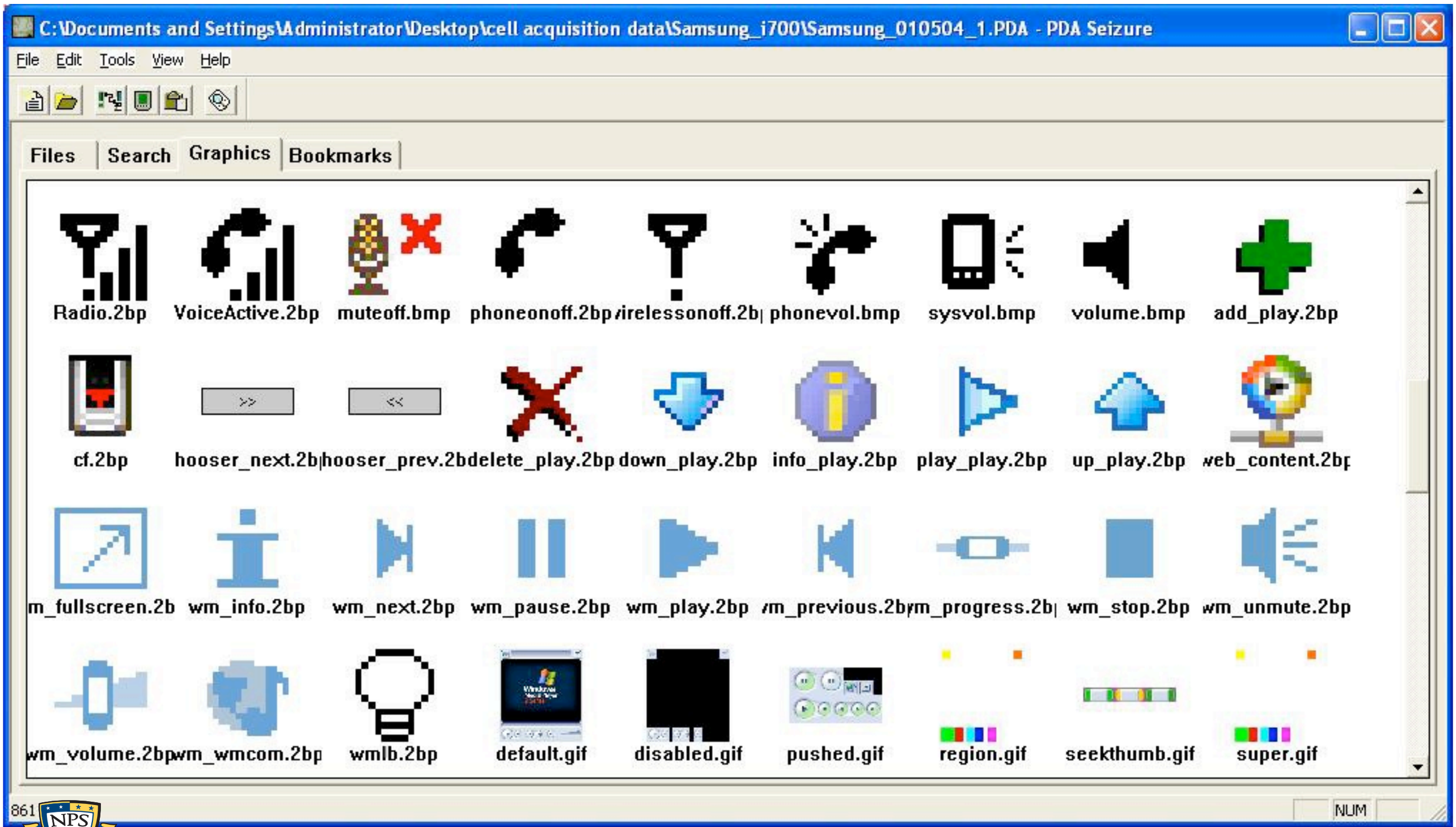
861 NUM

Paraben's PDA Seizure

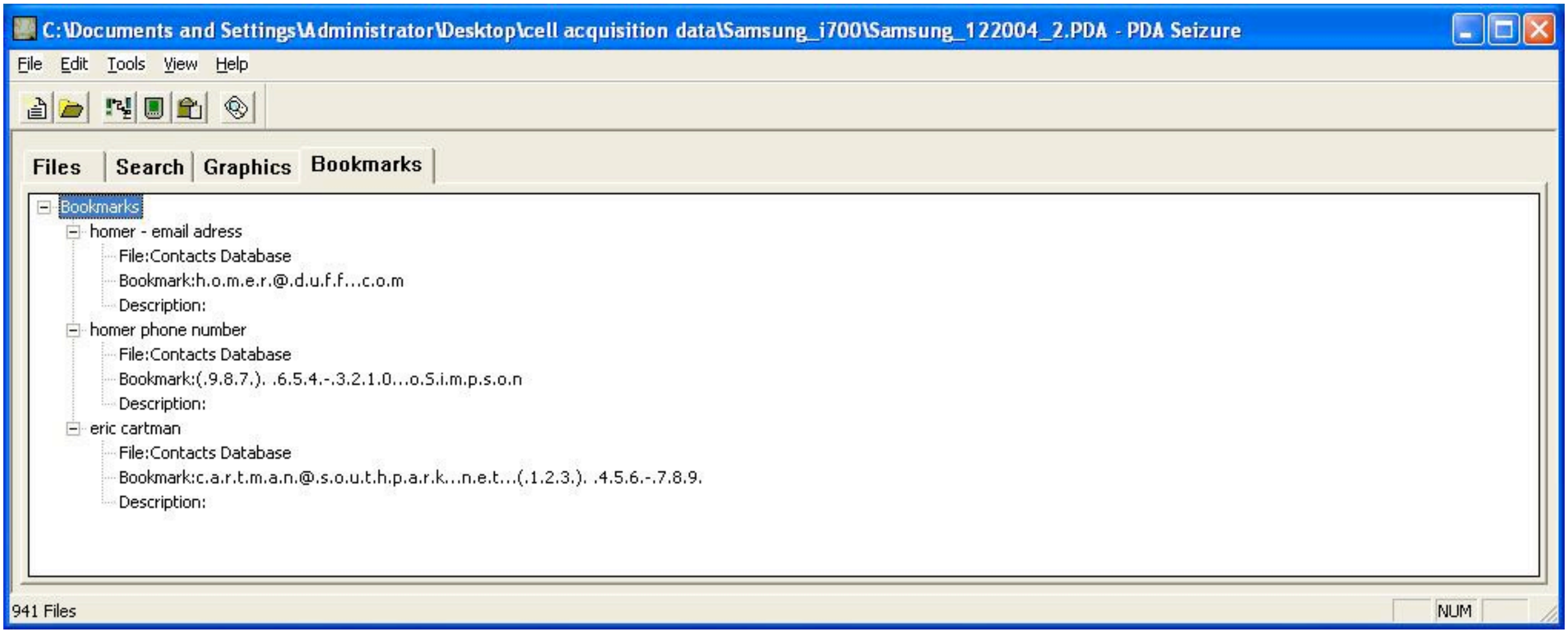


Source: NISTTIR 7250

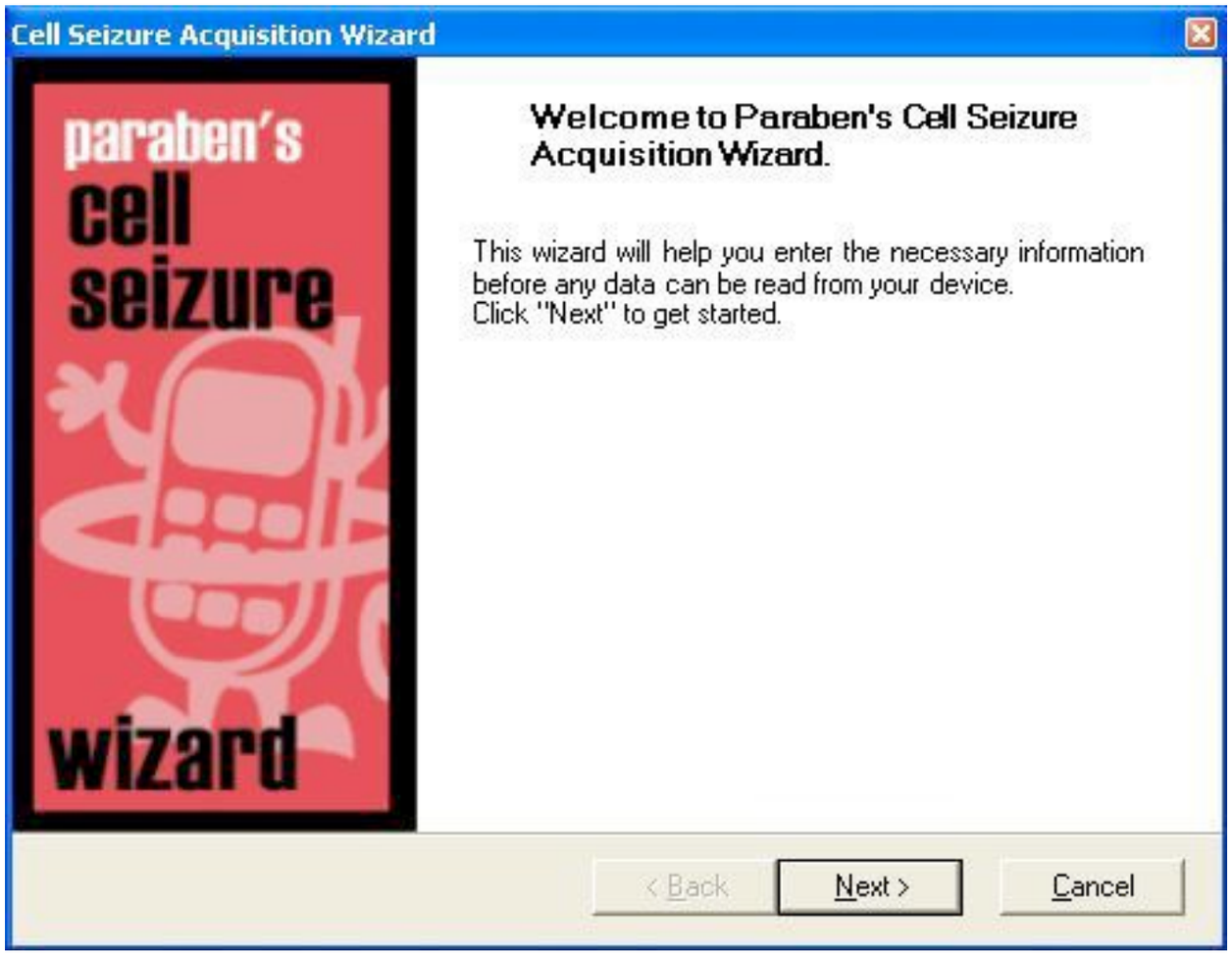
Paraben's PDA Seizure



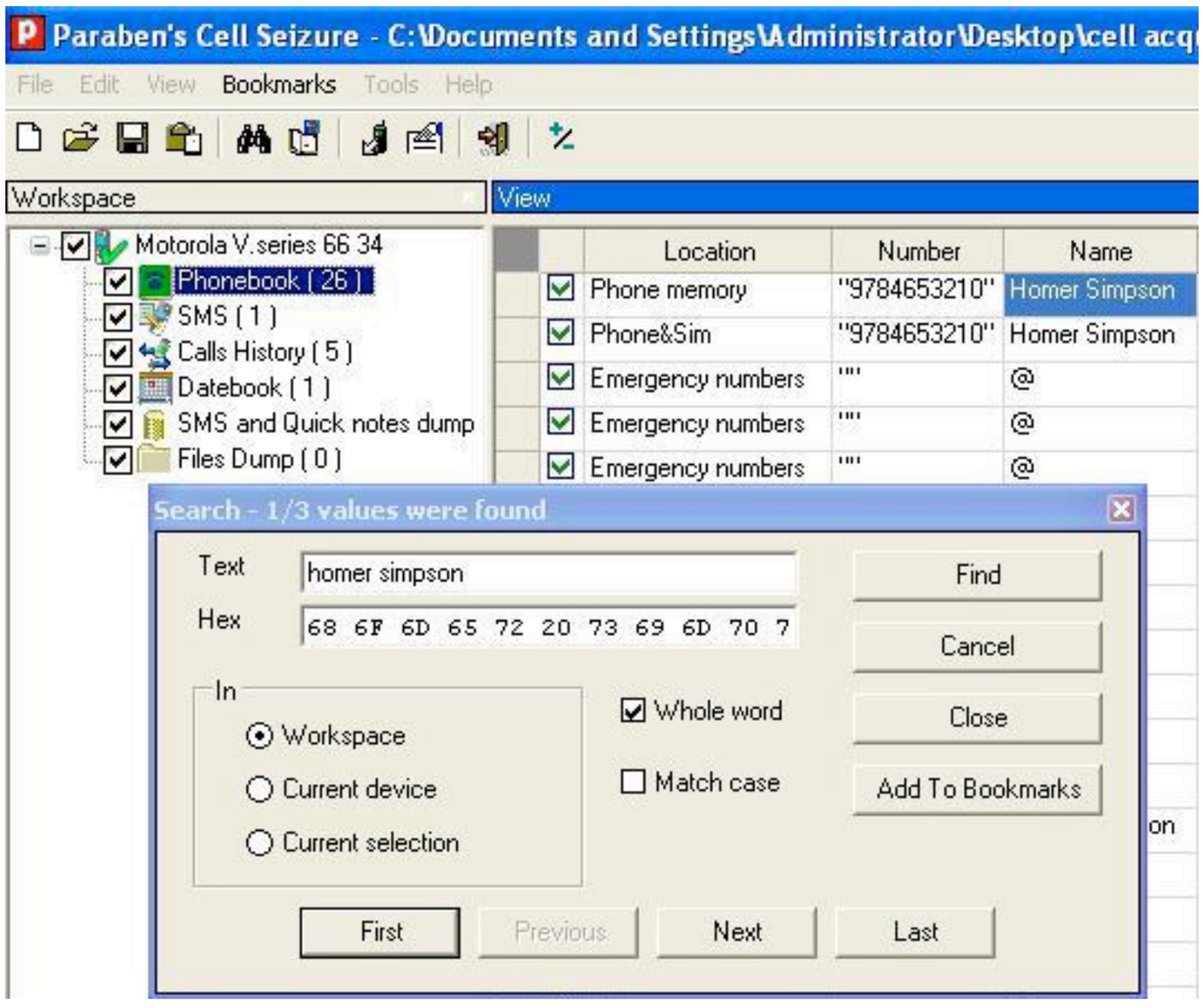
Paraben's PDA Seizure

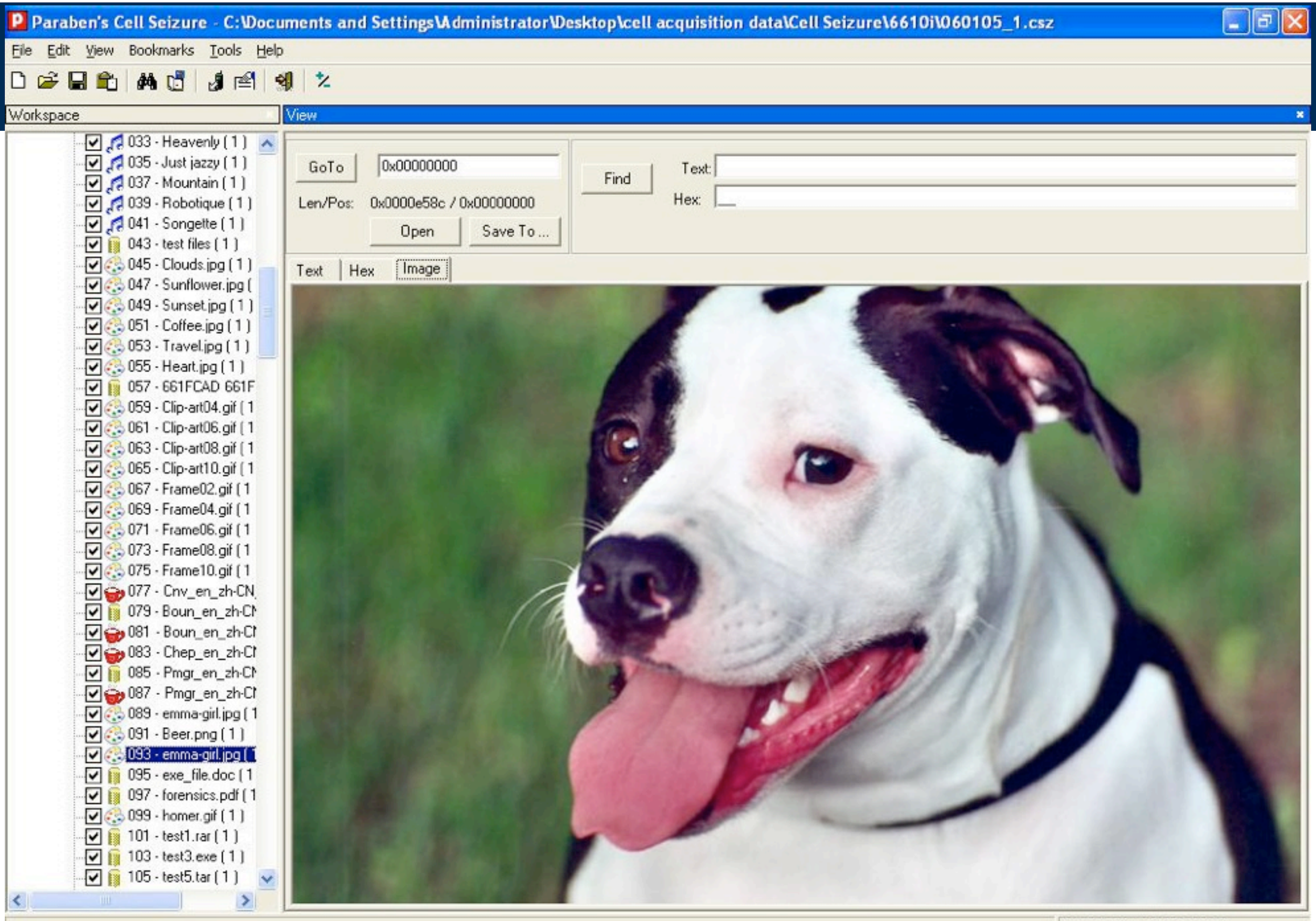


Paraben's Cell Seizure



Paraben's Cell Seizure




Paraben's Cell Seizure - C:\Documents and Settings\Administrator\Desktop\cell acquisition data\Cell Seizure\6610i\060105_1.csz
File Edit View Bookmarks Tools Help
Workspace
View
033 - Heavenly (1)
035 - Just jazzy (1)
037 - Mountain (1)
039 - Robotique (1)
041 - Songette (1)
043 - test files (1)
045 - Clouds.jpg (1)
047 - Sunflower.jpg (
049 - Sunset.jpg (1)
051 - Coffee.jpg (1)
053 - Travel.jpg (1)
055 - Heart.jpg (1)
057 - 661FCAD 661F
059 - Clip-art04.gif (1
061 - Clip-art06.gif (1
063 - Clip-art08.gif (1
065 - Clip-art10.gif (1
067 - Frame02.gif (1
069 - Frame04.gif (1
071 - Frame06.gif (1
073 - Frame08.gif (1
075 - Frame10.gif (1
077 - Cnv_en_zh-CN
079 - Boun_en_zh-CN
081 - Boun_en_zh-CN
083 - Chep_en_zh-CN
085 - Pmgr_en_zh-CN
087 - Pmgr_en_zh-CN
089 - emma-girl.jpg (1
091 - Beer.png (1)
093 - emma-girl.jpg (
095 - exe_file.doc (1
097 - forensics.pdf (1
099 - homer.gif (1)
101 - test1.rar (1)
103 - test3.exe (1)
105 - test5.tar (1)
GoTo
0x00000000
Len/Pos: 0x0000e58c / 0x00000000
Open
Save To ...
Find
Text:
Hex:
Text
Hex
Image
0x0000e58c / 0x00000000

Paraben's Cell Seizure Report Wizard


```
 Paraben's Cell Seizure Exported Workspace
      Motorola V.series 66

      Properties
Name------Value------
Manufacturer : Motorola :
Model : V.series 66 :
Serial number :  IMEI449276812531841.... :
MD5 : 47bae0d246ffa06a341b3c62ef61c7ff :
      Phonebook
Location------Number------Name------
Phone memory : "9784653210" : Homer Simpson :
Phone&Sim : "9784653210" : Homer Simpson :
      Properties
Name------Value------
MD5 : d41d8cd98f00b204e9800998ecf8427e :
      SMS
Number------Status------Date/Time------Message------
 "2404016148" :  "STO UNSENT" :  : 休 :
      Properties
Name------Value------
MD5 : d41d8cd98f00b204e9800998ecf8427e :
      Calls History
Name------Number------Direction------
Homer Simpsons/W : "9874653210" : Dialled calls :
 : "301975XXXX" : Dialled calls :
 : "301975XXXX" : Dialled calls :
      Properties
Name------Value------
MD5 : d41d8cd98f00b204e9800998ecf8427e :
      Datebook
Ricks birthday : 0 : 0 : 2000-01-30 00:00 : 1440 :  : non reoccurring :
      Properties
Bookmarks     Homer Simpson : Homer Simpson     Datebook entry : Ricks birthday
```

Cell Phone Forensics: References & Resources

USG References are out-of-date but useful nonetheless:

- Guidelines on Cell Phone Forensics (NIST SP 800-101)
  - —*May 2007*
  - —*http://csrc.nist.gov/publications/drafts/Draft-SP800-101.pdf*
- Cell Phone Forensic Tools: An Overview and Analysis (NISTIR 7387)
  - —*March 2007 (!)*
  - —*http://csrc.nist.gov/publications/nistir/nistir-7387.pdf*
- PDA Forensic Tools: An Overview and Analysis (NISTIR 7100)
  - —*August 2004 (!!)*
  - —*http://csrc.nist.gov/publications/nistir/nistir-7100-PDAForensics.pdf*

Recently published books are better, but not peer reviewed:

- iPhone Forensics: Recovering Evidence, Personal Data, and Corporate Assets byJonathan A. Zdziarski (Paperback - Sep 19, 2008)
- iPhone and iOS Forensics: Investigation, Analysis and Mobile Security for Apple iPhone, iPad and iOS Devices by Andrew Hoog and Katie Strzempka (Paperback - Jun 30, 2011)
- Android Forensics: Investigation, Analysis and Mobile Security for Google Android, Andrew Hoog (Paperback - June 29, 2011)

Android Forensic Targets

You've got an Android Phone. Now what?

SIM:

- Identity information.
- Possibly Address Book or SMS records from a previous phone
- On-board Flash (256M-2GB)

- Android file system (YAFFS2)
- Call history; messages; position information; network information; etc
- Downloaded applications & application data

Removable Flash (1GB-32GB)

- Downloaded applications & application data
- Media (songs; video; images); Documents
- Information from other computers (remember, phone can be a "thumb drive")

RAM (256M-1GiB)

- Linux; Dalvik (Java) VM; user programs
- May be only way to recover encryption keys, passwords, etc.

Two approaches for Android Forensics: Online & Offline

Online Analysis: Use Android to analyze Android

Enable USB debugging and debug with Android Debug Bridge (adb)

—*http://developer.android.com/guide/developing/tools/adb.html*

- Load an application that extracts data to your analysis machine
- RAM

—*Physical Dump of NAND flash*

Offline Analysis: Analyze Android as a storage system

- Analyze SDCard as a traditional FAT file system
- Logical analysis of YAFFS2 files

—*Less to get, but easier to get at*

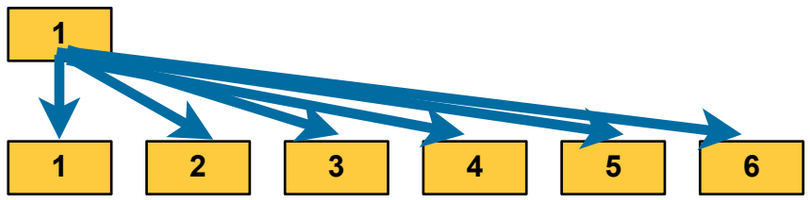
Which approach you choose depends on:

- Your goals — conviction, discovery, research
- Your skill level & available tools
- Legal requirements (i.e.: will the results be used in court?)

Flash memory is very different from traditional RAM.

Defining characteristics:

- Memory written in blocks (100s-1000s of bits per block — think “sectors”)
  - —*Must be erased before it can be written*
- Memory erased in *pages* (10,000s of bits per page — think 4K pages)
- Each bit has limited lifetime (typically 1000 — 100,000 cycles)
- Therefore, writes must be *wear leveled*



NOR flash (not always present)

- True random access (direct execution)
- Low-density (expensive)
- Boot code can execute directly out of NOR

NAND flash (always present)

- Block-oriented access
- High density (Single Layer Cells & Multi Layer Cells)
- ROM boot code (in the microprocessor) can copy NAND into RAM and execute.

There are two-approaches for remapping.

File Level — Flash File System

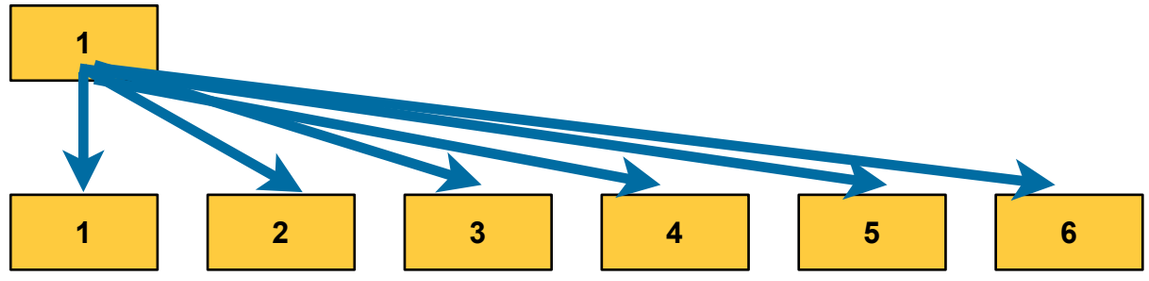
- Operating system directly controls writing & erasing.
- Files may be proactively moved to assist in leveling
- JFFS2 (Journaling Flash File System #2); YAFFS (Yet Another Flash File System); YAFFS2

Block Level — Flash Translation Layer

- Flash device appears as a block device
- Operating system rewrites as normal
- “Flash Translation Layer” transparently remaps & erases as necessary
- Used by all SD cards and SSDs

“TRIM” Command

- Tells FTL that a sector will not be read again
- Lets OS give SD/SSD “hint.”
- Implemented in Windows 7 and Linux ext4



Wear leveling means you can recover data *after it is deleted and overwritten*.

Assume this sequence of events:

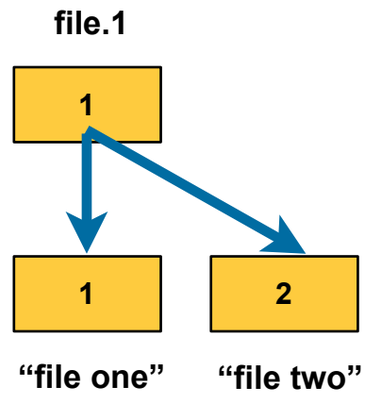
```
echo “file one” > file.1
echo “file two” > file.2
dd if=file1 of=file.2
```

These commands are executed at the *logical layer*

YAFFS2 would rewrite the directory entry for “file.2” to point at the new flash pages

- *A SSD or SDCard would rewrite the FTL so that the logical block # pointed to by the file.2 directory entry pointed to the new data*

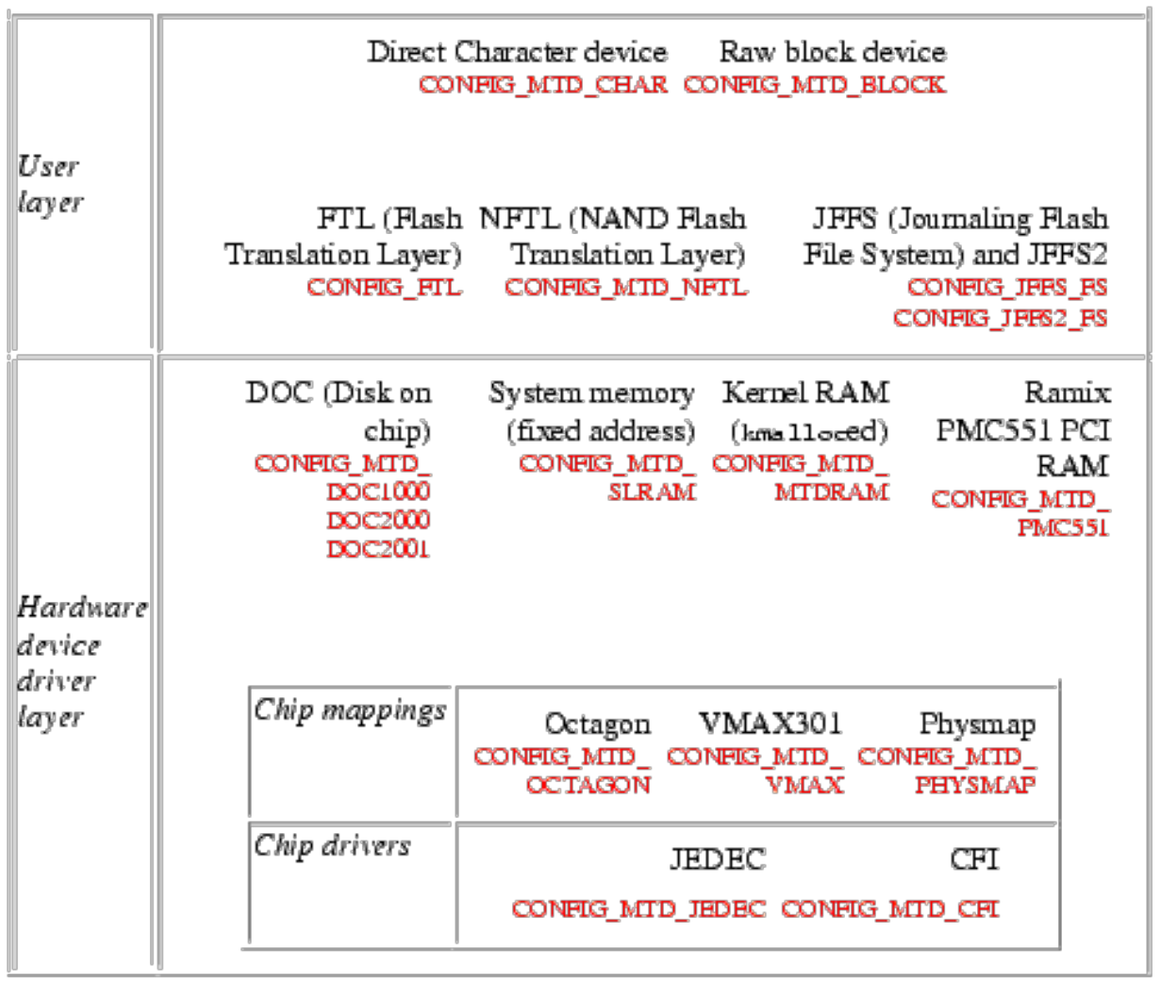
If you can access the *physical* layer, *you can recover the previous contents of file.2*



Android uses the Linux Memory Technology Device (MTD) to access flash memory.

The MTD has a Flash Translation Layer...

- ... but flash file systems (JFFS, JFFS2, YAFFS and YAFFS2) go directly to the hardware layer.



http://www.stlinux.com/howto/Flash/MTD

“Logical” vs. “Physical” dump.

A logical dump is a dump of the *records* or *files*

From data providers

- From walking the file system
- adb pull /*dir* local # don’t pull /proc

A physical dump is a dump of *sectors* or *pages*

YAFFS and YAFFS2:

—*raw is the individual flash pages*

—*16-bytes of Out-of-Band information stored every 512, 1024, or 2048 bytes must be removed*

—*Requires a YAFFS/YAFFS2 implementation to extract files*

—*FAT32 (or NTFS)*

—*raw is the individual disk “sectors” (512 or 4096 bytes)*

—*Requires FAT32 implementation to extract files*

—*Mount with a loop-back device to access allocated files*

- Use SleuthKit, EnCase, or FTK to access *deleted files*.

Sleuthkit — A user-level forensic file system

SleuthKit accesses files in an image *without mounting the disk*:

- mmls *image.raw* — list partitions
- fls [-o offset] *image.raw* [inode] — list files (optionally from a directory)
- icat [-o offset] *image.raw inode* — output the contents of an inode to stdout

```
$ ls -l nps-2009-canon2-gen5.raw
-rw-r--r--  1 simsong  admin  31129600 Jan  6  2009 nps-2009-canon2-gen5.raw
$ mmls nps-2009-canon2-gen5.raw
DOS Partition Table
Offset Sector: 0
Units are in 512-byte sectors

     Slot    Start        End          Length       Description
00:  Meta    0000000000   0000000000   0000000001   Primary Table (#0)
01:  -----   0000000000   0000000050   0000000051   Unallocated
02:  00:00   0000000051   0000060799   0000060749   DOS FAT16 (0x04)
$ fls -o 51 nps-2009-canon2-gen5.raw
r/r 3:      CANON_DC    (Volume Label Entry)
d/d 4:      DCIM
v/v 971779: $MBR
v/v 971780: $FAT1
v/v 971781: $FAT2
d/d 971782: $OrphanFiles
$
```

Using Sleuthkit for Android Forensics

Approach #1: MicroSD card

- Remove the MicroSD card and examine with SleuthKit
- Important: Use a *write blocker* to prevent modification to the SD card
- Advantage: Easy-to-do; no change to SD card
- Disadvantage: Will not read encrypted .apk files; shutting down may wipe important info

Approach #2: Analyze the Android device via USB

- Attach the Android device to your computer and select “USB Storage.”
- One or more partitions corresponding to the Android device *may appear*
- Question: Can we use a *write blocker* to prevent modification? (I don’t know)
- Advantage: Easy-to-do
- Disadvantage: May change Android device *even with write blocker*

Approach #3: Dump the Android device and analyze offline.

File formats typical on Android Phones

SQLite data files

- Public domain database holds SQL Schema, Tables, Rows, Columns
- Journal stored in secondary file
- Most of today's tools ignore the journal and deleted data

Internal log (circular buffer in memory)

- Log Collector (http://code.google.com/p/android-log-collector/)
- logcat

```
adb shell logcat > log.txt
```

- aLogCat

Text log files

- Some third party programs (e.g. DropBox) may store text logs
- Does the base Android system create text log files?

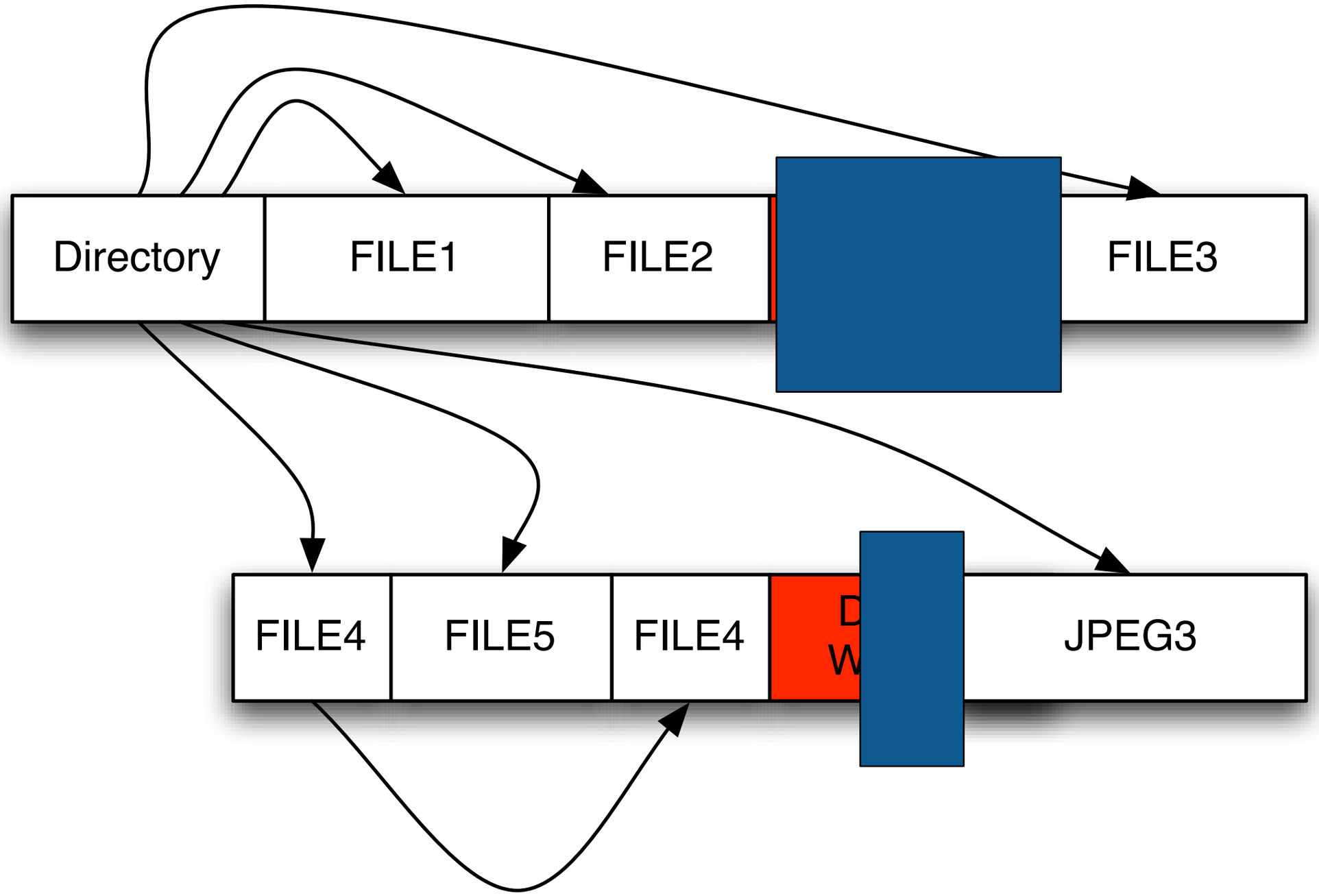
Android Forensics References

- “Recovery of Deleted Data from Flash Memory Devices”, Capt. James Regan, Master’s Thesis, Naval Postgraduate School, 2009. http://simson.net/clips/students/09Sep_Regan.pdf
- “Android Forensics: Simplifying Cell Phone Examinations,” Lessard & Kessler, Small Scale Digital Device Forensics Journal, Vol. 4, No. 1, September 2010, http://www.ssddfj.org/papers/SSDDFJ_V4_1_Lessard_Kessler.pdf
- http://viaforensics.com/category/android-forensics/
- http://viaforensics.com/android

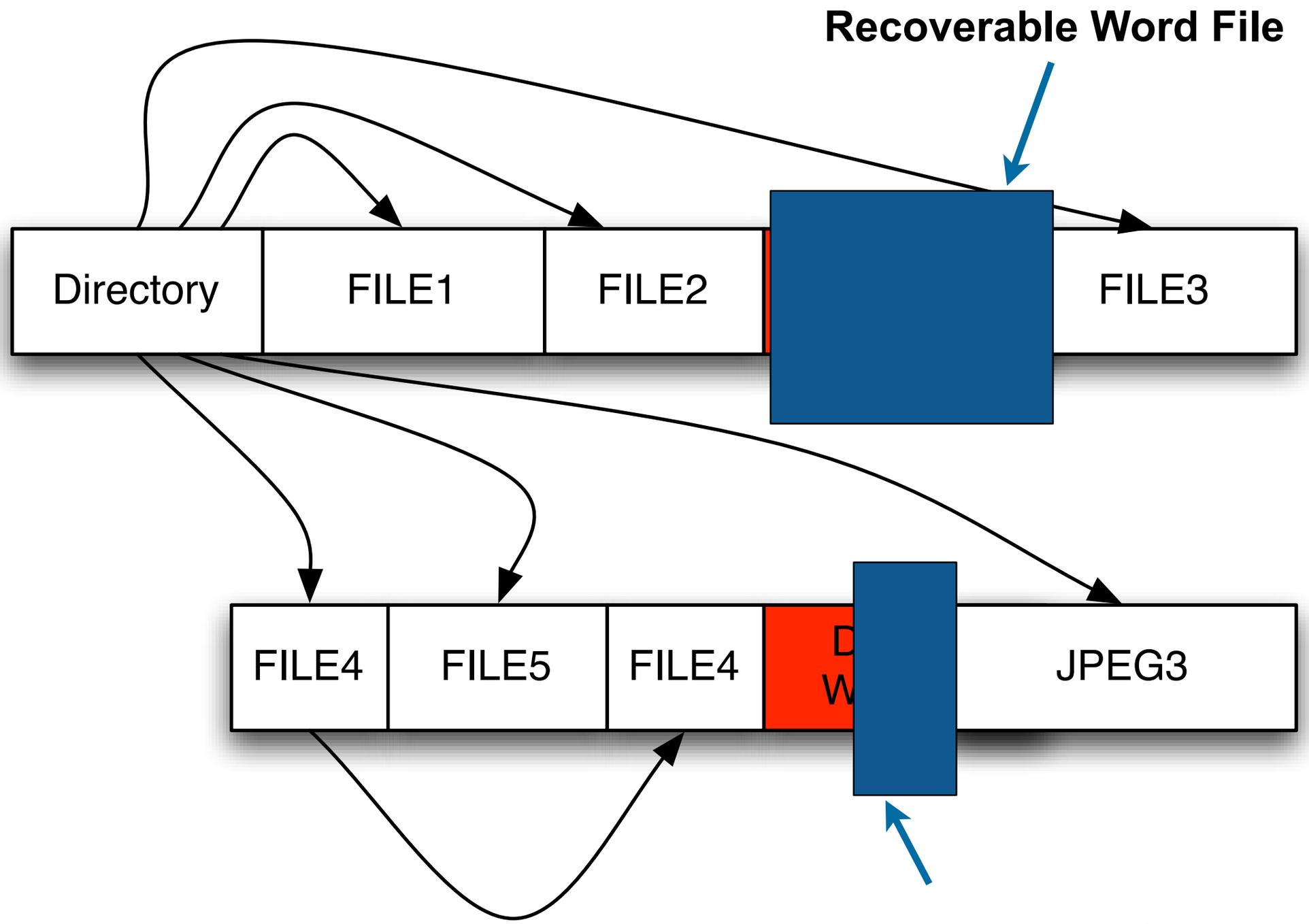
http://www.flickr.com/photos/12066488@N00/1397125588

File Carving

“Carving” searches for objects based on content, rather than on metadata.



“Carving” searches for objects based on content, rather than on metadata.



Recoverable embedded JPEG

File carving is a powerful tool for finding useful pieces of information.

What can be carved:

- Disks & Disk Images
- Memory
- Files of unknown format (to find embedded objects)

Objects that can be recovered:

- Images
- Text files & documents
- Cryptographic Keys

Why carve?

- Directory entries are overwritten
- Directory entries are damaged
- File formats aren’t known

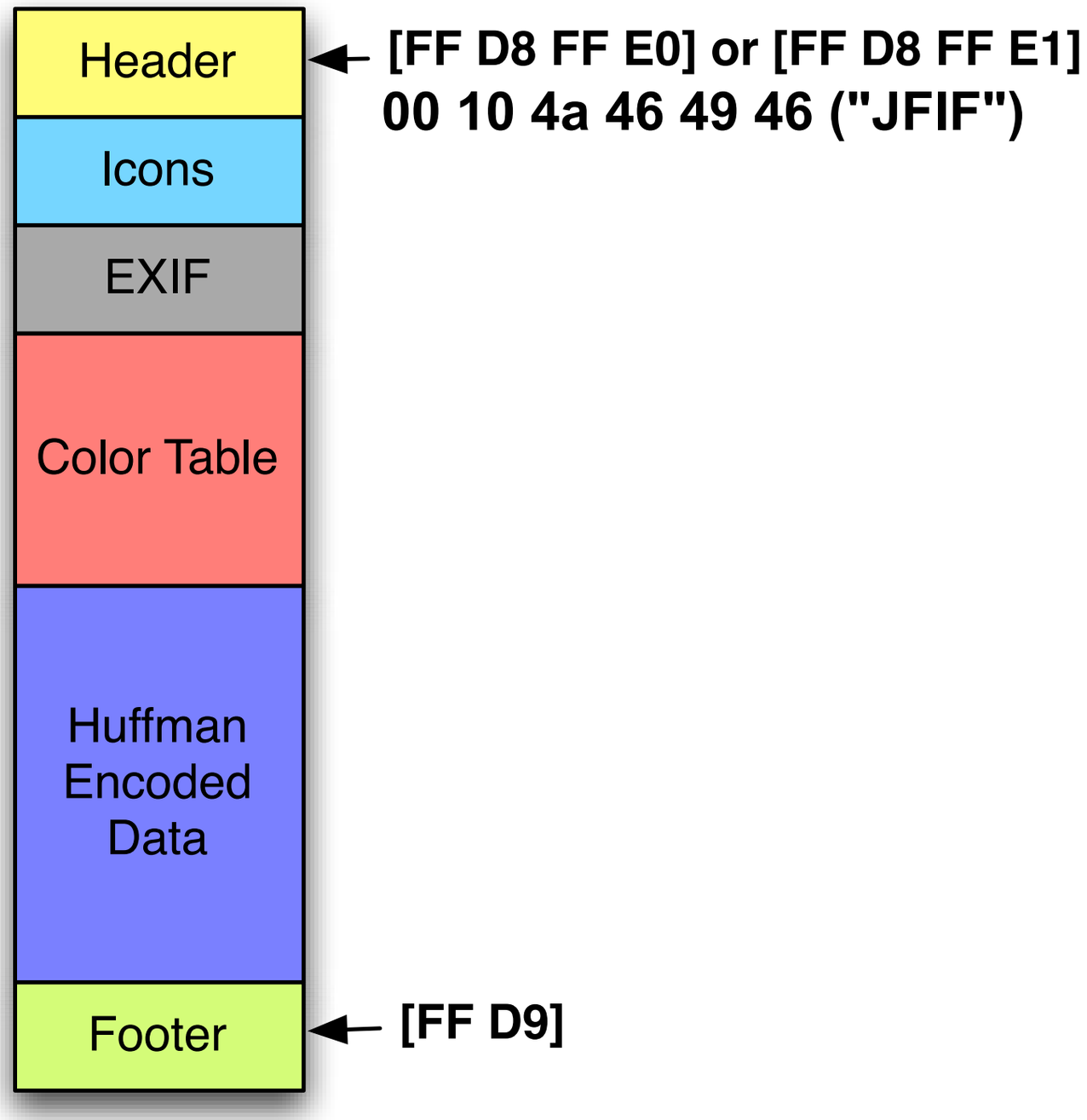
Example: Carving JPEG Files

JPEGs are container files

- Standard Header
- Standard Footer
- Embedded Images

Carving strategy:

- Find all headers
- Find all footers
- Save sectors to files



Header/Footer carving involves saving the data between a known header & known footer.

This strategy is used by foremost and scalpel.

Disk Sectors ➔

Possible explanations:

- This file may be fragmented.
- The file may have been overwritten.

If the file is fragmented, it can be recovered with fragment recovery carving

Header/Footer carving involves saving the data between a known header & known footer.

This strategy is used by foremost and scalpel.

Disk Sectors ➔

Possible explanations:

- This file may be fragmented.
- The file may have been overwritten.

If the file is fragmented, it can be recovered with fragment recovery carving

Header/Footer carving involves saving the data between a known header & known footer.

This strategy is used by foremost and scalpel.

Disk Sectors ➔

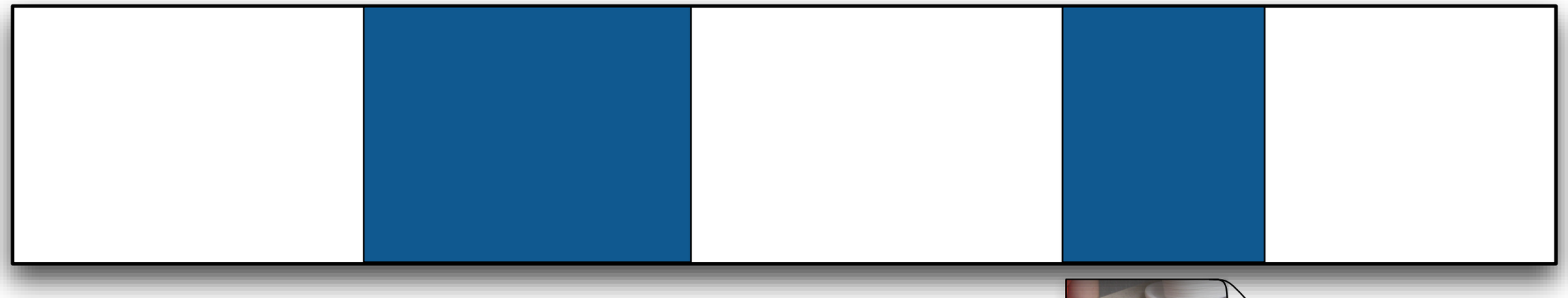
Possible explanations:

- This file may be fragmented.
- The file may have been overwritten.

If the file is fragmented, it can be recovered with fragment recovery carving

Header/Footer carving involves saving the data between a known header & known footer.

This strategy is used by foremost and scalpel.



Disk Sectors ➔

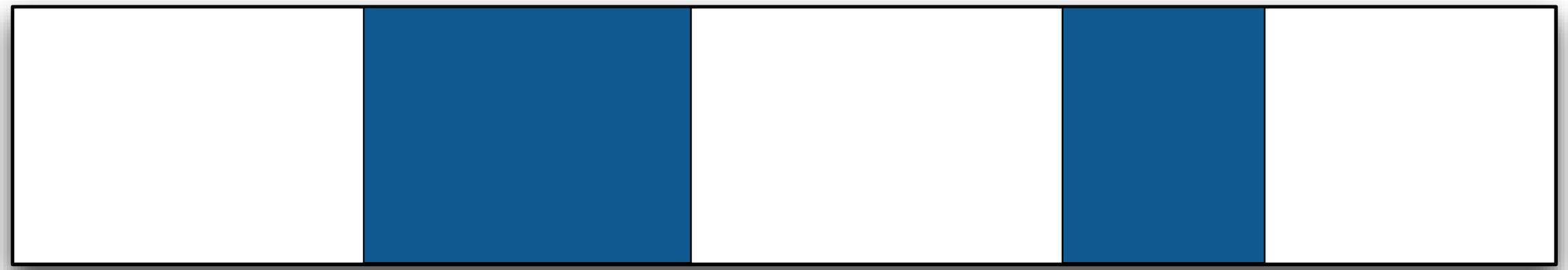
Possible explanations:

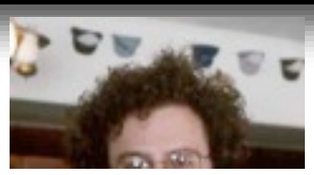
- This file may be fragmented.
- The file may have been overwritten.

If the file is fragmented, it can be recovered with fragment recovery carving

Header/Footer carving involves saving the data between a known header & known footer.

This strategy is used by foremost and scalpel.





Disk Sectors ➔

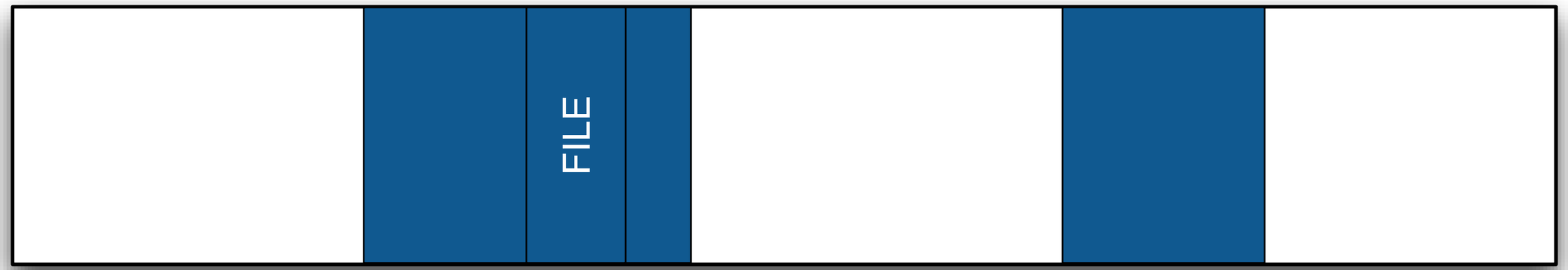
Possible explanations:

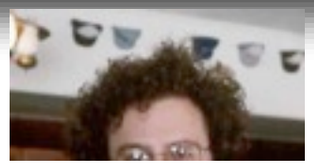
- This file may be fragmented.
- The file may have been overwritten.

If the file is fragmented, it can be recovered with fragment recovery carving

Header/Footer carving involves saving the data between a known header & known footer.

This strategy is used by foremost and scalpel.





Disk Sectors ➔

Possible explanations:

- This file may be fragmented.
- The file may have been overwritten.

If the file is fragmented, it can be recovered with fragment recovery carving

Is fragment reassembly carving important? We analyzed 400 hard drives to find out.

Today's file carvers cannot process fragmented files

Is this a problem? We don't know

We have disk images from many used hard drives

These drives simulate drives taken from production during a search

- ≈275 had file systems we could analyze when the study was done.

Files can be fragmented into two or more pieces.

	FAT[1]	NTFS	UFS
# File systems:	219	51	5
# Fragments	Number of Files		
(contiguous)	1,285,975	502,050	70,222
2	25,151	20,851	10,932
3	4,929	5,622	1,047
4	2,473	3,176	408
5–10	4,340	11,730	658
11–20	1,591	7,001	94
21–100	1,246	10,912	13
101–1000	185	5,672	0
1001–	2	567	0
Total Files:	1,325,892	567,581	83,374

Forensically important files are more likely to be fragmented than non-important files.

Log files:

- Written by appending

Microsoft OLE files:

- .doc, .xls, .ppt
- Files designed for update-in-place

Other files likely to fragment:

- Microsoft Windows Registry
- sqlite databases (untested)
- Big files (video)

What typically doesn't fragment:

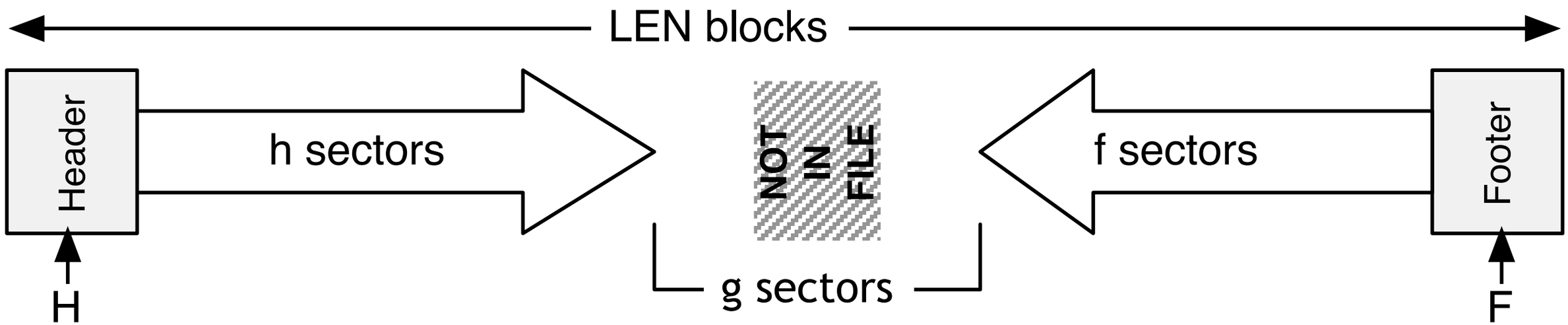
- JPEGs on disks with free space.

	file	Size of files with 2 fragments:		
Ext	count	avg	stddev	max
pnf	7,583	41,583	81,108	1,317,368
dll	7,479	221,409	384,758	9,857,608
html	3,417	28,388	66,694	2,505,490
jpeg	2,963	29,673	178,563	6,601,153
gif	2,566	22,133	99,370	3,973,951
exe	2,348	399,528	4,354,053	206,199,144
1	1,125	57,475	130,630	1,998,576
dat	780	291,407	673,906	7,793,936
z	716	74,353	340,808	6,248,869
h	690	16,444	12,232	110,592
inf	683	79,578	101,448	522,916
wav	575	1,949,459	6,345,280	39,203,180
swf	548	62,582	120,138	1,155,989
ttf	540	163,854	649,919	10,499,104
sys	513	1,276,323	12,446,966	150,994,944
txt	480	33,410	275,641	5,978,896
hlp	475	185,259	375,461	3,580,078
tmp	450	206,908	772,290	8,388,608
so	440	103,939	205,617	1,501,148
wmf	418	48,864	49,869	586,414
...	...	...	...	

Table 7: Most common files in corpus consisting of two fragments, by file extension.

Fragment Recovery Carving can reassemble fragmented files using validation.

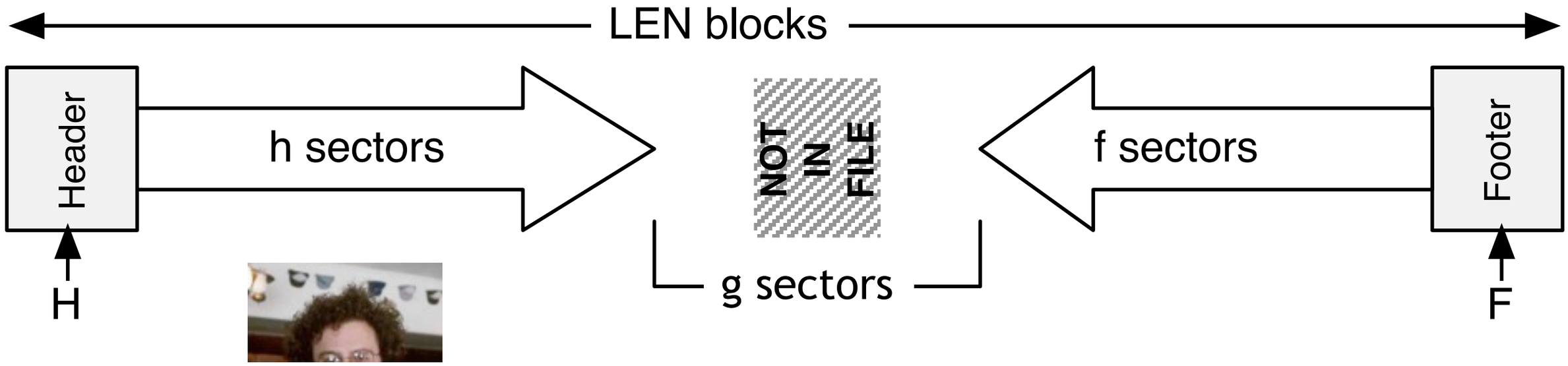
Fragment Recovery Carving:




```
LEN = S-F+1
for I in range(0,LEN):
    for J in range(0,LEN-I):
        data = blocks[S:S+I] + blocks[F-J:J]
        if valid(data)==True:  save(data)
```

Fragment Recovery Carving can reassemble fragmented files using validation.

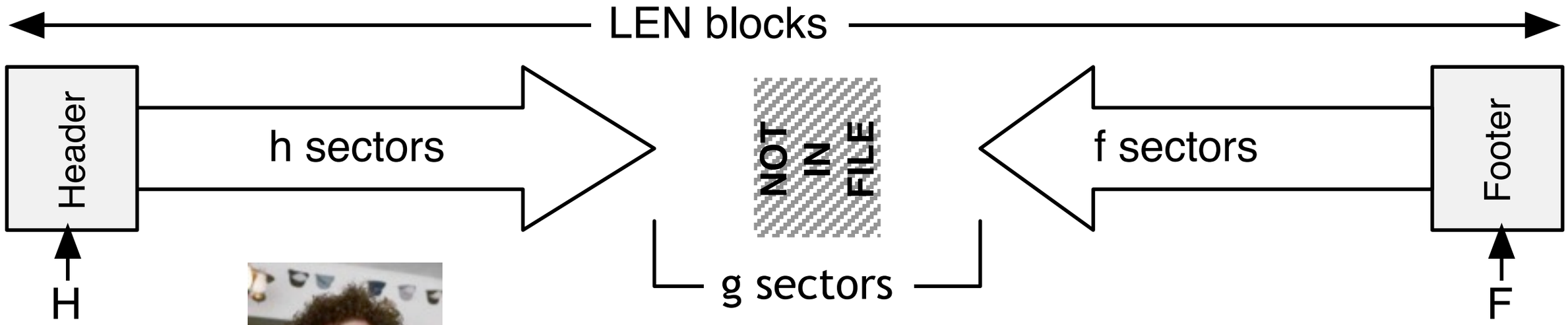
Fragment Recovery Carving:




```
LEN = S-F+1
for I in range(0,LEN):
    for J in range(0,LEN-I):
        data = blocks[S:S+I] + blocks[F-J:J]
        if valid(data)==True:  save(data)
```

Fragment Recovery Carving can reassemble fragmented files using validation.

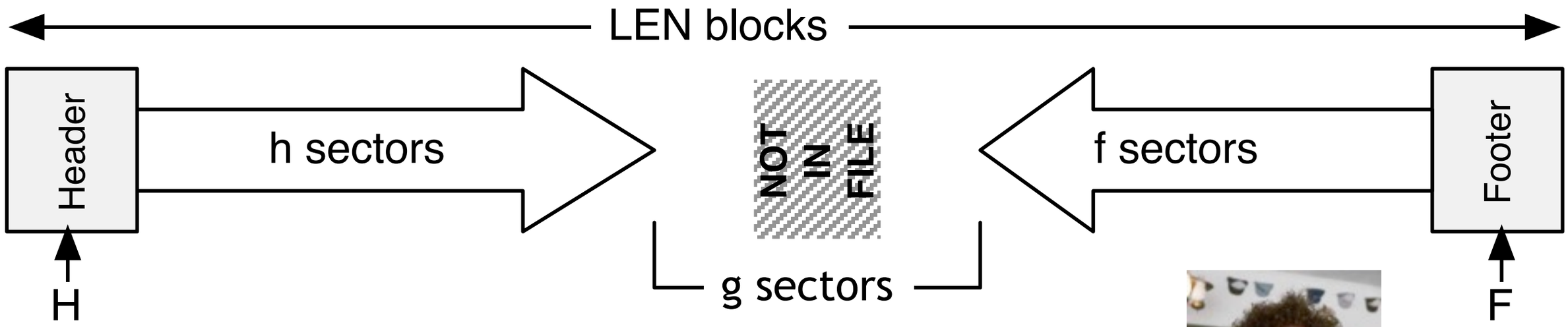
Fragment Recovery Carving:

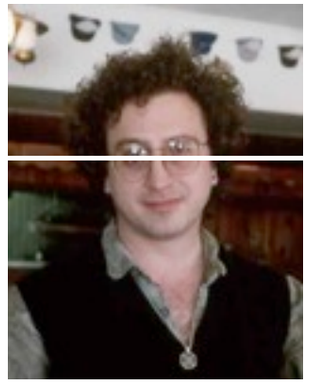


```
LEN = S-F+1
for I in range(0,LEN):
    for J in range(0,LEN-I):
        data = blocks[S:S+I] + blocks[F-J:J]
        if valid(data)==True:  save(data)
```

Fragment Recovery Carving can reassemble fragmented files using validation.

Fragment Recovery Carving:

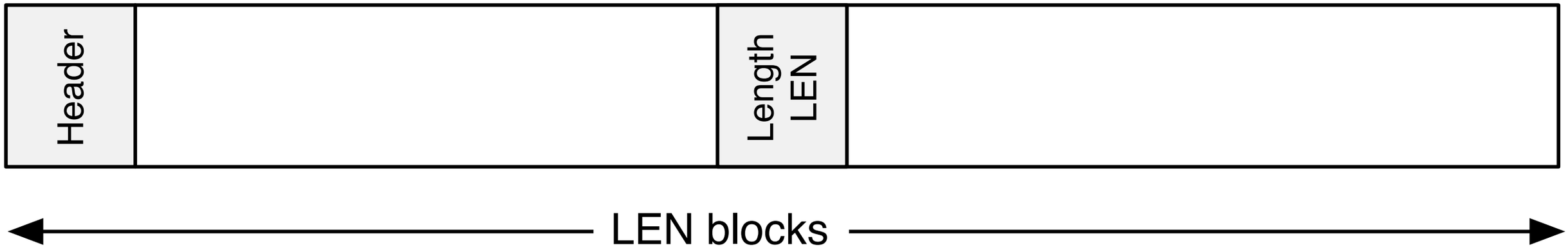




```
LEN = S-F+1
for I in range(0,LEN):
    for J in range(0,LEN-I):
        data = blocks[S:S+I] + blocks[F-J:J]
        if valid(data)==True:  save(data)
```

Header/Length Carving takes advantage of blocks that code a file's length.

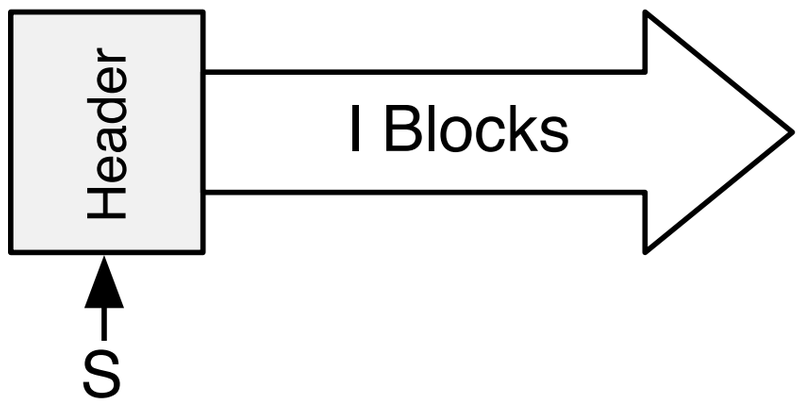
Header/Length sectors: (LEN blocks are found in ZIP & MSOffice)

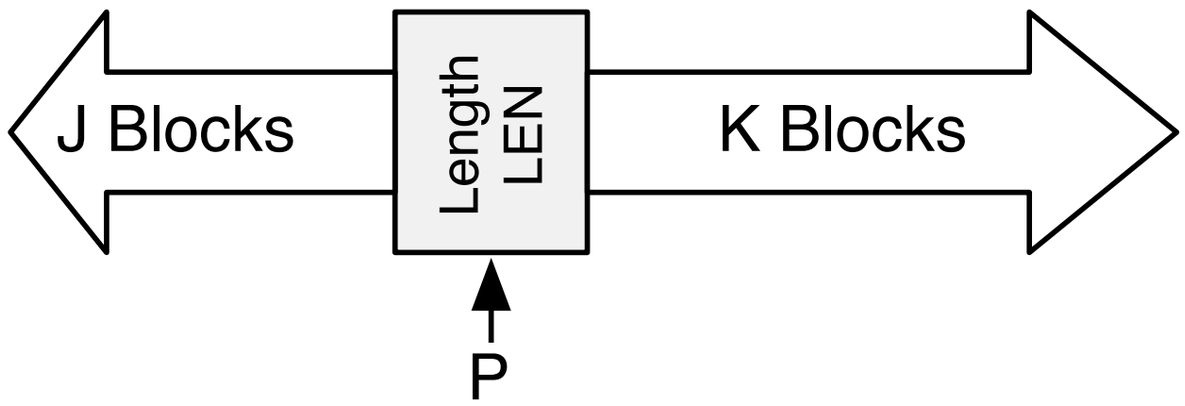


Header/Embedded Length Carving:

- Looks for structures that code length
- Works with MS Office and ZIP files

Header/Length Fragment Recovery Carving:






```
for I in range(0,LEN):
    for J in range(0,LEN-I):
        K = LEN - (I+J)
        data = blocks[S:S+I] + blocks[P-J:P+K]
        if valid(data)==True:  save(data)
```

Carving tools available today:

Open Source:

- PhotoRec - Recovers lost photos from hard drives
- Scalpel - Improved version of Foremost, by Golden G. Richard III
- Foremost - Developed by Jesse Kornblum and Kris Kendall at AFOSI
- bulk_extractor — Feature extractor

Proprietary:

- Adroit Photo Recovery — Amazing, but only works on JPEGs
- DataLifter - File Extractor Pro
- EnCase & FTK — both have limited carving functionality

Stream-based forensics with bulk_extractor

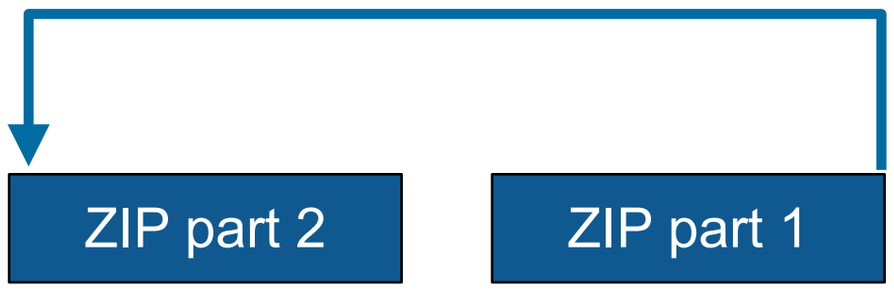
Stream-Based Disk Forensics:
Scan the disk from beginning to end; do your best.

1. Read all of the blocks in order
2. Look for information that might be useful
3. Identify & extract what's possible in a single pass

Advantages:

- No disk seeking
- Read the disk at maximum transfer rate
- Reads *all the data* — allocated files, deleted files, file fragments

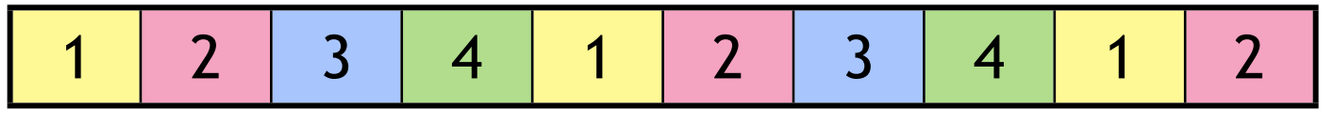
Disadvantages:



- Fragmented files won't be recovered:
  - —*Compressed files with part2-part1 ordering*
  - —*FIles with internal fragmentation (.doc)*
- A pass through the file system is needed to map contents to file names.

bulk_extractor: a high-speed disk scanner.

Written in C, C++ and Flex



- Command-line tool
- Linux, MacOS, Windows (compiled with mingw)

Key Features:

- Uses regular expressions and rules to scan for:
  - *email addresses; credit card numbers; JPEG EXIFs; URLs; Email fragments*
  - *Recursively re-analyzes ZIP components*
- Produces a histogram of the results
- Multi-threaded
- Disk is "striped" into pages
  - *Results stored in mostly-ordered “feature files”*

Challenges:

- Must work with evidence files of *any size* and on *limited hardware*
- Users can't provide their data when the program crashes
- Users are *analysts* and *examiners*, not engineers.

bulk_extractor output: text files of "features" and context.

email addresses from domexusers:

```
48198832  domexuser2@gmail.com    tocol>____<name>domexuser2@gmail.com/Home</name>____
48200361  domexuser2@live.com     tocol>____<name>domexuser2@live.com</name>____<pass
48413829  siege@preoccupied.net   siege) O'Brien <siege@preoccupied.net>_hp://meanwhi
48481542  danilo@gnome.org        Danilo __egan <danilo@gnome.org>_Language-Team:
48481589  gnom@prevod.org         : Serbian (sr) <gnom@prevod.org>_MIME-Version:
49421069  domexuser1@gmail.com    server2.name", "domexuser1@gmail.com");__user_pref("
49421279  domexuser1@gmail.com    er2.userName", "domexuser1@gmail.com");__user_pref("
49421608  domexuser1@gmail.com    tp1.username", "domexuser1@gmail.com");__user_pref("
```

Histogram:

```
n=579   domexuser1@gmail.com
n=432   domexuser2@gmail.com
n=340   domexuser3@gmail.com
n=268   ips@mail.ips.es
n=252   premium-server@thawte.com
n=244   CPS-requests@verisign.com
n=242   someone@example.com
```

bulk_extractor success:
City of San Luis Obispo Police Department, Spring 2010

District Attorney filed charges against two individuals:

- Credit Card Fraud
- Possession of materials to commit credit card fraud

Defendants:

- arrested with a computer
- Expected to argue that defends were unsophisticated and lacked knowledge

Examiner given 250GiB drive *the day before preliminary hearing*

In 2.5 hours Bulk Extractor found:

- *Over 10,000 credit card numbers on the HD (1000 unique)*
- *Most common email address belonged to the primary defendant (possession)*
- *The most commonly occurring Internet search engine queries concerned credit card fraud and bank identification numbers (intent)*
- *Most commonly visited websites were in a foreign country whose primary language is spoken fluently by the primary defendant.*

Eliminating false positives:
Many of the email addresses come with Windows!

Sources of these addresses:

- Windows binaries
- SSL certificates
- Sample documents

```
n=579    domexuser1@gmail.com
n=432    domexuser2@gmail.com
n=340    domexuser3@gmail.com
n=268    ips@mail.ips.es
n=252    premium-server@thawte.com
n=244    CPS-requests@verisign.com
n=242    someone@example.com
```

It's important to suppress email addresses not relevant to the case

Approach #1 — Suppress emails seen on many other drives

Approach #2 — Stop list from bulk_extractor run on clean installs

Both of these methods *white list* commonly seen emails

- Operating Systems have a LOT of emails. (FC12 has 20,584!)
- Is it wise to give Linux developers a free pass?

Approach #3: Context-sensitive stop list.

Instead of extracting just the email address, extract the context:

- Offset: **351373329**
- Email: **zeeshan.ali@nokia.com**
- Context: **ut_Zeeshan Ali <zeeshan.ali@nokia.com>, Stefan Kost <**

- Offset: **351373366**
- Email: **stefan.kost@nokia.com**
- Context: **>, Stefan Kost <stefan.kost@nokia.com>_____sin**

—Here "context" is 8 characters on either side of feature

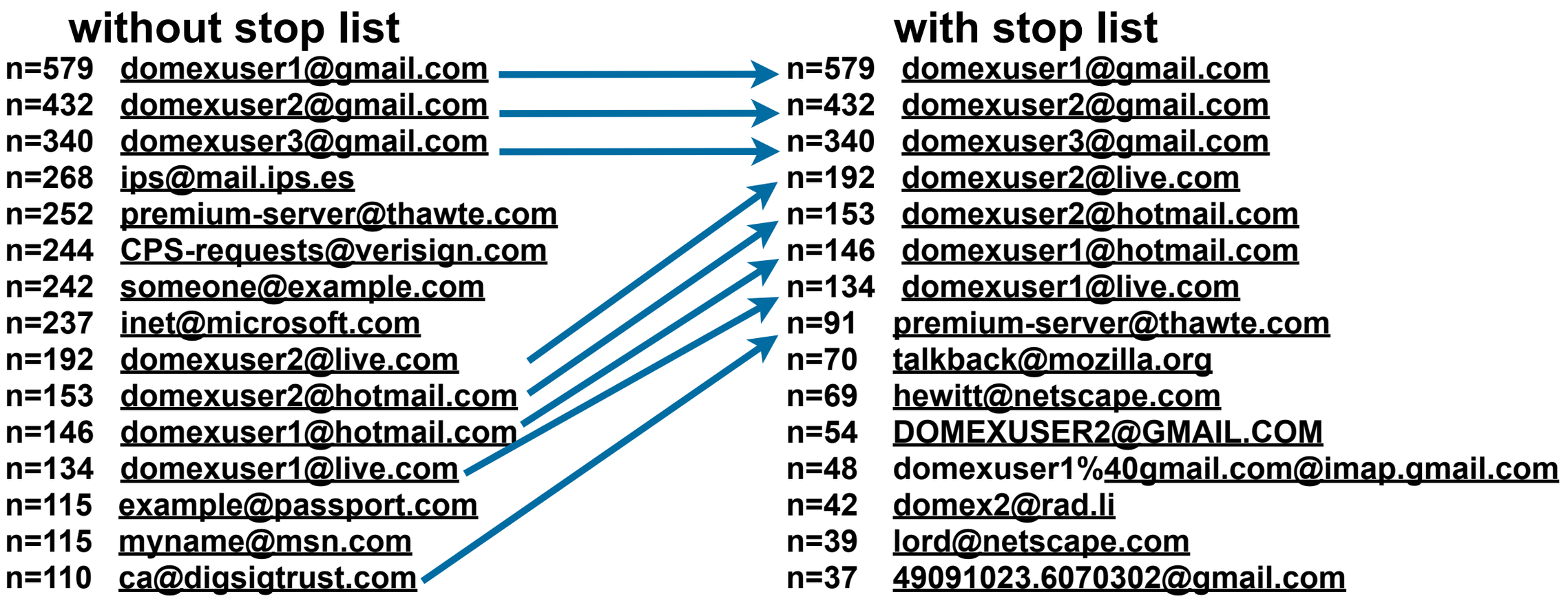
—

We created a context-sensitive stop list for Microsoft Windows XP, 2000, 2003, Vista, and several Linux.

Total stop list: 70MB (628,792 features; 9MB ZIP file)

Applying it to domexusers HD image:

- # of emails found: 9143 ➔ 4459



http://afflib.org/downloads/feature_context.1.0.zip

bulk_extractor: Implemented as a set of C++ classes

Forensic Buffers and Path:

- sbuf_t — Holds data, margin, and forensic path of each page
- pos0_t — Path of byte at sbuf[0]

```
100                     Offset at 100 bytes
100-GZIP-500            At offset 100, GZIP compressed, 500 bytes further in
```

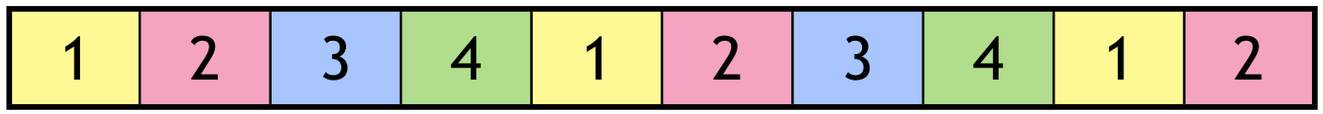
- feature_recorder — Holds output for each feature type

Plug-In Scanner System

- Each scanner is a C++ function that can be linked or loaded at run-time(*)
- Simple scanners look for features in bulk data and report them
  - *scan_accts, scan_aes, scan_bulk, scan_ccns2, scan_email, scan_exif, scan_find, scan_headers, scan_net, scan_wordlist*
- Scanners can instantiate files:
  - *scan_kml*
- Scanners can be recursive.
  - *scan_base64, scan_gzip, scan_hiberfile, scan_pdf, scan_zip*

bulk_extractor: Speed from multi threading

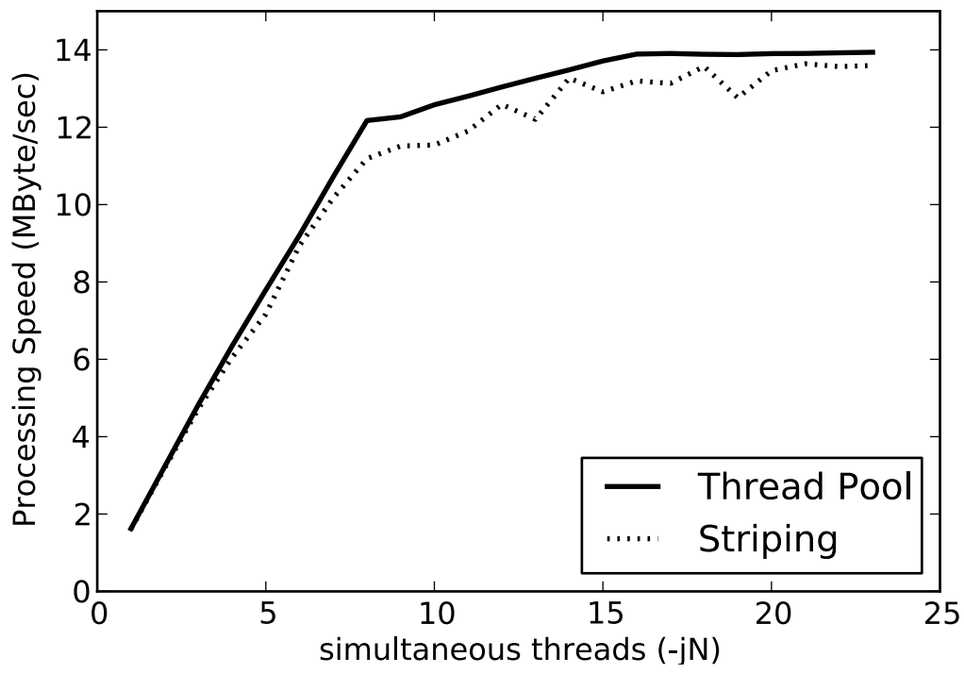
Primary thread:



- Iterator reads “pages” of forensic data and passes each page to a “worker.”
- Iterators available for:
  - *raw & splitraw files*
  - *AFF, E01*
  - *Directory Hierarchies*
  - *MD5 is computed automatically as data is read (source validation)*
- Generates DFXML file with:
  - *Tool compile and runtime provenance.*
  - *Status reports of what is found, errors, etc*

Worker Threads:

- One per core
- Automatically figures out how many cores you have.



Bulk_extractor's magic — opportunistic decompression

Most forensic tools recover:

- allocated files
- "deleted" files
- carving of unallocated area

bulk_extractor uses a different methodology:

- Carving and Named Entity Recognition
- Identification, Decompression and Re-Analysis of compressed data

This helps with:

- hibernation files and fragments (hibernation files move around)
- swap file fragments
- browser cache fragments (gzip compression)

Post-processing the feature files

The feature files are designed for easy, rapid processing

Tab-Delimited

—*path, feature, context*

- Text (UTF-8)

bulk_diff.py: prepares difference of two bulk_extractor runs

Designed for timeline analysis

- Developed with analysts
- Reports “what’s changed.”
  - —*Actually, “what’s new” turned out to be more useful*
  - —*“what’s missing” includes data inadvertantly overwritten*

identify_filenames.py: Reports files responsible for features

Requires DFXML run (fiwalk) for disk image

- Currently a two-step process; could be built in to bulk_extractor

Anti-Forensics: Techniques, Detection and Countermeasures

Simson L. Garfinkel
Naval Postgraduate School

What is Anti-Forensics?

Computer Forensics:

- "Scientific Knowledge for collecting, analyzing, and presenting evidence to the courts" (USCERT 2005)

Anti-Forensics:

- tools and techniques that frustrate forensic tools, investigations and investigators

Goals of Anti-Forensics:

- Avoid detection
  - —*Of a crime*
  - —*Of the anti-forensics tool*
- Attack the forensic tool:
  - —*Reveal the presence of the forensic tool or forensic process*
- Disrupt information collection or increase the examiner's time
- Cast doubt on a forensic report or testimony

Physical destruction is a simple anti-forensic technique.

Overwriting can be used to destroy data.

Overwriting:

- Eliminate data or metadata
- Examples: disk sanitizers; Microsoft Word metadata “washers,” timestamp eliminators

Issues:

- You must guarantee that data is *overwritten* and not simply *remapped*
- *Problem with flash file systems*
- How many times do you need to overwrite?
  - —*DoD standard is 3 times*
  - —*Guttman says 2 times with a modern hard drive*
  - —*Simson says once is enough*
  - —

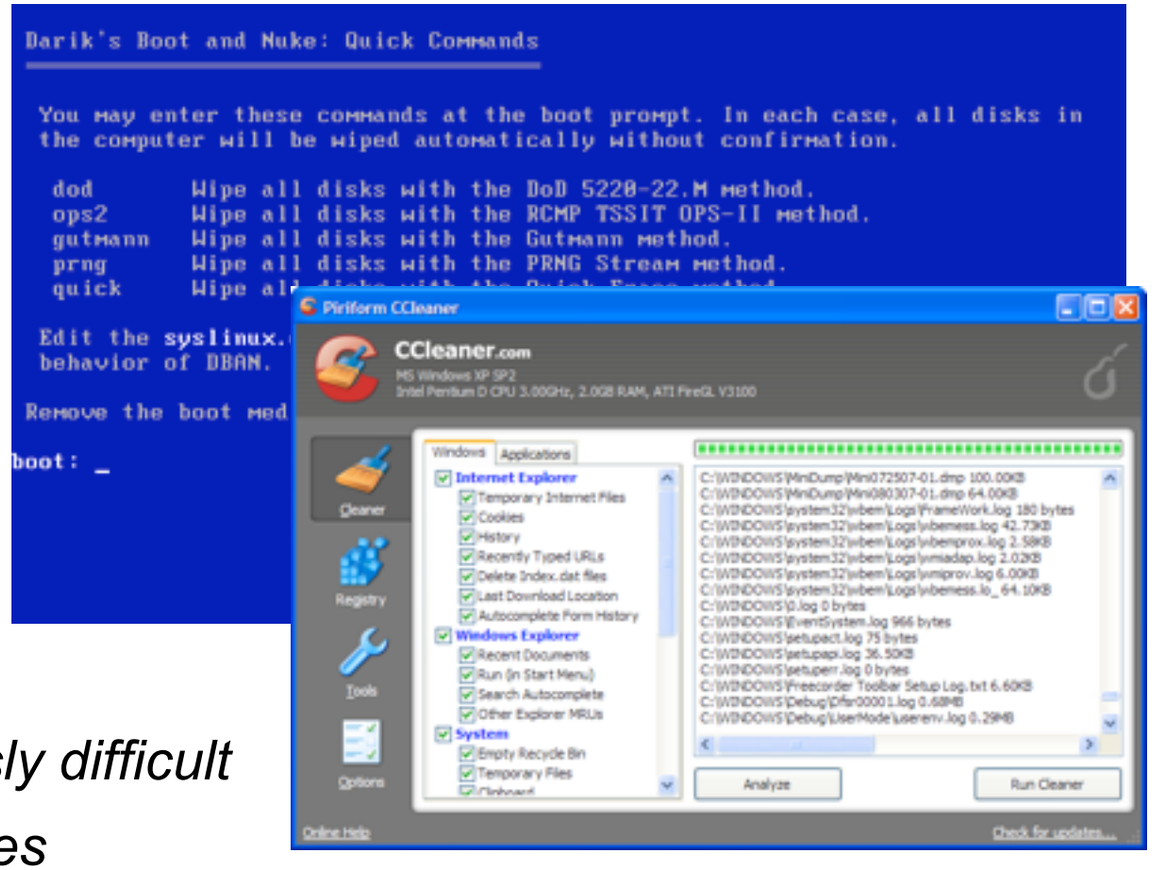
Overwriting: Two Approaches

Overwrite Everything — DBAN — Darik's Boot and Nuke

- http://www.dban.org/
- A single pass is sufficient
- Overwrite what matters

Windows temp files?

- Cookies?
- Pornography?
- http://www.ccleaner.com/
- http://sourceforge.net/projects/eraser/
- http://heidi.ie/eraser
  - —*Overwriting just what matters is notoriously difficult*
  - —*Selective overwriting leaves obvious traces*
  - —*See "Evaluating Commercial Counter-Forensic Tools," Matthew Geiger*
    - http://www.dfrws.org/2005/proceedings/geiger_couterforensics.pdf
    - http://www.first.org/conference/2006/papers/geiger-matthew-papers.pdf



Steganography: Hide data where tools won't look for it.

Data Hiding in File System Structures

- Slacker — Hides data in slack space
- FragFS — Hides in NTFS Master File Table
- RuneFS — Stores data in “bad blocks”
- KY FS — Stores data in directories
- Data Mule FS — Stores in inode reserved space

Data Hiding "out of the map"

- Host Protected Areas (HPA) & Device Configuration Overlay (DCO)
- Bad block areas of hard drives
- Graphics RAM

Approach Two: Cryptography or steganography.

Cryptographic File Systems

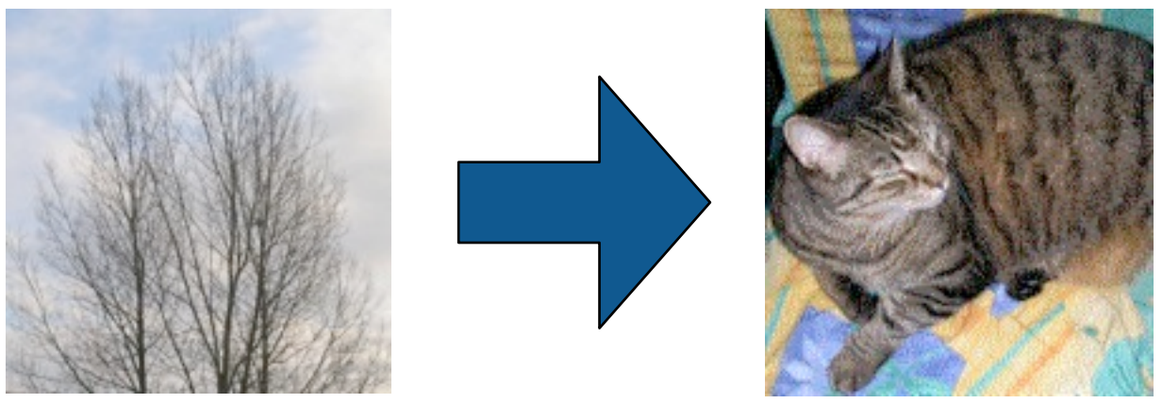
- Built in: FileVault & EFS
- Add-on: BestCrypt, TrueCrypt, FreeOTFE

Encrypted Network Protocols (SSL, SSH, Onion Routing*)

Program Packers (PECompact, Burneye) & Rootkits

Steganography

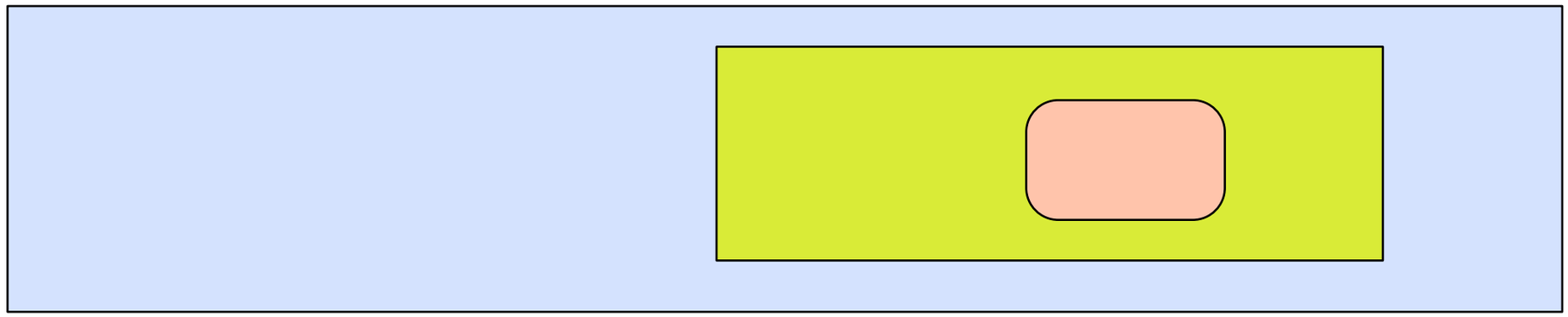
- OpenStego (Images)
- MP3Stego (Music)



***Onion routing also protects from traffic analysis**

Cryptographic File Systems are increasingly a problem for forensic investigators

Many allow hiding an encrypted file system inside an encrypted file system:



- Easy-to-use transparent encryption (FileVault, EFS, IronKey, TrueCrypt) makes crypto easier for both legitimate users and the bad guys
- The law on forcing people to reveal keys is unclear

TrueCrypt has a “deniable encryption” scheme

It’s design to hide the existence of an encrypted volume from someone who might torture you

- I recommend *not using TrueCrypt* because you can’t prove that you aren’t using deniable encryption

Anti-Forensics 3: Minimizing the Footprint

Overwriting and Data Hiding are easy to detect

Tools leave tell-tale signs; examiners know what to look for

- Statistical properties are different after data is overwritten or hidden

AF tools that minimize footprint avoiding leaving traces for later analysis

Memory injection and syscall proxying

- Live CDs, Bootable USB Tokens
- Virtual Machines—VMWare, QEMU, etc
- Anonymous Identities and Storage

Memory Injection and Userland Execve: Running a program without loading the code.

Memory Injection loads code without having the code on the disk

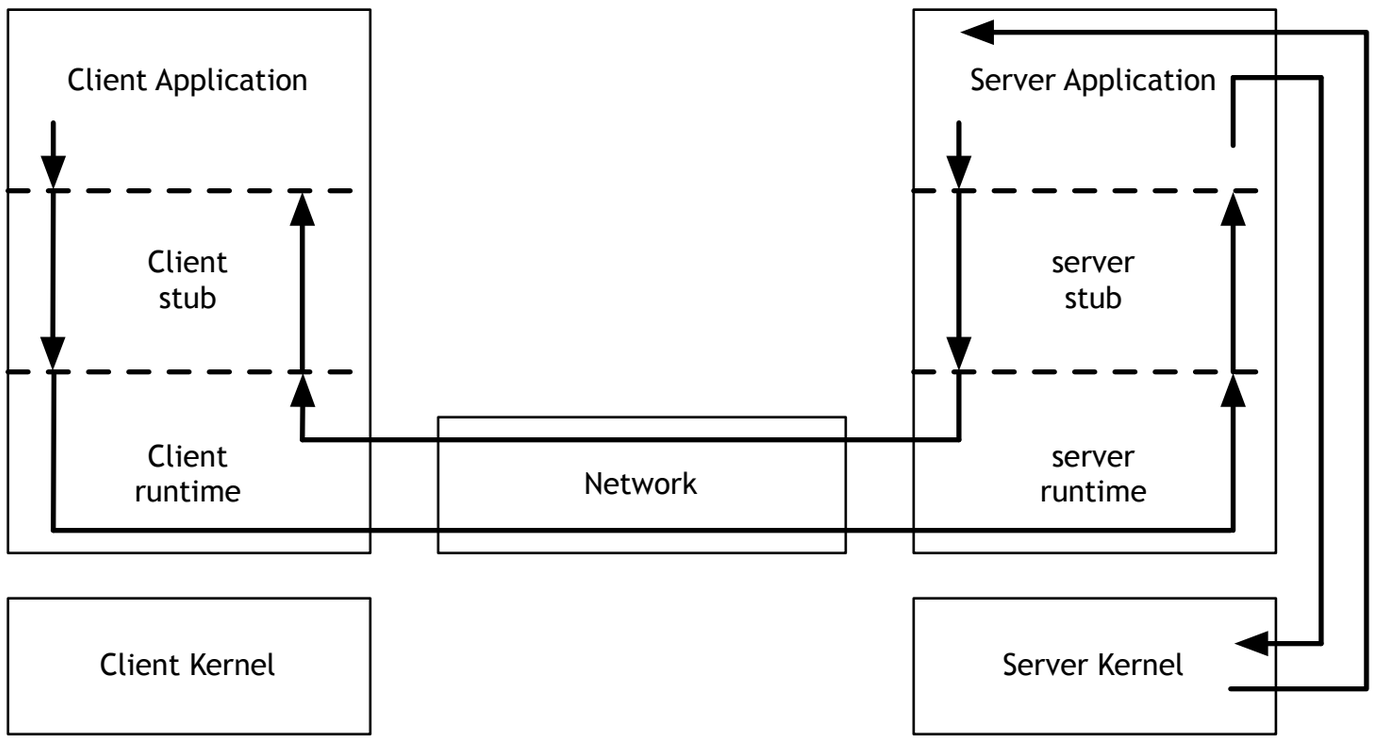
Buffer overflow exploits — run code supplied as (oversized) input

Userland Execve
— Runs program without using execve()
— Bypasses logging and access control
— Works with code from disk or read from network

Syscall proxying:
Running a program without the code!

Syscall Proxying

- Program runs on one computer, syscalls executed on another
- Program not available for analysis
- May generate a lot of network traffic
- Developed by Core Security; used in Impact



Live CDs, Bootable USB Tokens, Virtual Machines: Running code without leaving a trace.

Most forensic information is left in the file system of the running computer

These approaches keep the attacker's file system segregated:
— In RAM (CDs & Bootable USB Tokens)
— In the Virtual Machine file (where it can be securely deleted)

Anonymous Identities and Storage: The attacker's data may be anywhere.

Attackers have long made use of anonymous e-mail accounts. Today these accounts are far more powerful

Yahoo and GMail both have 2GB of storage

- APIs allow this storage to be used as if it were a file system

Amazon's Elastic Compute Cloud (EC2) and Simple Storage Service (S3) provide high-capability, little-patrolled services to anyone with a credit card

- EC2: 10 ¢/CPU hour (Xen-based virtual machines)
- S3: 10 ¢/GB-Month

With BGP, it's possible to have "anonymous IP addresses."

- Announce BGP route
- Conduct attack
- Withdraw BGP address
- Being used by spammers today

Attacking the Investigator: AF techniques that exploit CFT bugs.

Craft packets to exploit buffer-overflow bugs in network monitoring programs like tcpdump, snort and ethereal

Create files that cause EnCase to crash

Successful attacks provide:

- Ability to run code on the forensic appliance
- Erase collected evidence
- Break the investigative software
- Leak information about the analyst or the investigation
- Implicate the investigator

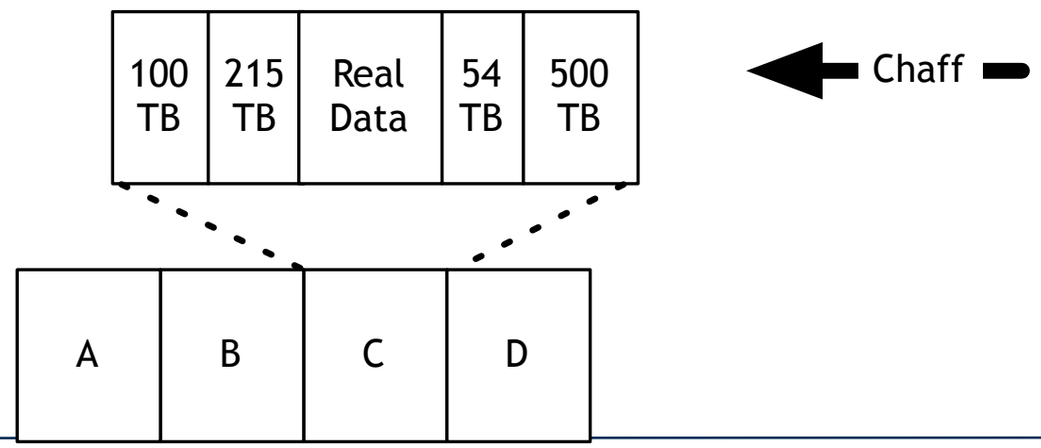
Attacking the Investigator: Denial-of-Service Attacks against the CFT

Any CFT resource whose use is determined by input can be overwhelmed

Create millions of files or identities

- Overwhelm the logging facility
- Compression bombs — 42.zip

The clever adversary will combine this chaff with real data, e.g.:



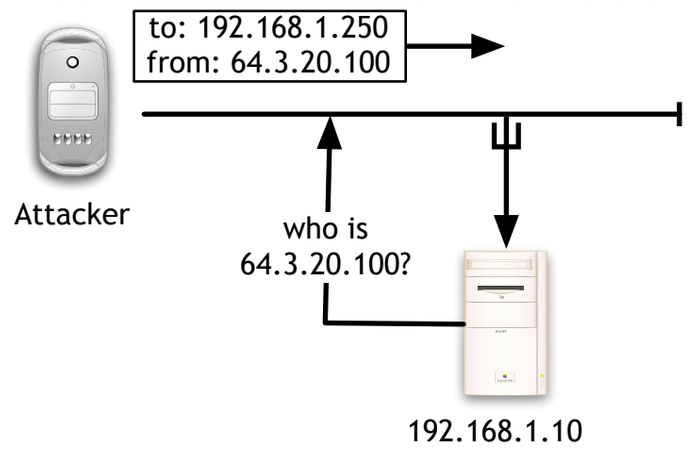
Anti-Forensic Tools can detect Computer Forensic Tools: cat-and-mouse.

SMART (Self-Monitoring, Analysis and Reporting Technology) drives report:

- Total number of power cycles
- Total time hard drive has been on

Network Forensics can be detected with:

- Hosts in “promiscuous” mode responding differently
  - —— *to PINGs*
  - —— *to malformed packets*
  - —— *to ARPs*
- Hosts responding to traffic not intended to them (MAC vs. IP address)
- Reverse DNS queries for packets sent to unused IP addresses



Countermeasures for Anti-Forensics

Improve the tools — many CFTs are poorly written

Save data where the attacker can't get at it:

- — Log hosts
- — CD-Rs

Develop new tools:

- — Defeat encrypted file systems with keyloggers
- — Augment network sniffers with traffic analysis

Research directions in Computer Forensics

Environmental Data Survey Projects

- Phone systems
- Hard drives & data storage devices
- Network hosts and traffic

Theory and Algorithm Development:

- Theoretical basis to forensics (Brian Carrier 2006 PhD)
- Cross-Drive Analysis (Garfinkel)
- Carving Fragmented Objects with Validation

Tool Development

- Easy-to-use tools
- Batch tools
- Data correlation