

Simson L. Garfinkel, Ph.D. Naval Postgraduate School

May 1, 2007 Computers, Freedom & Privacy 2007

### A bit about me

Tech Journalist: 1985--2002

Entrepreneur: 1995--2002

Vineyard.NET

Sandstorm Enterprises, Inc.

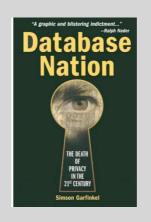
MIT EECS: 2003--2005

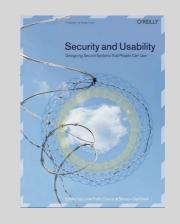
Harvard Center for Research on Computation and Society: 2005-2007

Naval Postgraduate School: 2007-



"Used Hard Drives Reveal Secrets."







### Forensics: Dual Meaning

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





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

(American Heritage Dictionary, 4<sup>th</sup> Edition)

## Computer Forensics: Multiple Meanings



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

  (Computer Forensics: Incident Response Essentials, Warren Kruse and Jay Heiser.)
- 2. "The scientific examination, analysis, and/or evaluation of digital evidence in legal matters."

(Scientific Working Group on Digital Evidence, <a href="http://www.swgde.org">http://www.swgde.org</a>)

So what's digital evidence?

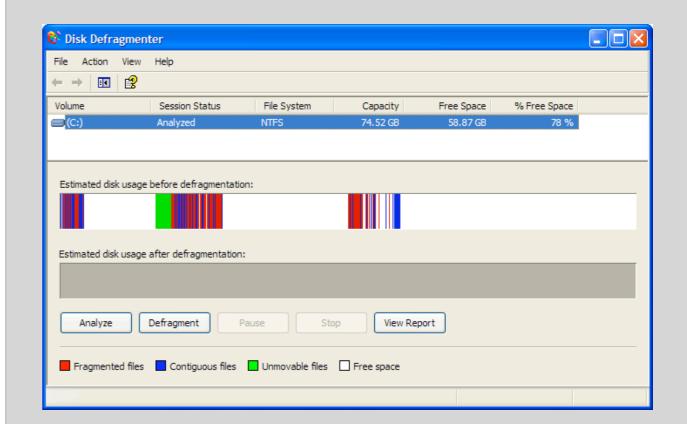
### Computer Forensics is like a magic camera

### Tools can go "back in time..."

- View previous versions of files
- Recover "deleted" files
- Find out what was typed
- Discover visited websites

### Why does this work?

- Computers keep extensive logs
- Most data is not encrypted
- free() doesn't erase memory
- DELETE doesn't erase files
- FORMAT doesn't wipe disks



## Digital Evidence is evidence found in digital systems.

Brian Carrier's PhD has several definitions for digital evidence:

- "Information stored or transmitted in binary form that may be relied upon in court" [Int02]
- "Information of probative value that is stored or transmitted in binary form" [Sci05]
- "Information and data of investigative value that is stored on or transmitted by a computer" [Ass05]
- "Any data stored or transmitted using a computer that support or refute a theory of how an offense occurred or that address critical elements of the offense such as intent or alibi" [Cas04]

All of these definitions assume a **legal process**. Forensics can be used for much more.

## Computer Forensics are typically used after a crime is suspected

### Computer break-ins:

- Determine how a computer was compromised.
- Determine extent of damage

### Make a claim about the computer's owner:

- Possession of contraband information
- Copyright infringement
- Theft of intellectual property
- Confirm/disprove an alibi



### Forensics can also be used for auditing

Evaluate the privacy properties of a system
Understand what's actually going over a network
Audit application performance & security
Spot-check regulatory compliance:

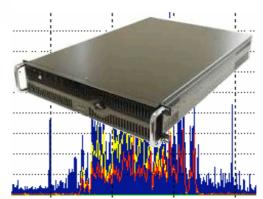
- Data disposal policies
- Data flow across boundaries

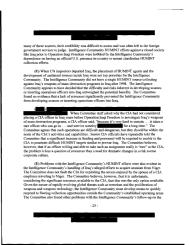
Audit internal information flows

# This tutorial looks at the range of forensic techniques currently in use

- 1. The Forensic Process
- 2. Legal Standards
- 3. Specific Forensic Techniques
  - Disk Forensics
  - Network Forensics
  - Document Forensics
  - Memory Forensics
  - Cell Phone Forensics
  - Software Forensics
- 4. Anti-Forensics
- 5. Civil and Criminal Applications

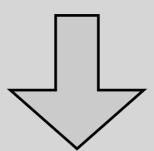






```
printf("%d, %f", i, f);
    i++; f+=3.0;
    g = fmod(f,i);
```







## The Forensic Process

# Computer Forensics turns computer systems into courtroom testimony.

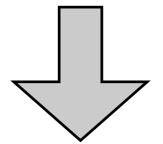
### Five basic steps:

- 1.Preparation
- 2.Collection
- 3.Examination
- 4.Analysis
- 5.Reporting

#### Source:

Electronic Crime Scene Investigation Guide, National Institute of Justice







## Step 1: Preparation Identify potential sources of evidence

Computer system components:

- Hard drives
- Memory / flash / configuration
- Physical configuration

Web Pages on other computers

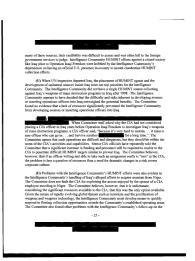
**Files** 

Communication networks











Each source may need its own personnel, tools, training & procedures. One of the most difficult tasks is determining what to include & exclude.

### Step 2:

### Collect and Preserve the evidence

If the activity is ongoing, your choices include:

- Passive Monitoring
- Experimental Probing

If the activity is over, choices include:

- Make a copy
- Seizure

#### Issues to consider:

- What tools are used? Are they validated?
- Is the copy accurate? Is it complete?
- How can you prove that the copy wasn't modified at a later time?



## Step 3: Examination.

### Make evidence "visible" and eliminated excess.

### Disk Analysis:

- Examine partitions and file systems
- Resident files & delete files
- "Slack space" at end of files
- Unallocated space between files

#### File based evidence:

- Document text
- Deleted text
- Metadata (creation date; author fields; etc.)

#### **Network Evidence:**

- Device configuration
- Categorize packets; discard what isn't needed

## Step 4: Analyze to determine "significance and probative value"

Build a hypothesis about what happened.

Look for evidence to prove or disprove hypothesis.

### **Examples:**

- Hypothesis: Suspect is arrested on suspicion of child pornography
- Evidence: Known child pornography on suspect's hard drive

- Hypothesis: Suspect broke into a telephone company computer and stole confidential documents.
- Evidence: Hacker tools; confidential information from telco.

### **BUT:**

### Investigators rarely look for count-evidence

Build a hypothesis about what happened.

Look for evidence to prove or disprove hypothesis.

### Examples:

- Hypothesis: Suspect is arrested on suspicion of child pornography
- Evidence: Known child pornography on suspect's hard drive
- Counter Evidence: Root kit allowing remote access
- Hypothesis: Suspect broke into a telephone company computer and stole confidential documents.
- Evidence: Hacker tools; confidential information from telco.
- Counter Evidence: Documents publicly available

## Step 5: Reporting and Testimony

There are many kinds of testimony:

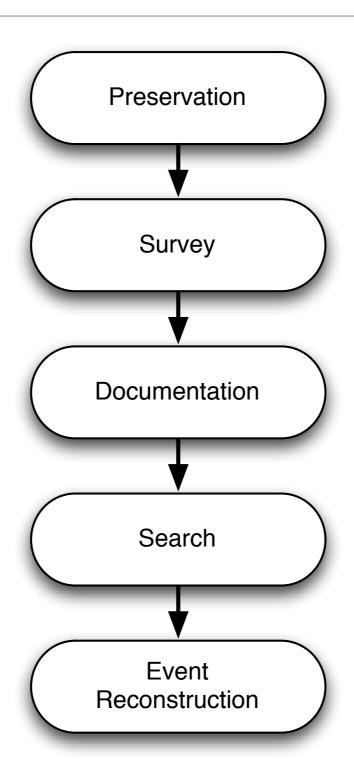
- Written reports
- Depositions
- Courtroom testimony



Testimony needs to include several key points:

- The tools used and procedures that were followed
- What was found
- Examiner's interpretation of what it means

# The Digital Crime Scene Investigation model has five similar steps.



## This isn't really what happens in reality. Instead, investigations are guided by "hypotheses."

Goal of most investigations is to explain evidence that is observed.

- Investigations asked to answer questions about previous states or events.
- Investigator encounters the machine.
- Investigator uses tools to extract and preserve information from machine "Because the observation of the data is indirect, a hypothesis must be formed that the actual data is equal to the observed data"
  - "Hypotheses also need to be formulated about the data abstractions that exist and the previous states and events that occurred."
- The investigator searches for data that supports or refutes the hypotheses.
- Information may be used for confirming/eliminating a hypotheses even if the information itself is inadmissible in court.

A Hypothesis-Based Approach to Digital Forensic Investigations, Brian D. Carrier, PhD Thesis, June 2006



Legal Standards

US Federal Rules of Evidence Daubert

## US Federal Rules of Evidence article VIII 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

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

### 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 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 "Daubert Standard" is supposed to keep "junk science" out of the courts.

Daubert turns federal judges "gatekeepers." Daubert v. Merrell Dow Pharmaceuticals, 509 US 579 (1993)

- Evidence must be "relevant"
- Evidence must be "reliable" (ie, scientific)
  - Subject to peer review (has been published)
  - Generally accepted by the relevant professional community
  - Standards for the technique's operation
  - Known error rate

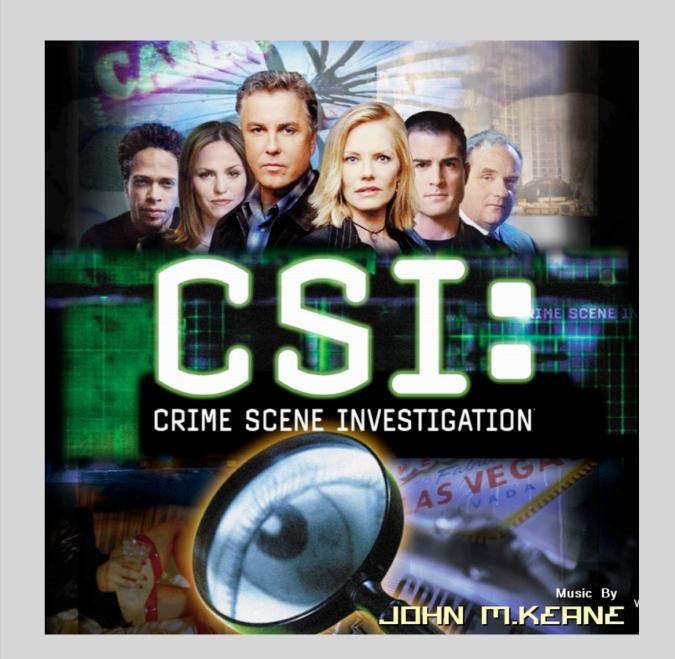
Surprisingly, digital evidence may not meet this standard. [Carrier 2006, pp. 1-4]

# The "CSI Effect" causes victims and juries to have unrealistic expectations.

#### Prosecutors & Jurors:

- Think it's impossible to delete anything.
- Expect highly produced presentations.
- Have no tolerance for ambiguity.

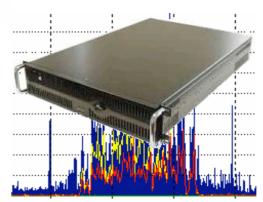
\_

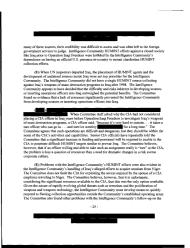


# This tutorial looks at the range of forensic techniques currently in use

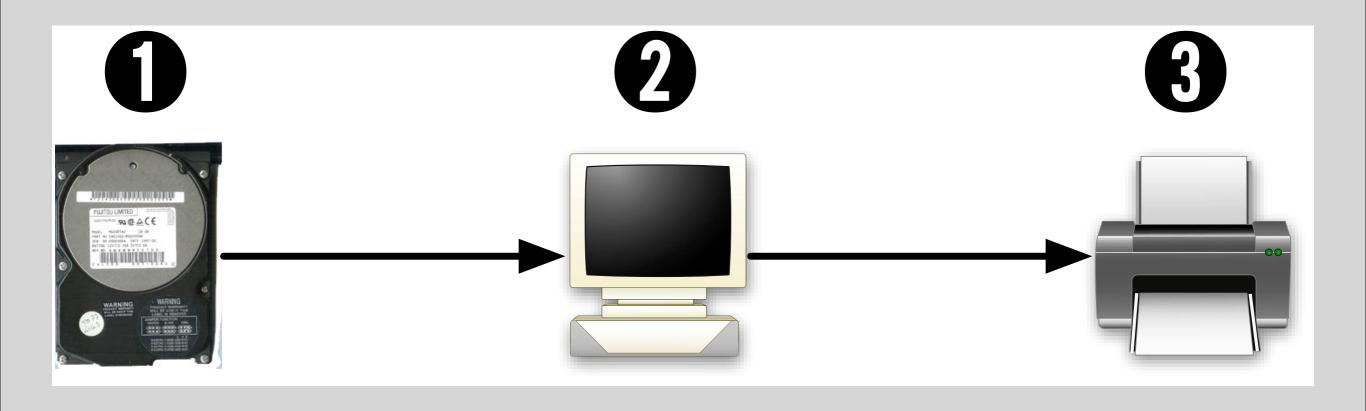
- √ The Forensic Process
- √ Legal Standards
- 3. Specific Forensic Techniques
  - Disk Forensics
  - Network Forensics
  - Document Forensics
  - Memory Forensics
  - Cell Phone Forensics
  - Software Forensics
- 4. Anti-Forensics
- 5. Civil and Criminal Applications







```
printf("%d, %f", i, f);
    i++; f+=3.0;
    g = fmod(f,i);
```



## Disk Forensics

## Hard drive forensics: Typical tasks

#### Recover:

- Deleted files
- Child pornography

#### Recreate:

- Timelines when did the computer do what?
- Flow of information
- Evidence of Inappropriate use

### Gather Intelligence:

- Names of associates
- Meeting places

## Hard drive forensics: Tools of the trade

### Local acquisition:

- Write-Blockers prevent modification
- Create an "image file"

#### Mirror Disks:

Work with a "mirror" of original disk

### Network acquisition

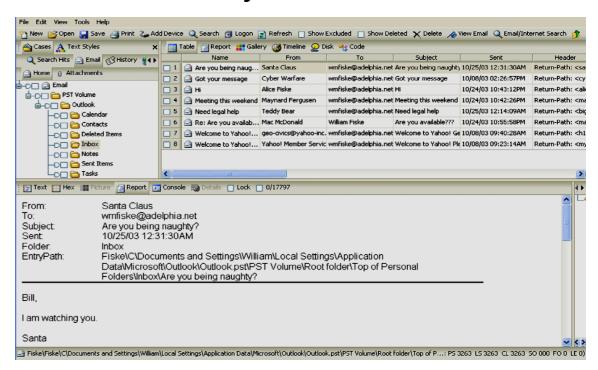
"Encase Enterprise" allows remote forensics on live system

### **GUI-Based Programs:**

- Forensic Tool Kit
- Encase (Guidance Software)
- Forensic Toolkit (Accessdata)







# The important thing about disk imaging: get the data off the suspect drive, onto your drive.

### Imaging options:

- dd if=/dev/hda of=diskfile.img
- aimage /dev/hda diskfile.img
- LinEn

#### Most tools will:

- copy the raw device to a file
- Compute MD5 & SHA1

#### Some tools will:

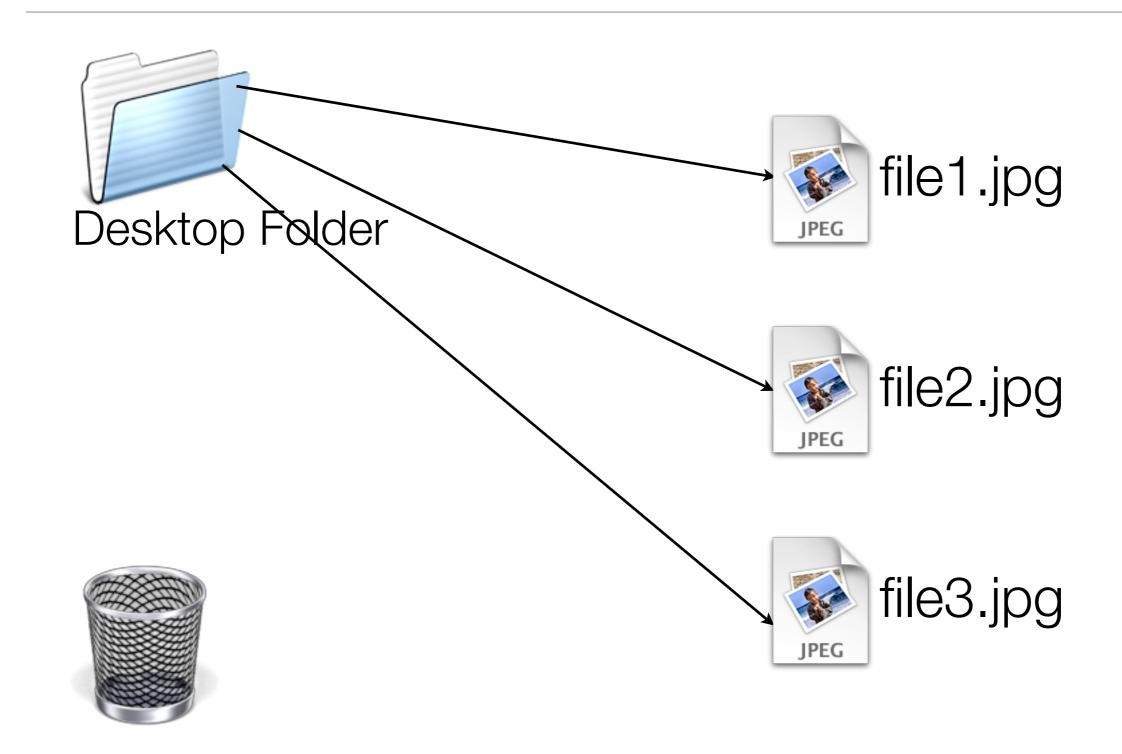
- Compress image
- Capture metadata like s/n
- Record investigative notes

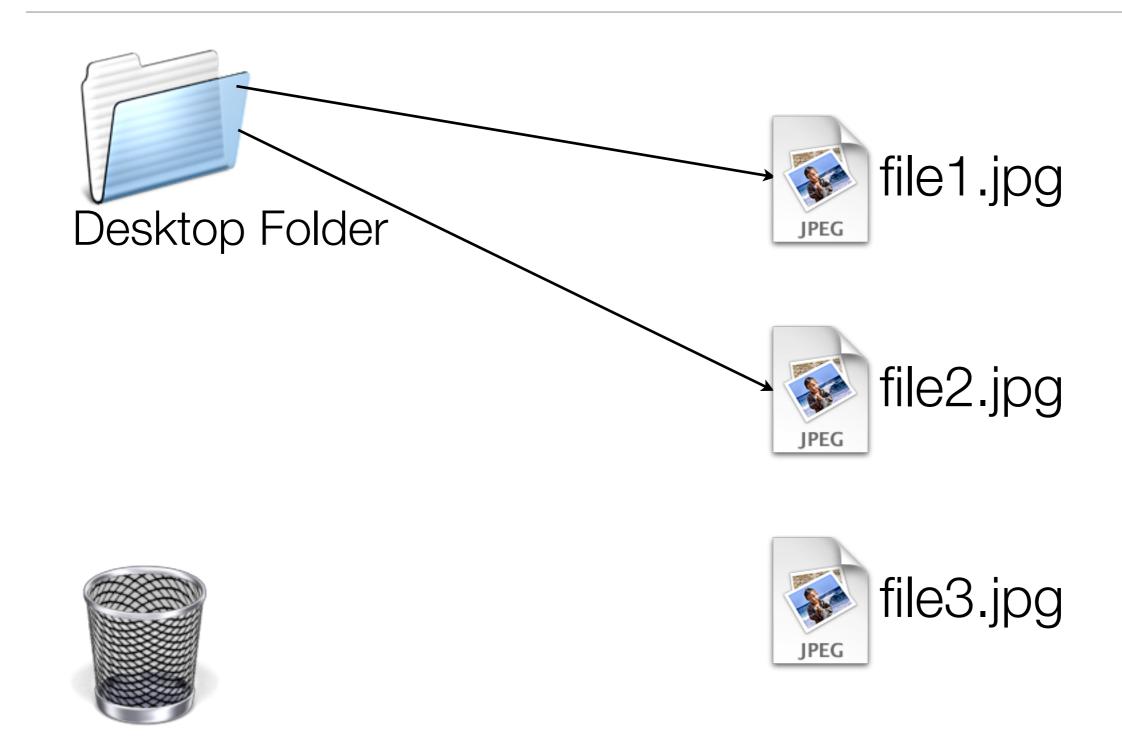


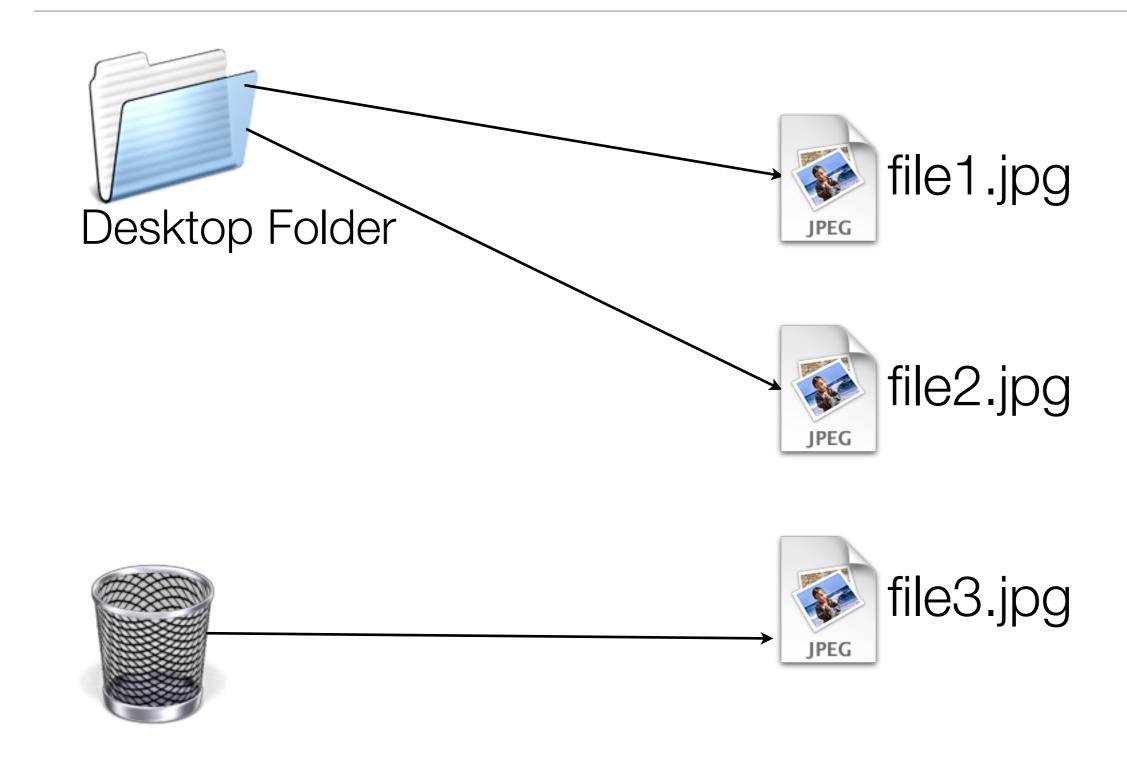
## Once data is imaged, the investigator has many options:

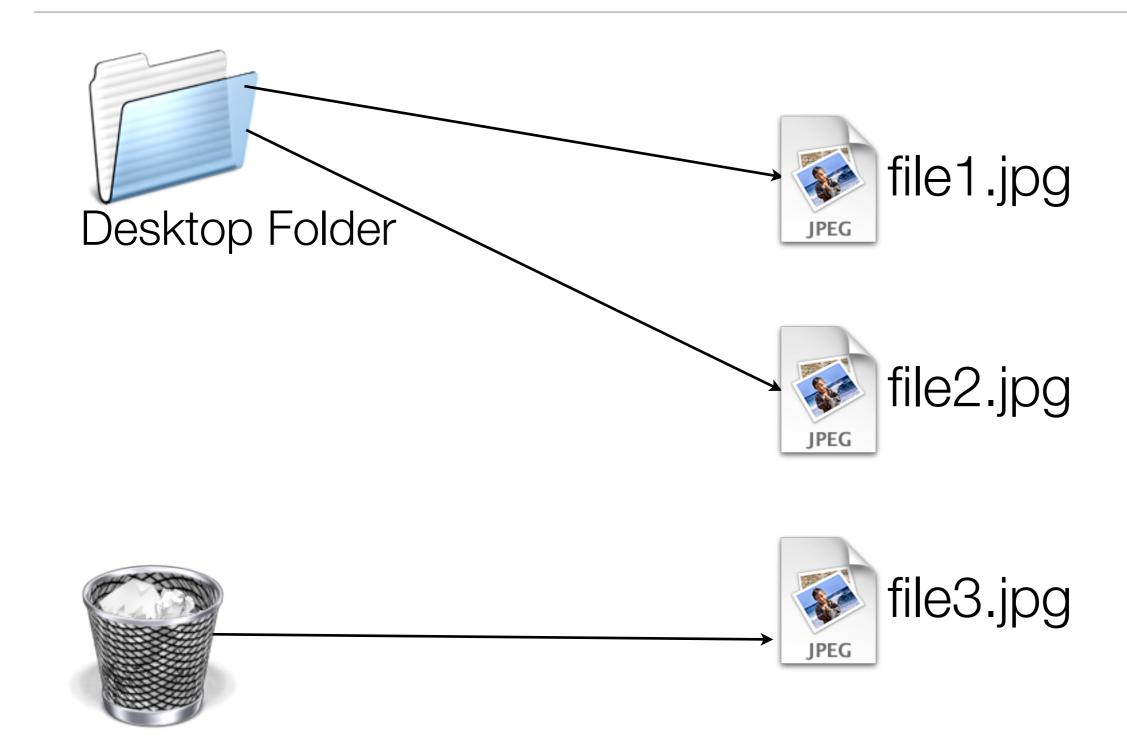
### Typical "first steps" include:

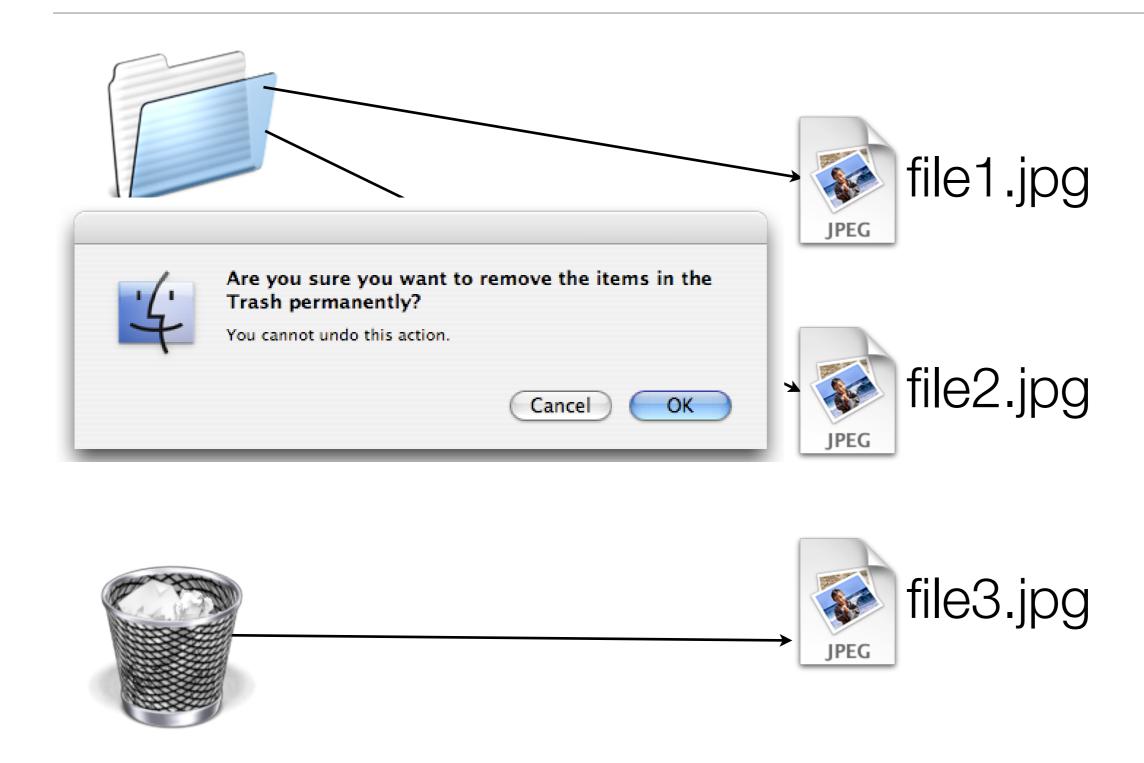
- Inventory all files (resident & deleted) on disk
- Show files modified during a certain time period
- Search disk for files with "known bads" (hacker tools, child porn)
- Scan for key words

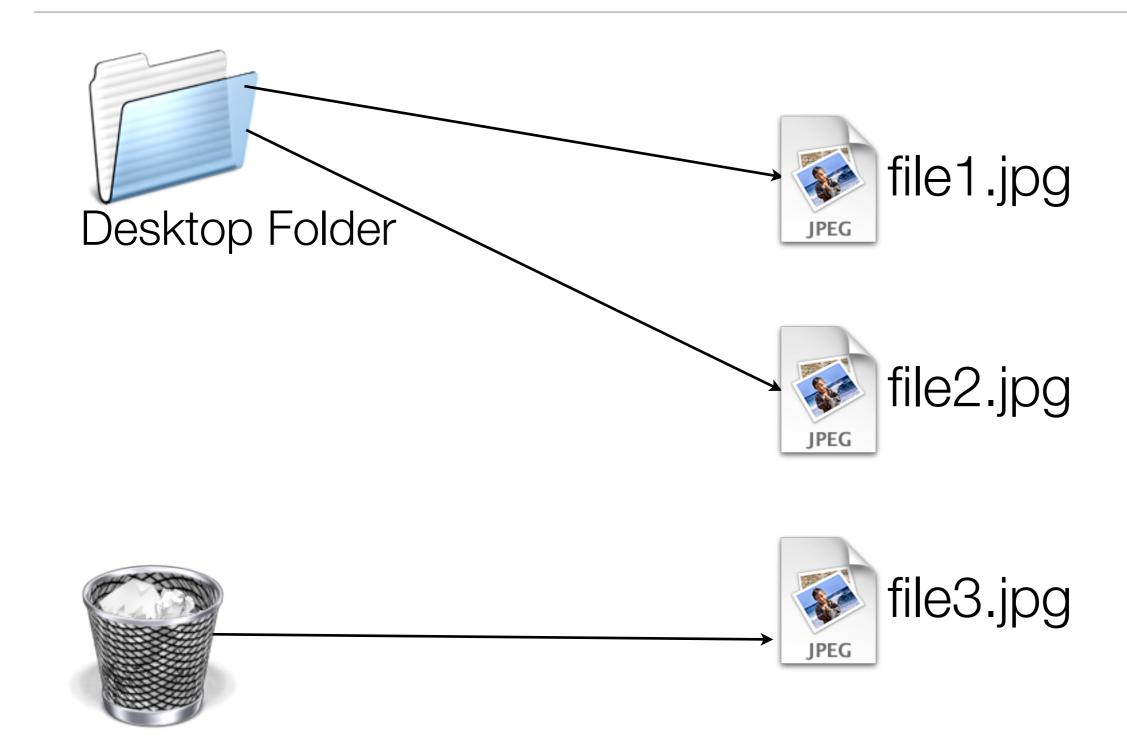


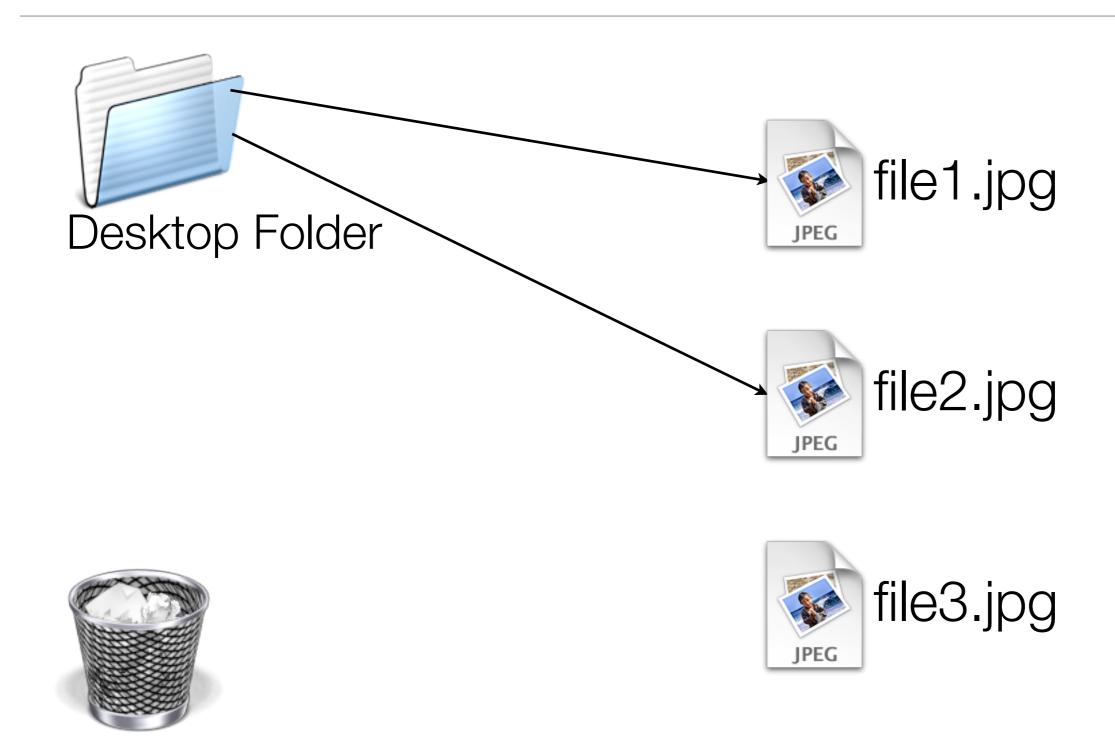


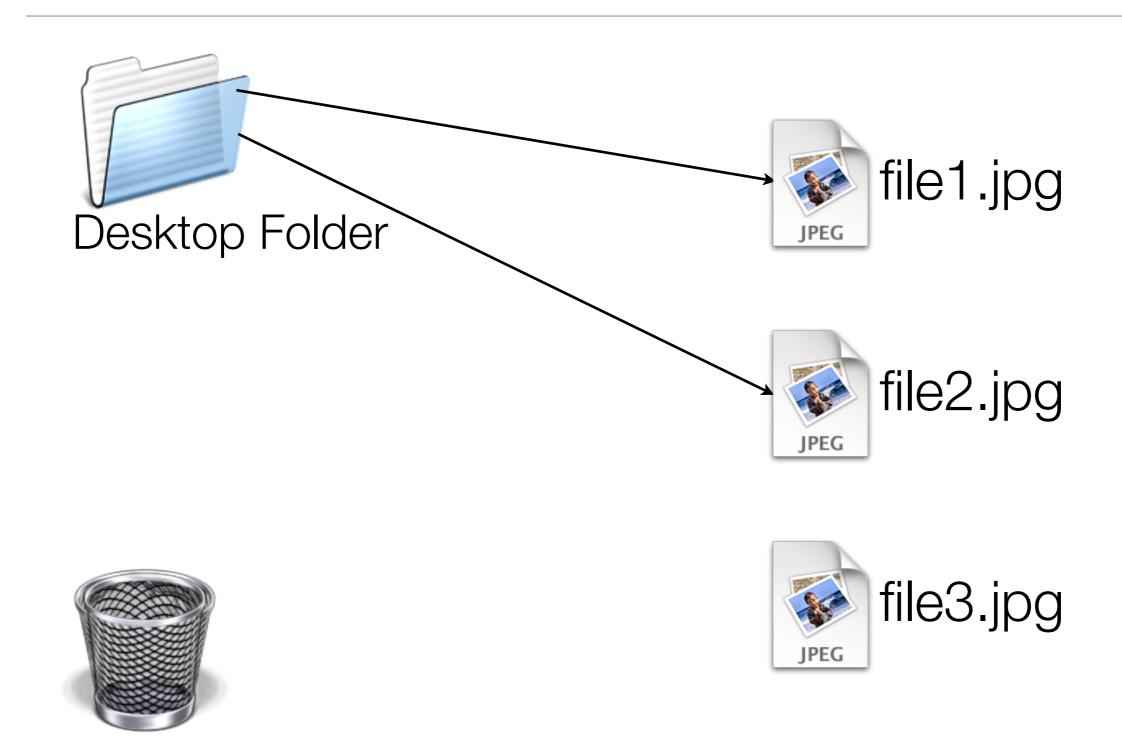


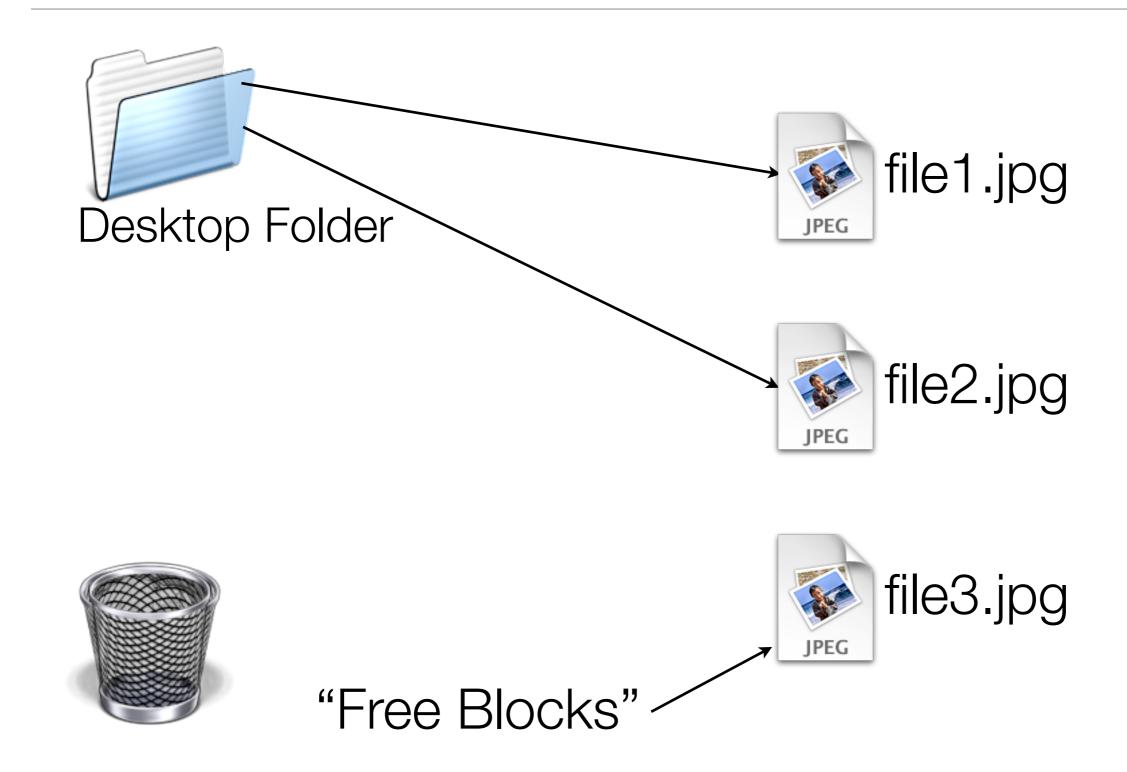




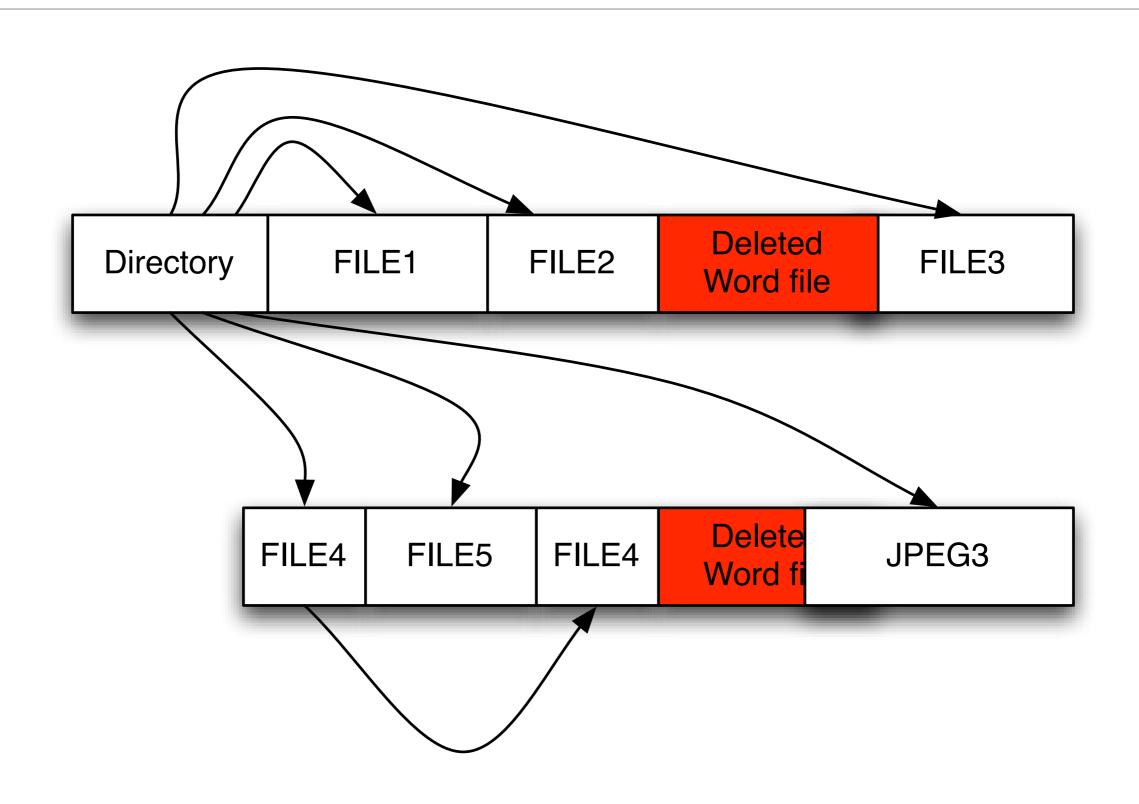




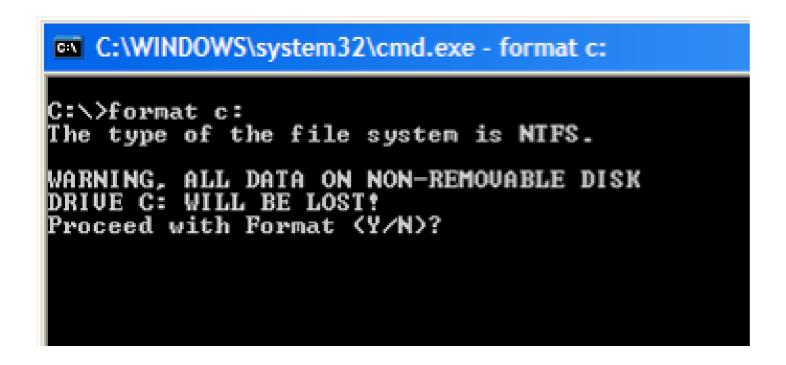


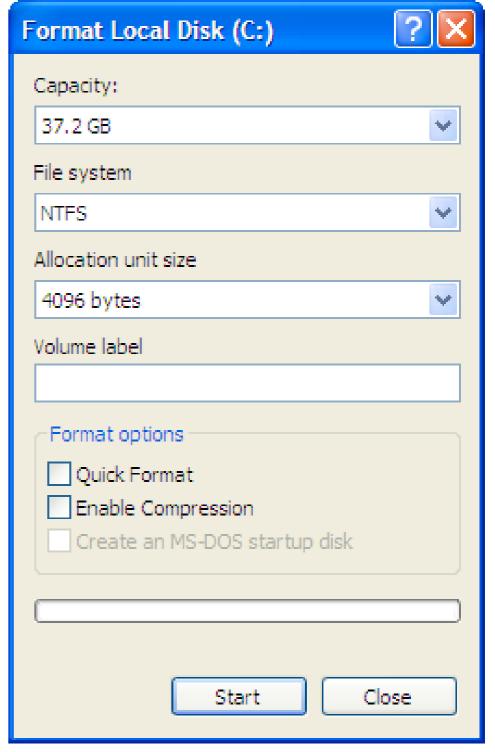


As a result, a typical disk has many kinds of files and data segments on it:



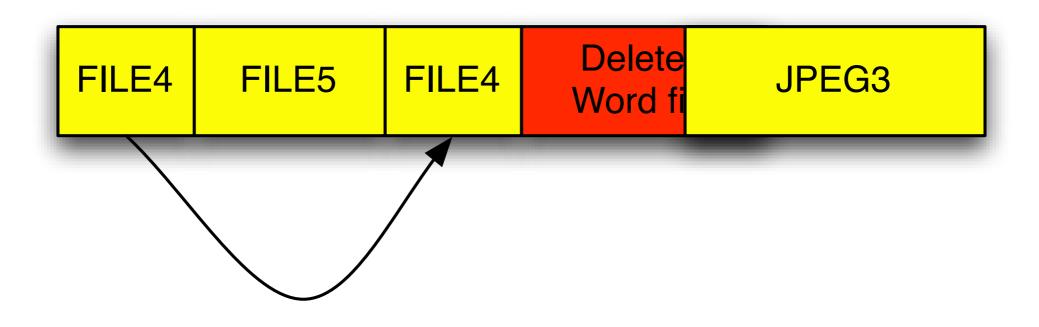
## Formatting a disk just writes a new root directory.





## Formatting a disk just writes a new root directory.





## Example: Disk #70: IBM-DALA-3540/81B70E32

Purchased for \$5 from a Mass retail store on eBay

Copied the data off: 541MB

#### Initial analysis:

• Total disk sectors: 1,057,392

• Total non-zero sectors: 989,514

• Total files: 3

#### The files:

## Image this disk to a file, then use the Unix "strings" command:

```
% strings 70.img | more
Insert diskette for drive
and press any key when ready
Your program caused a divide overflow error.
If the problem persists, contact your program vendor.
Windows has disabled direct disk access to protect your lo
To override this protection, see the LOCK /? command for m
The system has been halted. Press Ctrl+Alt+Del to restart
You started your computer with a version of MS-DOS incompatible
version of Windows. Insert a Startup diskette matching this
OEMString = "NCR 14 inch Analog Color Display Enchanced SV
Graphics Mode: 640 x 480 at 72Hz vertical refresh.
XResolution = 640
YResolution = 480
```

## % strings cont...

ling the Trial Edition

\_\_\_\_\_\_

IBM AntiVirus Trial Edition is a full-function but time-li evaluation version of the IBM AntiVirus Desktop Edition pr may have received the Trial Edition on a promotional CD-RO single-file installation program oveer a network. The Tria is available in seven national languages, and each language provided on a separate CC-ROM or as a separa

EAS.STCm

EET.STC

ELR.STCq

ELS.STC

## % strings 70.img cont...

MAB-DEDUCTIBLE

MAB-MOOP

MAB-MOOP-DED

**METHIMAZOLE** 

INSULIN (HUMAN)

COUMARIN ANTICOAGULANTS

CARBAMATE DERIVATIVES

**AMANTADINE** 

MANNITOL

**MAPROTILINE** 

CARBAMAZEPINE

CHLORPHENESIN CARBAMATE

**ETHINAMATE** 

FORMALDEHYDE

MAFENIDE ACETATE

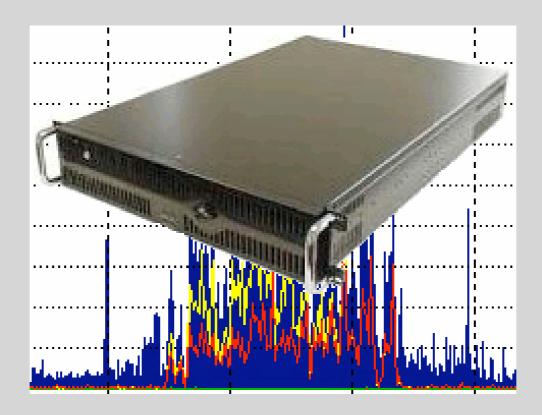
# Roughly 1/3 of the discarded hard drives have significant amounts of confidential data.

From sampling 150 hard drives collected between 1998 and 2002, we found:

- Thousands of credit cards
- Financial records
- Medical information
- Trade secrets
- Highly personal information

[Garfinkel & Shelat 03]

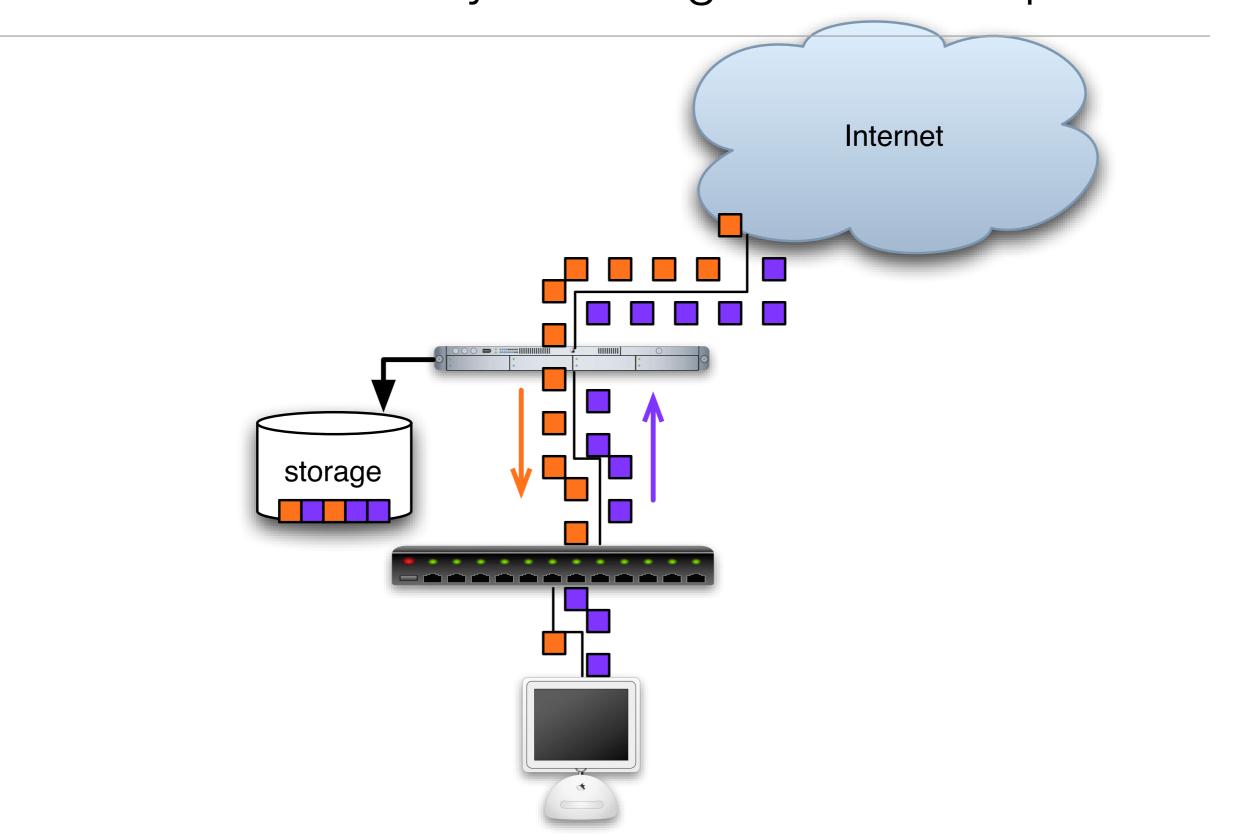




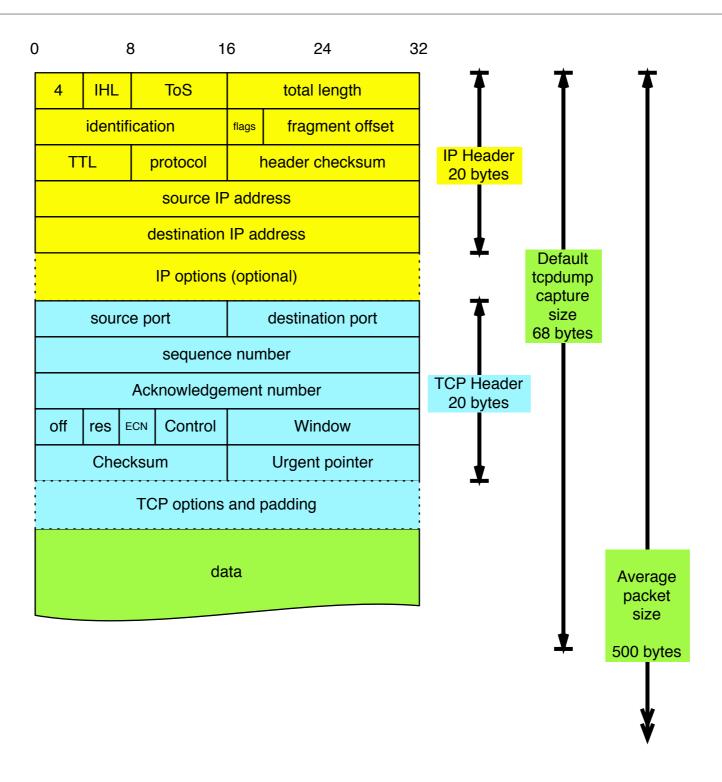
## Network Forensics

packets flows logfiles "Network Forensics" has many different meanings. Internet Capture and Analysis of: packets in flight packets after the fact just packet headers network flows log files storage

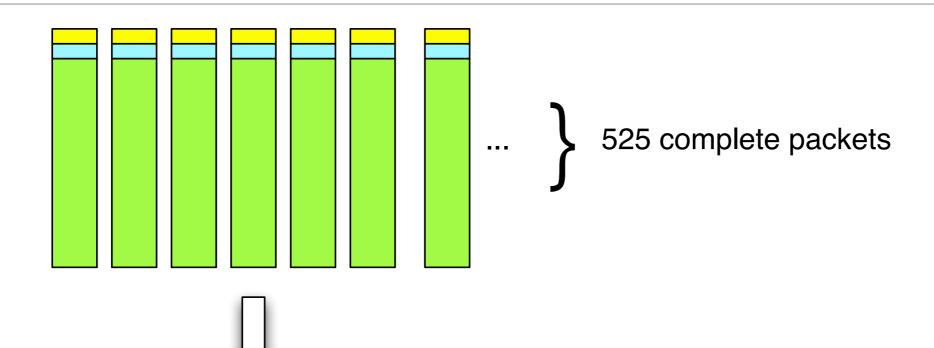
Packets can be analyzed in flight or after capture.



# Systems can capture the entire packet or just the packet header

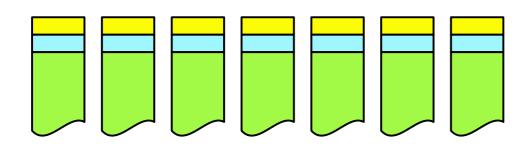


## Complete packets allows for reconstruction.

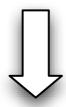




With just headers, you can only get source, destination, size, timestamps, ports, etc.



... } 525 packet headers



```
10:52:16.294858 IP 192.168.1.102.58754 > www2.cnn.com.http: S
10:52:16.370616 IP www2.cnn.com.http > 192.168.1.102.58754: S
10:52:16.370700 IP 192.168.1.102.58754 > www2.cnn.com.http: .
10:52:16.371114 IP 192.168.1.102.58754 > www2.cnn.com.http: P
10:52:16.455120 IP www2.cnn.com.http > 192.168.1.102.58754: .
10:52:19.956986 IP i7.cnn.net.http > 192.168.1.102.58755: .
10:52:19.961475 IP i7.cnn.net.http > 192.168.1.102.58755: .
10:52:19.981228 IP cnn1.dyn.cnn.com.http > 192.168.1.102.58766: 10:52:19.983731 IP cl4.cnn.com.http > 192.168.1.102.58761: P
```

## Packet headers can be used to reconstruct "flows"

```
10:52:16.294858 IP 192.168.1.102.58754 > www2.cnn.com.http: S
10:52:16.370616 IP www2.cnn.com.http > 192.168.1.102.58754: S
10:52:16.370700 IP 192.168.1.102.58754 > www2.cnn.com.http: .
10:52:16.371114 IP 192.168.1.102.58754 > www2.cnn.com.http: P
10:52:16.455120 IP www2.cnn.com.http > 192.168.1.102.58754: .
10:52:19.956986 IP i7.cnn.net.http > 192.168.1.102.58755: .
10:52:19.961475 IP i7.cnn.net.http > 192.168.1.102.58755: .
10:52:19.983731 IP cl4.cnn.com.http > 192.168.1.102.58766:
```



```
Count Source > Destination
 46 i7.cnn.net.http
                        > 192.168.1.102.58755
 34 192.168.1.102.58755 > i7.cnn.net.http
 26 69.22.138.51.http > 192.168.1.102.58776
  24 www2.cnn.com.http
                        > 192.168.1.102.58754
 21 192.168.1.102.58776 > 69.22.138.51.http
 19 192.168.1.102.58765 > i7.cnn.net.http
 17 64.236.29.63.http
                        > 192.168.1.102.58758
 17 192.168.1.102.58754 > www2.cnn.com.http
 16 i7.cnn.net.http
                        > 192.168.1.102.58765
 14 192.168.1.102.58759 > 64.236.29.63.http
 13 72.32.153.176.http > 192.168.1.102.58769
 13 192.168.1.102.58769 > 72.32.153.176.http
 13 192.168.1.102.58758 > 64.236.29.63.http
 12 64.236.29.63.http > 192.168.1.102.58759
 10 64.236.29.63.http > 192.168.1.102.58778
 10 64.236.29.63.http > 192.168.1.102.58757
```

Many switches and routers will report "netfow"

data directly.

Each Cisco NetFlow record contains:

Total bytes & packets

S&D IP addresses

• S&D ports (UDP or TCP)

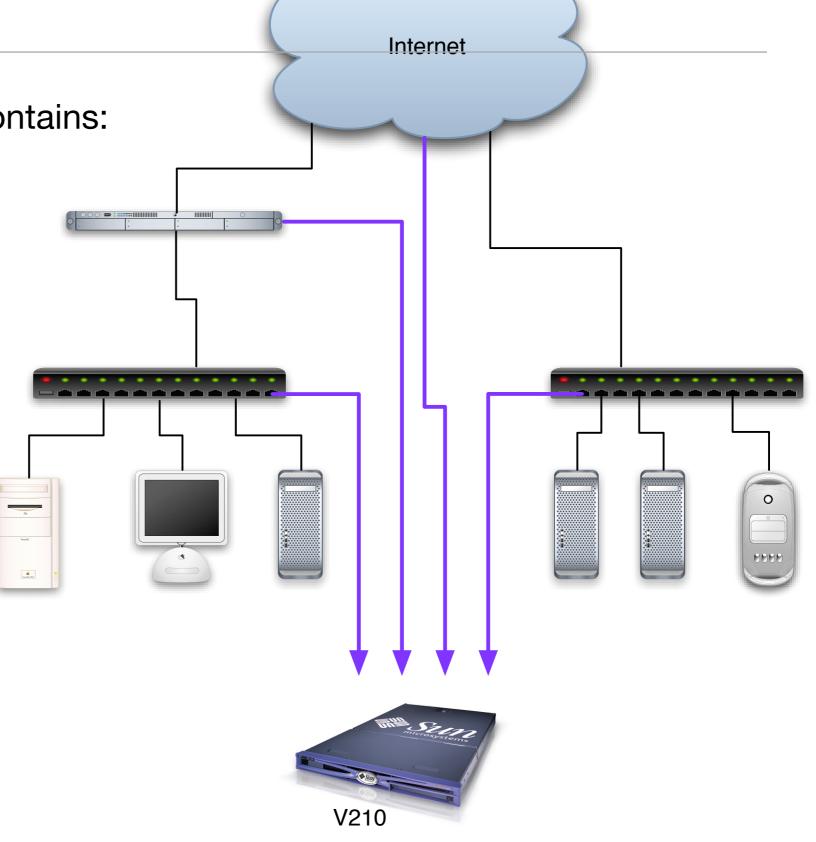
• flags

start & end time

min & max packet size

VLANs & ifaces

Vendor proprietary data



# Each computer and router generates log files. Here's what's on my MacBook:

#### Date & Time of:

- OS installation
- Calendar syncs
- Wake from sleep & time slept
- Every program that crashed
- Every file installed
- Every log-in and log-out

#### Other information:

- Daily amount of free space
- Every 802.11 network found
- Every associated network

\_



Log files are kept on each host; they can be aggregated into a central location

Internet A central repository makes the logs more resistant to attack. Logfile Repository

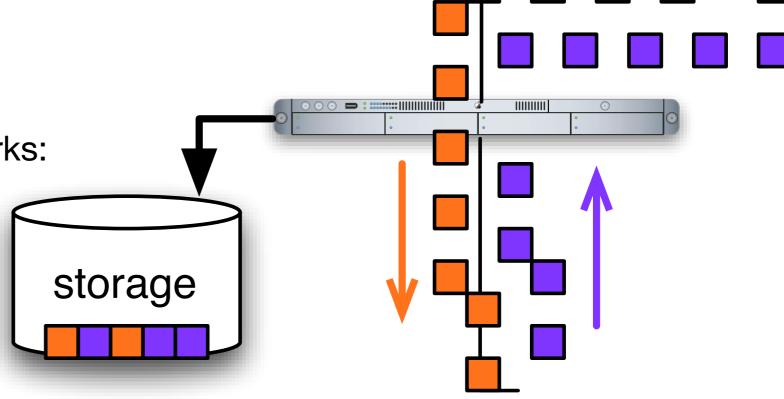
# Some vendors call this "deep packet inspection" or "deep packet analysis."

Primary use is to discover inappropriate data transfer or service use:

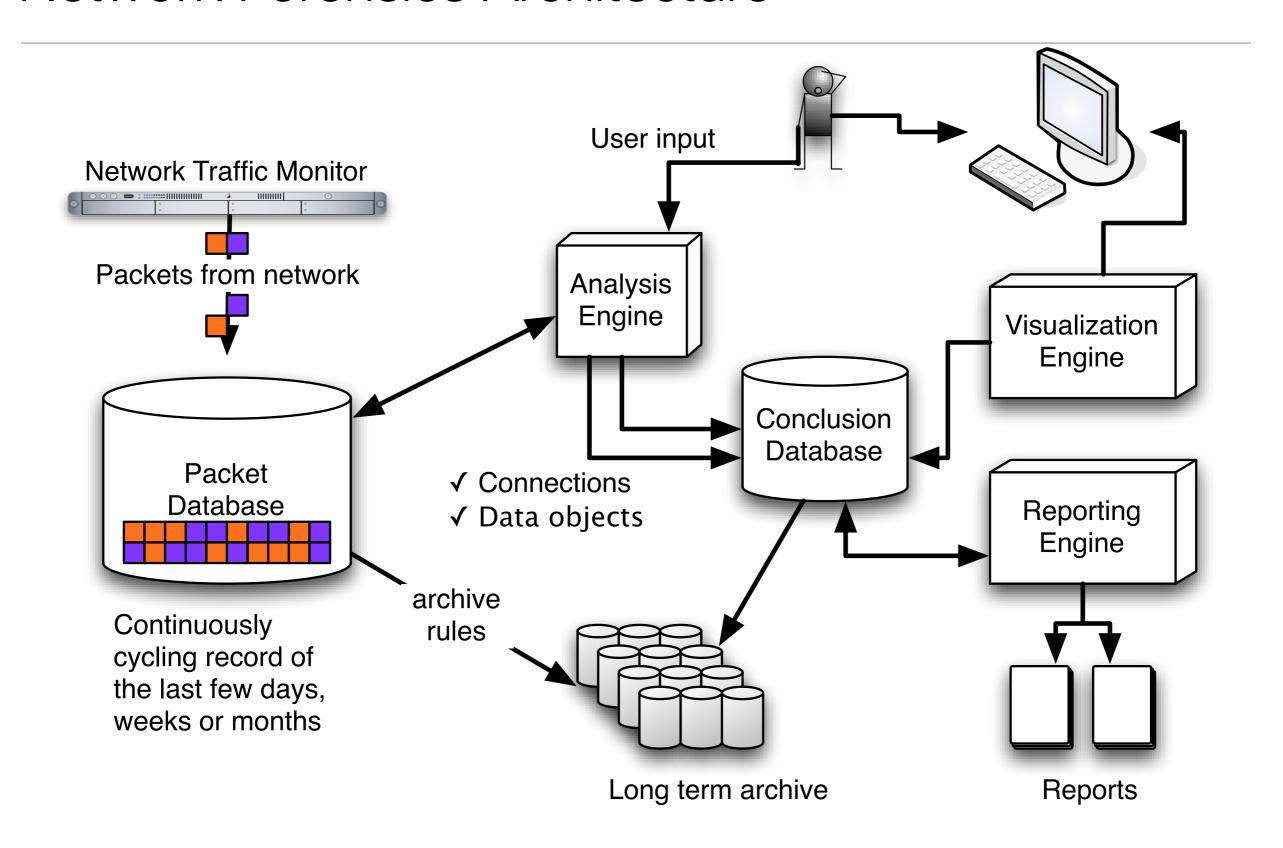
- Use of outside chat or web mail services.
- Leaking protected health Information.
- Restrict information

Also good for debugging networks:

- Duplicate requests
- Incomplete transactions
- Discovery of vulnerabilities without scanning
- Cleartext usernames & passwords



## Network Forensics Architecture



## Packet monitoring is similar to wiretapping.

#### Passive Monitoring Options:

- Use an ethernet "hub" with a packet sniffer.
- Set up a switched monitoring port.
- Full-duplex networks may require two monitoring ports.

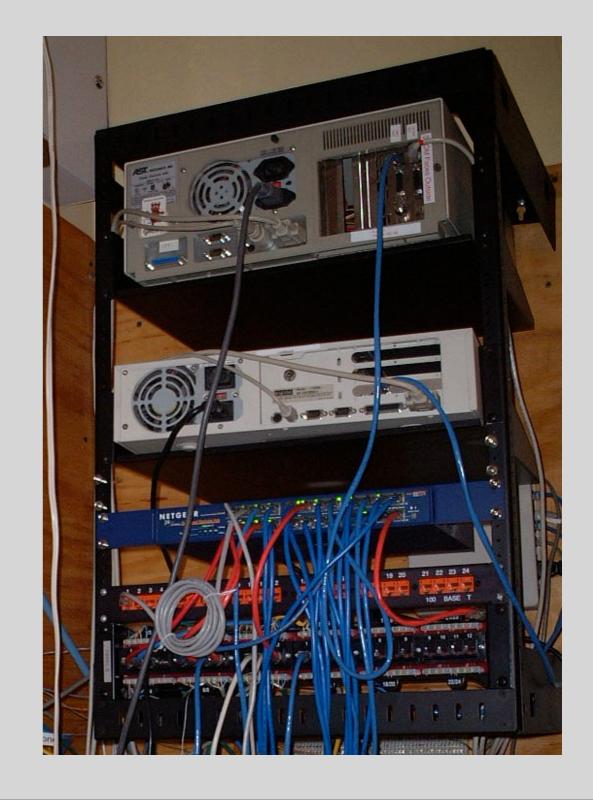
#### **Active Monitoring Options:**

- Monitor with a proxy or router.
- Monitor packets at endpoints

#### Critical uses:

- Attack assessment
- Policy enforcement

"A DVR for an Internet connection."



## Internet Wiretapping History

- 1983 Netwatch Graphical display of Internet Traffic
- 1990 First reports of hostile packet sniffers
- 1995 Ardita (Harvard FAS monitored by FBI)
- 1997 FBI / DOJ / Carnivore
- 1999 Emergence of commercial tools
- 2003 Cisco Systems adds "Lawful Intercept Controls" to switches to allow eavesdropping on VoIP conversations "without detection"
- 2007 FBI reportedly adopts large-scale Internet surveillance techniques.

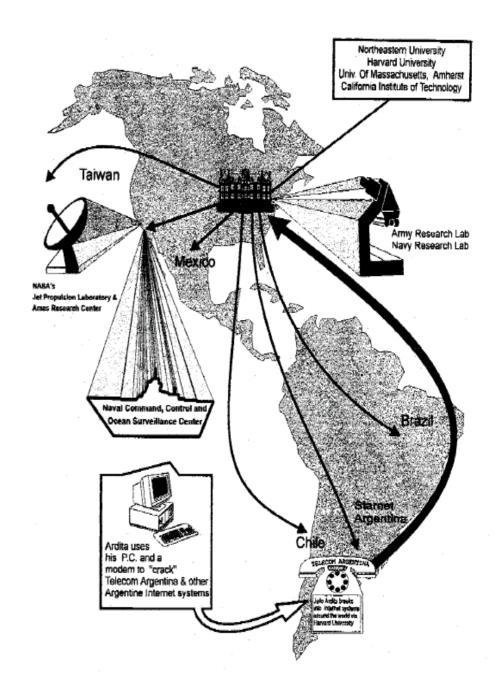
# 1996: Julio Caesar Ardita used Harvard FAS as a jump-off point

From Harvard, Ardita penetrated military and commercial systems throughout the world.

FBI installed TCP/IP stream reassembler with keyword trigger developed by US Army

Details at:

http://www.simson.net/ref/1996/ardita.pdf



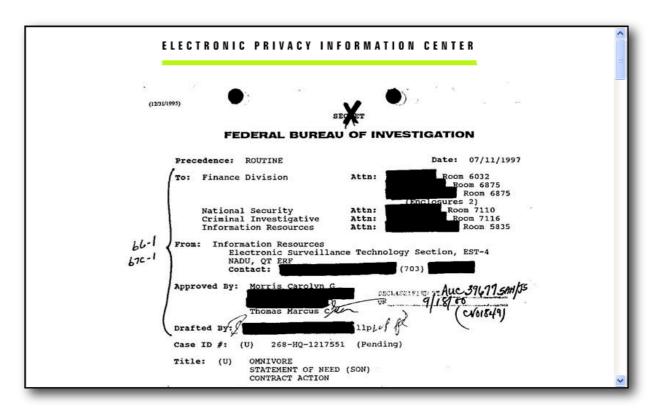
## 1997:

## US Department of Justice develops "Omnivore"

#### Hodge-podge of technologies:

- Monitoring of IP and
- Intercepts stored on ZIP disks
- Solaris X.86
- Triggers on:
  - SMTP username
  - RADIUS

\$2,315,000 development cost



http://www.epic.org/privacy/carnivore/omnivoreproposal.html

# 1998: Omnivore renamed "Carnivore" ("gets at the meat")

### **Targeting Techniques:**

- email usernames, RADIUS username
- IP address, DHCP mac address

### Analysis:

- Logins & Logouts
- Email "pen register" (SMTP & RFC822)
- telnet

Apparently designed for medium-sized dial-up ISPs.

Renamed Digital Collection System 2000 (DSC2000)

Reportedly abandoned in favor of commercial and open source tools

## Is it reasonable to capture all the packets?

In 1991, Los Alamos captured all information in and out of the lab's T1 on DAT tape:

8 gigabytes/day (50%)

Disks have gotten bigger faster than network connections have gotten faster.

This is an engineering problem.

Connection	GB/Day (50%)
T1	8 GB
10 Mbit	54 GB
T3	170 GB
OC3	512 GB
OC12	2,000 GB

## Network Forensic issues:

### Scaling issues:

- Amount of data
- Quality of data (lost packets)

### Analysis issues:

- String search
- Correlation

### Ultimate goal of work:

- Reconstruction
- Exploration

#### Vendor:

- Open Source
- Commercial

## Full-content "deep analysis" solutions:

#### **Open Source**

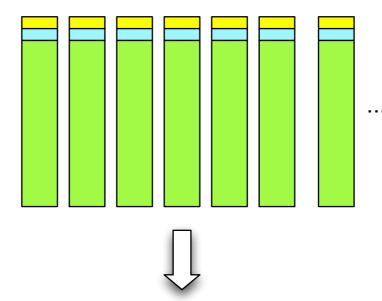
- Wireshark
- Snort
- Squil

#### Commercial in-memory:

- NFR
- Intrusic
- McAfee
- NetWitness

#### Commercial archiving systems:

- CA eTrust Network Forensics
- Chronicle Solutions
- NIKSUN NetDetector
- Sandstorm NetIntercept
- Network Intelligence



525 complete packets



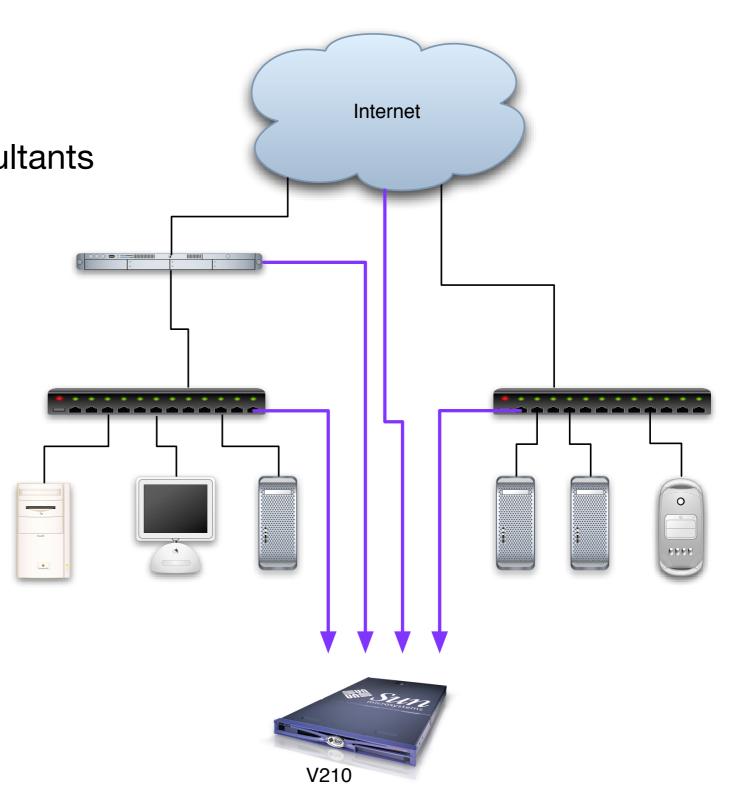
## Flow-based systems: "blind" to data

### Advantages:

- More economical
- Finds rogue servers and consultants

#### Can't discover:

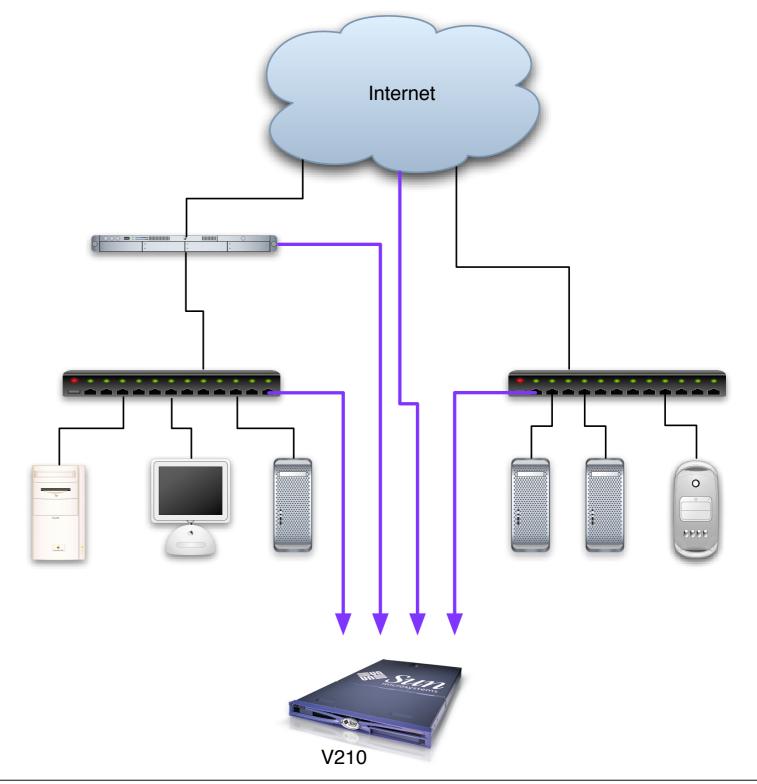
- Missing encryption
- Inappropriate encryption
- Protocols on wrong ports



## Flow-based vendors

Arbor Networks
GraniteEdge Networks
Lanscope
Mazu Networks
Q1 Labs

...and many more



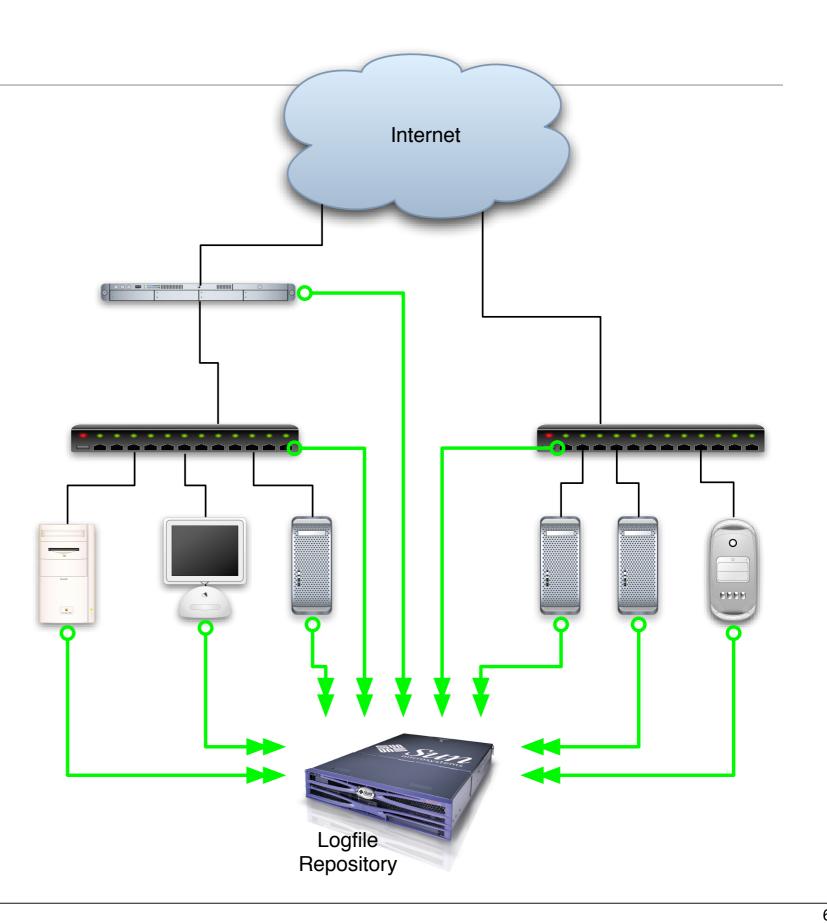
## Log files: options

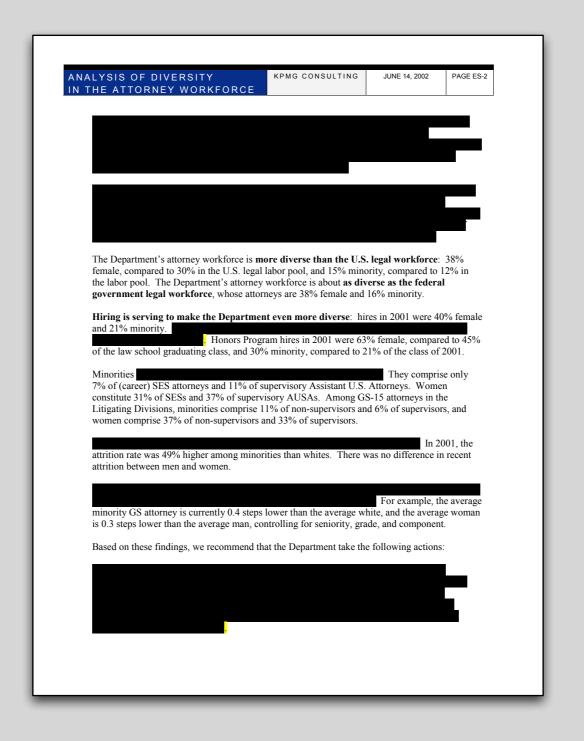
### Open Source Options:

syslog

### Commercial Options:

- LogLogic
- Netforensics
- Q1 Labs
- Many other options...





#### **Document Forensics**

#### Uses for Document Forensics

Which computer generated this document?

Who edited this document?

What was changed? When?

Is this document "authentic?"

## Approaches for Data and Document Analysis:

#### Look for hidden data:

- Deleted information; previous versions
- GIDs embedded in Microsoft Word document

#### Look for characteristic data:

- Indicates authorship
- Indicates program used to create document.

#### Look for inconsistent data:

Indicates possible tampering.

# Privacy and Security violations result when improperly sanitized documents are released.

#### Adobe PDF files:

- The New York Times published a PDF file containing the names of Iranians who helped with the 1953 coup. (2000) (<a href="http://cryptome.org/cia-iran.htm">http://cryptome.org/cia-iran.htm</a>)
- Multinational Force-Iraq report (2005)

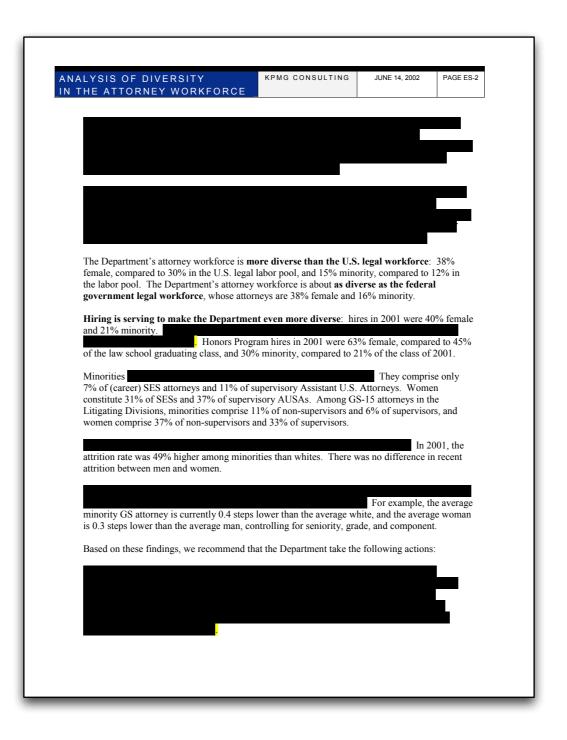
#### Microsoft Word Files:

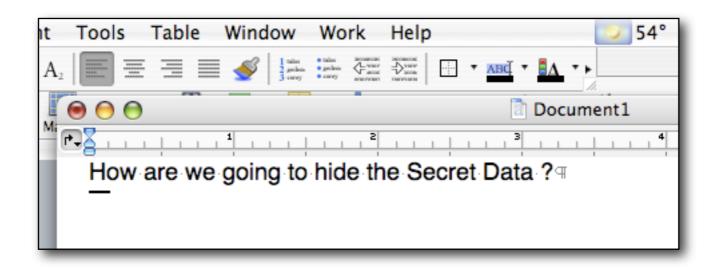
- SCO Word file revealed its anti-Linux legal strategy. (2004)
- Intelligence report by Blair Government was found to be plagiarized from a postgraduate student at the Monterey Institute of International Studies based on transaction log (2003) (http://www.computerbytesman.com/ privacy/blair.htm)

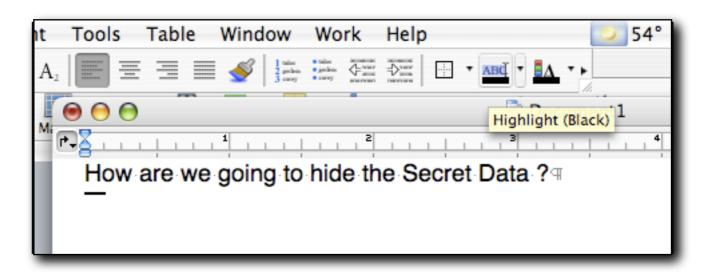
## Why is data left in documents?

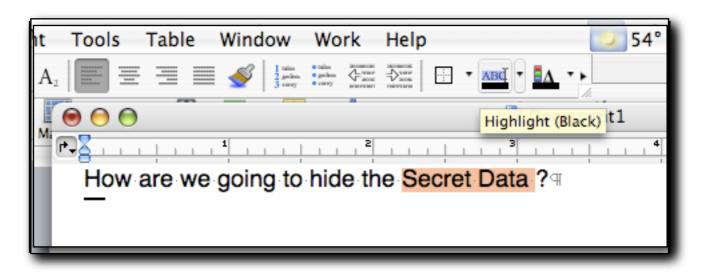
- 1. Confusion between "covering data" and removing it.
- 2. Failure to implement "complete delete."
- 3. Information that is written but never read.

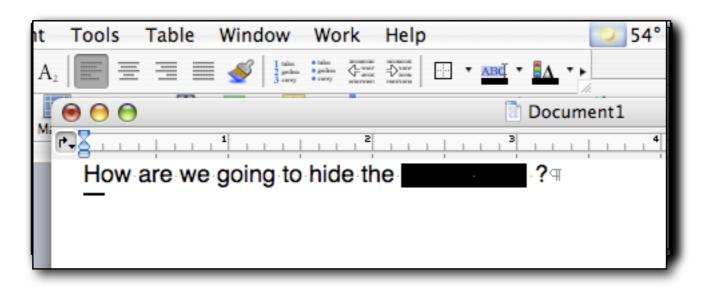
## Most Acrobat leakage is a result of Microsoft Word.











When Microsoft Word generates the PDF file, "Secret Data" is covered with the black box



## Tools for recovering hidden data in Acrobat files:

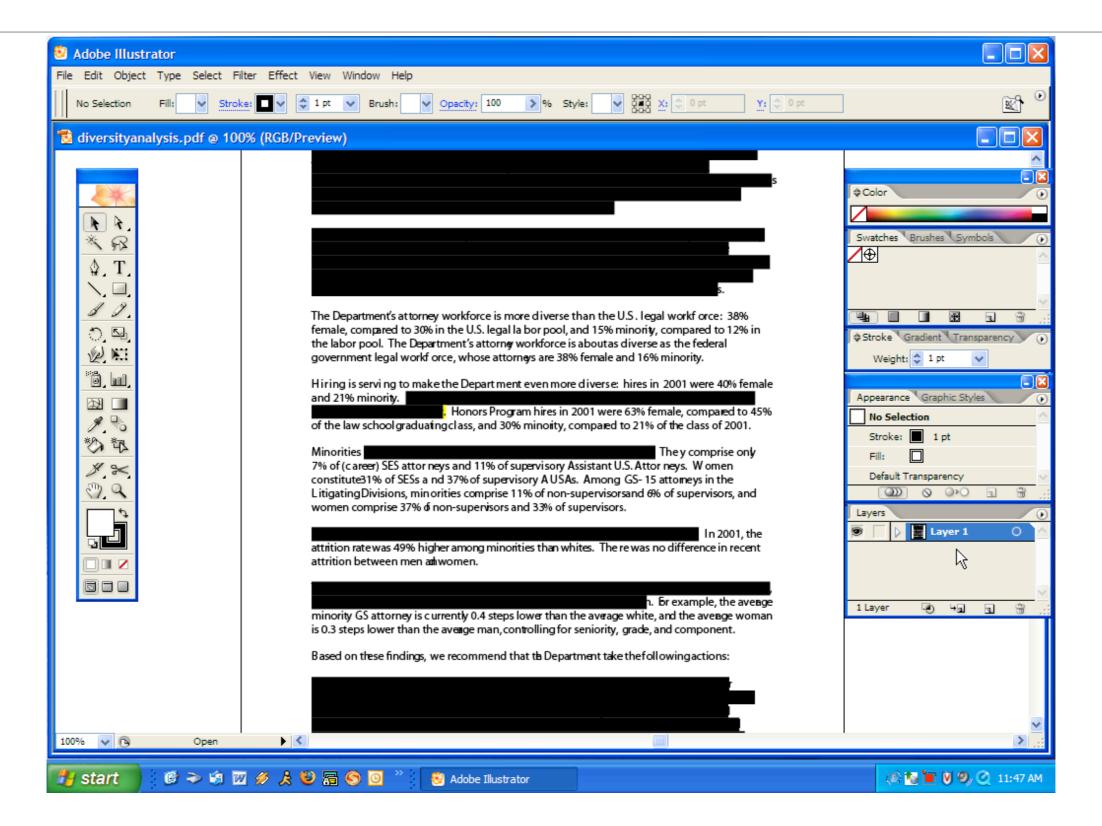
#### Adobe Illustrator

- Move the boxes
- Turn the boxes yellow

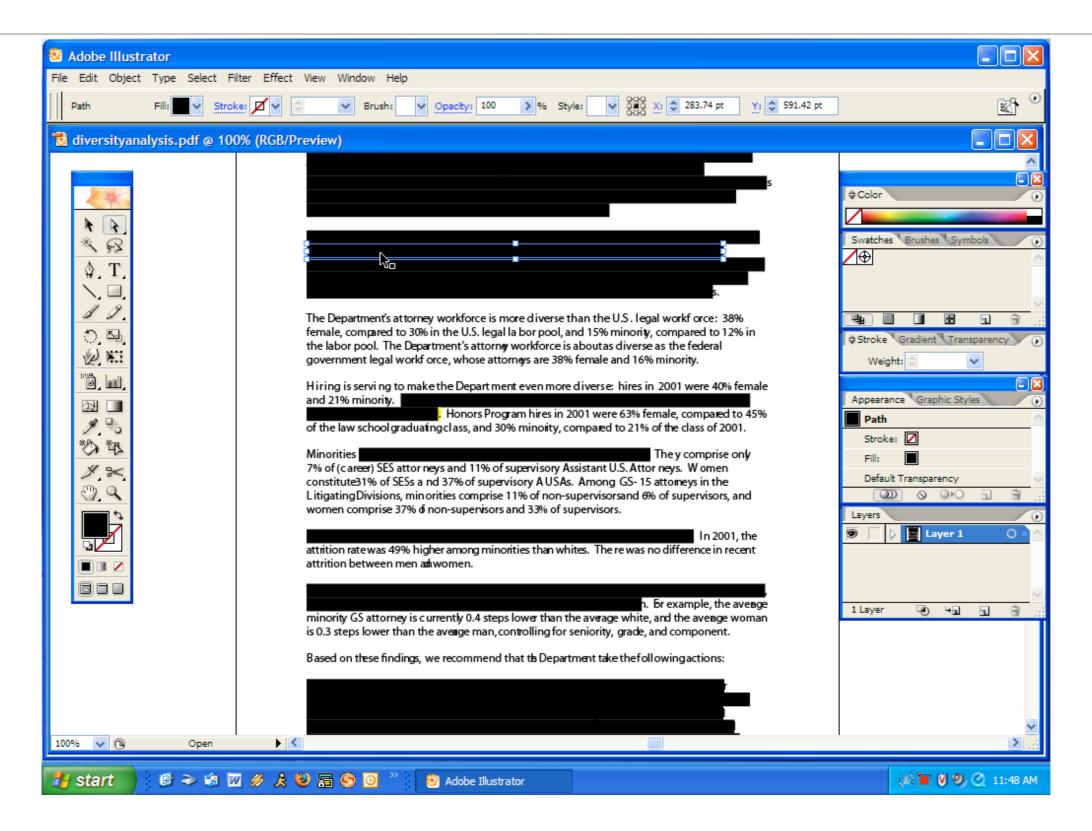
#### Adobe Acrobat Reader

Select and copy the text

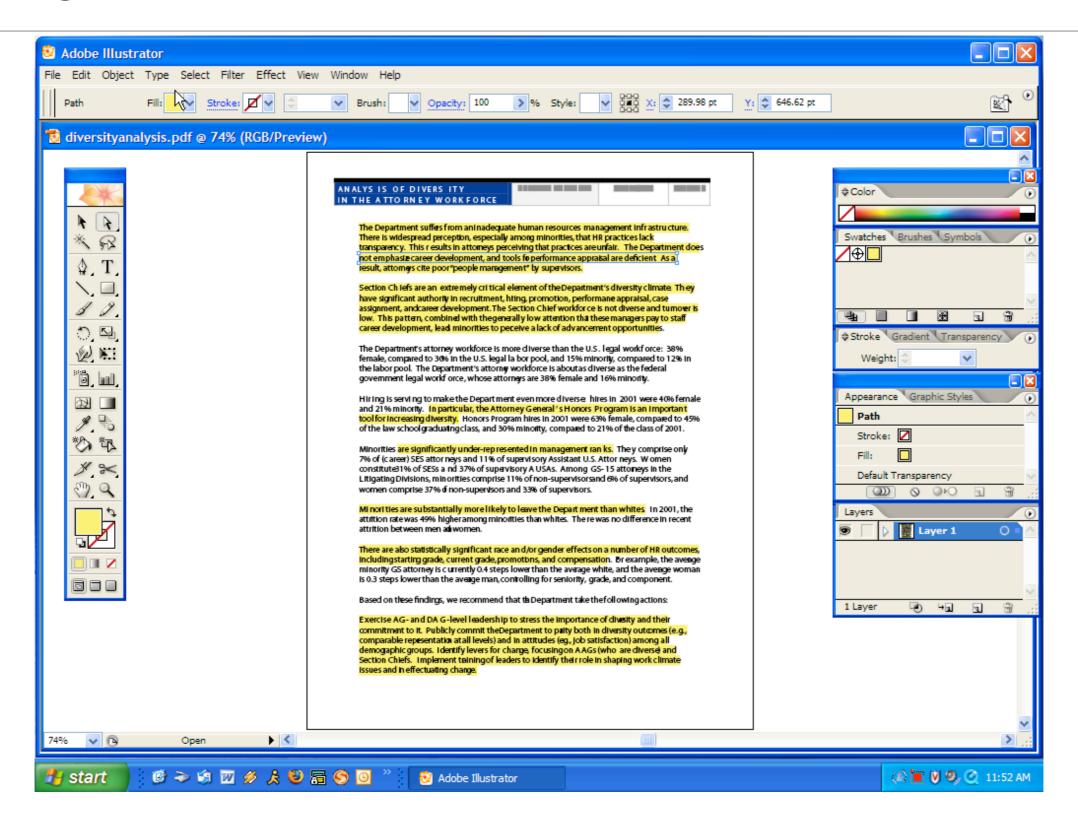
### Adobe's Illustrator can read and edit PDF files.



#### Select each "block box."



## Change the color of the box to yellow.



# Behold the "redacted" data.

#### ANALYS IS OF DIVERS ITY IN THE ATTORNEY WORKFORCE

KPMG CO NS ULT ING JUNE 14, 2002

PAG E ES- 2

The Department suffers from an inadequate human resources management infrastructure. There is widespread perception, especially among minorities, that HR practices lack transparency. This r esults in attorneys perceiving that practices are unfair. The Department does not emphasize career development, and tools for performance appraisal are deficient. As a result, attorneys cite poor "people management" by supervisors.

Section Chiefs are an extremely critical element of the Department's diversity climate. They have significant authority in recruitment, hing, promotion, performane appraisal, case assignment, and career development. The Section Chief workforce is not diverse and turnover is low. This pattern, combined with the generally low attention that these managers pay to staff career development, lead minorities to peceive a lack of advancement opportunities.

The Department's attorney workforce is more diverse than the U.S. legal workf orce: 38% female, compared to 30% in the U.S. legal la bor pool, and 15% minority, compared to 12% in the labor pool. The Department's attorney workforce is about as diverse as the federal government legal workf orce, whose attorneys are 38% female and 16% minority.

Hiring is serving to make the Department even more diverse: hires in 2001 were 40% female and 21% minority. In particular, the Attorney General's Honors Program is an important tool for increasing diversity. Honors Program hires in 2001 were 63% female, compared to 45% of the law school graduating class, and 30% minority, compared to 21% of the class of 2001.

Minorities are significantly under-represented in management ranks. They comprise only 7% of (career) SES attor neys and 11% of supervisory Assistant U.S. Attor neys. Women constitute 31% of SESs and 37% of supervisory AUSAs. Among GS-15 attorneys in the Litigating Divisions, minorities comprise 11% of non-supervisors and 6% of supervisors, and women comprise 37% of non-supervisors and 33% of supervisors.

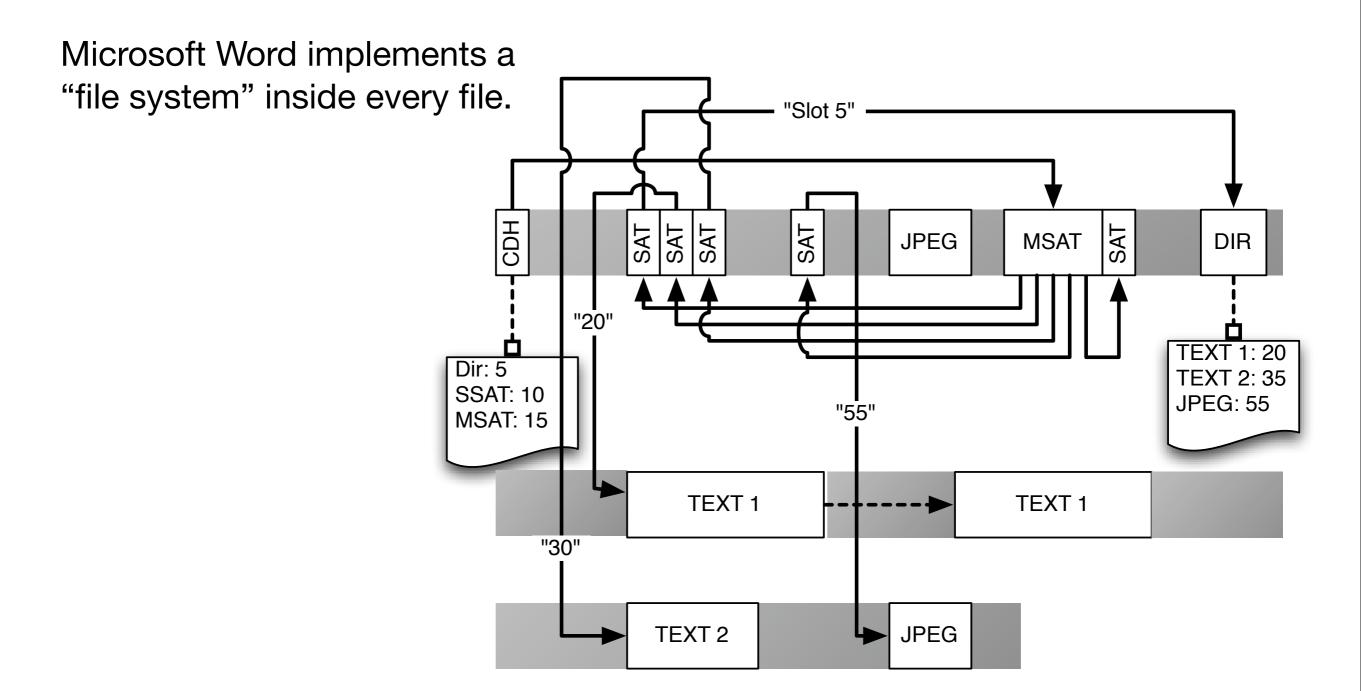
Mi nori ties are substantially more likely to leave the Depart ment than whites. In 2001, the attrition rate was 49% higher among minorities than whites. The re was no difference in recent attrition between men and women.

There are also statistically significant race an d/or gender effects on a number of HR outcomes, including starting grade, current grade, promotions, and compensation. Be example, the aveage minority GS attorney is currently 0.4 steps lower than the average white, and the aveage woman is 0.3 steps lower than the aveage man, controlling for seniority, grade, and component.

Based on these findings, we recommend that the Department take the following actions:

Exercise AG- and DAG-level leadership to stress the importance of divesity and their commitment to it. Publicly commit the Department to paity both in diversity outcomes (e.g., comparable representation at all levels) and in attitudes (eg., job satisfaction) among all demographic groups. Identify levers for change, focusing on AAGs (who are diverse) and Section Chiefs. Implement training of leaders to identify their role in shaping work climate issues and in effectuating change.

Data can be left in a Word document in unallocated sectors.



## Tools for recovering hidden Word data:

#### Unix strings(1) command reveals:

- Deleted text
- Names and/or usernames of author and editors
- Paths where document was saved
- GUID of system on which it was saved
   Note: Text may be UTF16 (remove NULLs or use more intelligent processing)

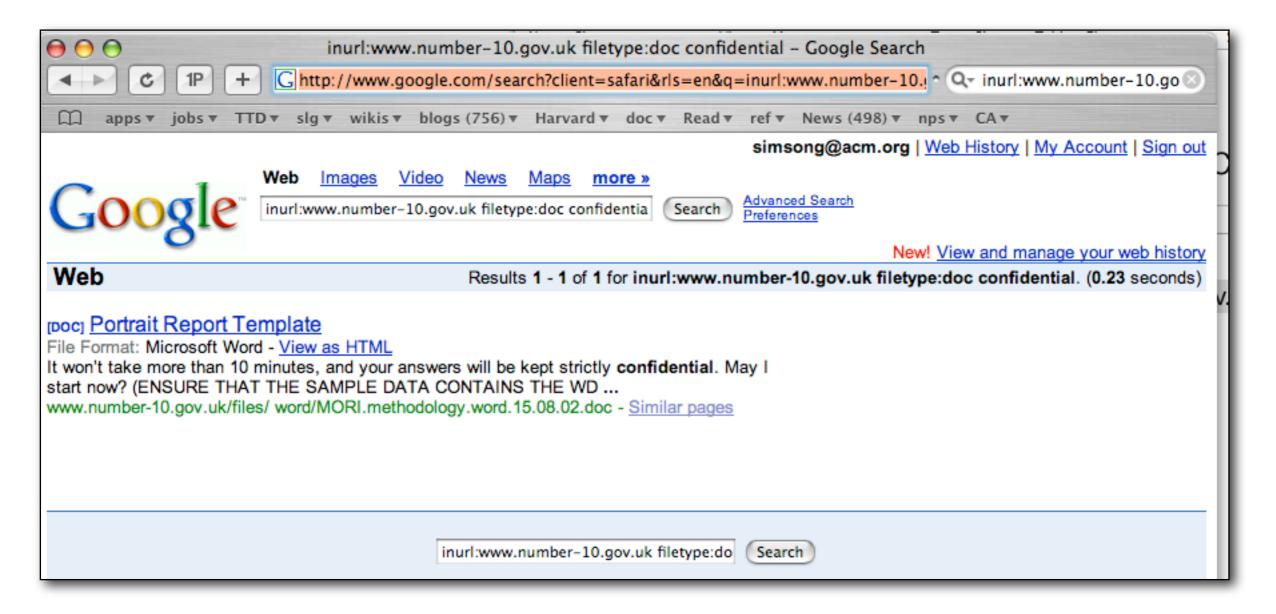
#### Other tools:

- Antiword (<a href="http://www.winfield.demon.nl">http://www.winfield.demon.nl</a>/)
- catdoc
- wvText
- MITRE's Heuristic Office File Format Analysis toolkit (HOFFA)

## Tools for finding Microsoft Word files

#### Use Google!

• inurl:<u>www.number-10.gov.uk</u> filetype:doc confidential



# Case study of inconsistent data: State of Utah vs. Carl Payne

```
Ip:NP:6445:::::
smtp:*NP:6445:::::
uucp:NP:6445:::::
nuucp:NP:6445:::::
listen:*LK*:::::
nobody:NP:6445:::::
setup:ANImj3G8/T3m2:6445:::::
ftp:NP:6445:::::
carl:*1rwuFse0eS/S6:9807:::::
majo:NP::::::
```

## State of Utah vs. Carl Payne

#### State's Claims:

- Victim ISP suffered devastating attack on November 6th, 1996.
  - All files erased
  - All router configurations cleared.
- Carl Payne, one of the company's founders, had a falling out with the company and was terminated on October 30th, 1996.
- Payne had the necessary knowledge to carry out the attack.
- Payne created a "back door" on his last week of employment.
- Payne's accounts were used for the attack.

## State of Utah vs. Carl Payne

#### State's Evidence:

- 140 pages of printouts made by a local expert on the day of the attack.
- Testimony of the expert.
- Testimony of the Fibernet employees

#### Payne's Defense:

- "I didn't do it."
- All of Payne's account passwords had been changed when he was terminated.
- Alibi defense: was having breakfast with a friend when attack took place.

## /etc/shadow (printed November 6, 1996)

```
root:0rdtD.YmG4mNA:9818::::::
daemon:NP:6445:::::
bin:NP:6445:::::
sys:NP:6445:::::
adm:NP:6445:::::
lp:NP:6445:::::
smtp:*NP:6445:::::
uucp:NP:6445:::::
nuucp:NP:6445:::::
listen:*LK*:::::
nobody:NP:6445:::::
noaccess:NP:6445:::::
setup:ANImj3G8/T3m2:6445:::::
ftp:NP:6445:::::
carl: *1rwuFse0eS/S6:9807:::::
majo:NP::::::
news::::::
dbowling:*n.56DqWPfcZ6w:9807:::::
hart:YqEuyT.mD8buc::::::
usenet:*Lq.mMF7KaEdd.:9800:::::
```

# Solaris /etc/shadow: setup:ANImj3G8/T3m2:64455::::::

Field 1: Username

Field 2: Encrypted Password

Field 3: Password Aging

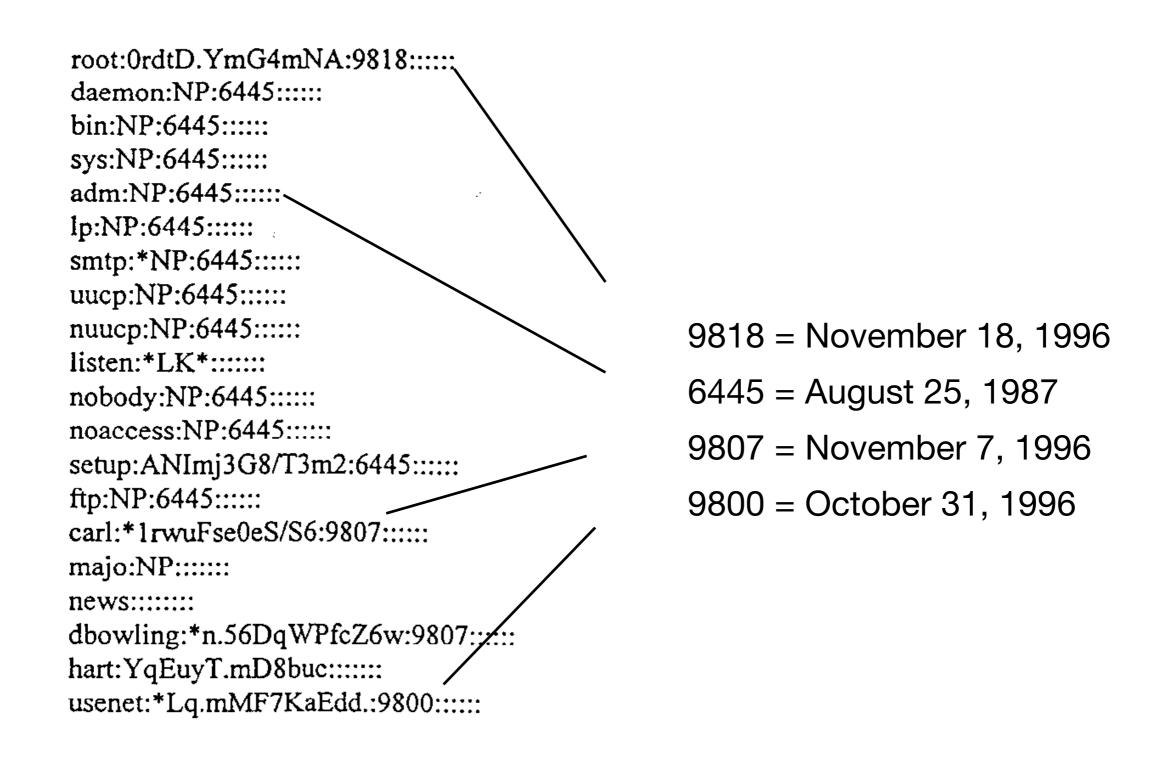
Number of days since January 1, 1970

Source: Solaris Documentation

## Decoding "6645"

August 25, 1987

## /etc/shadow (Printed November 6, 1996 by prosecution expert witness)



## Lessons of Utah vs. Payne

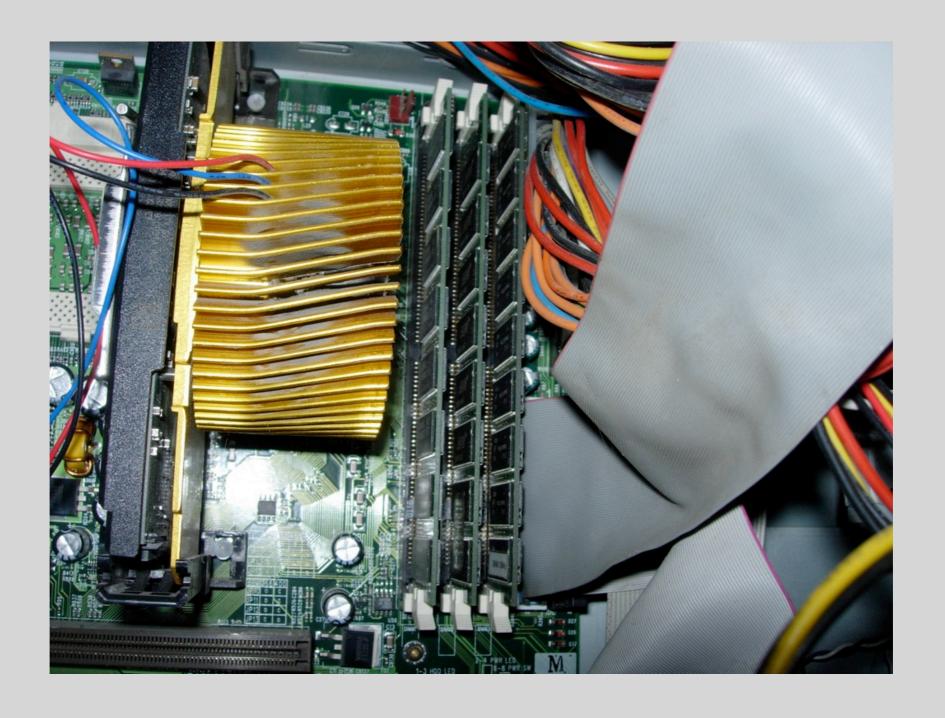
Not all "Evidence" is equal (Chain-of-custody is vital)

Evidence may not prove what you think it proves

Computer evidence lends itself to forgery

Most data isn't tampered...

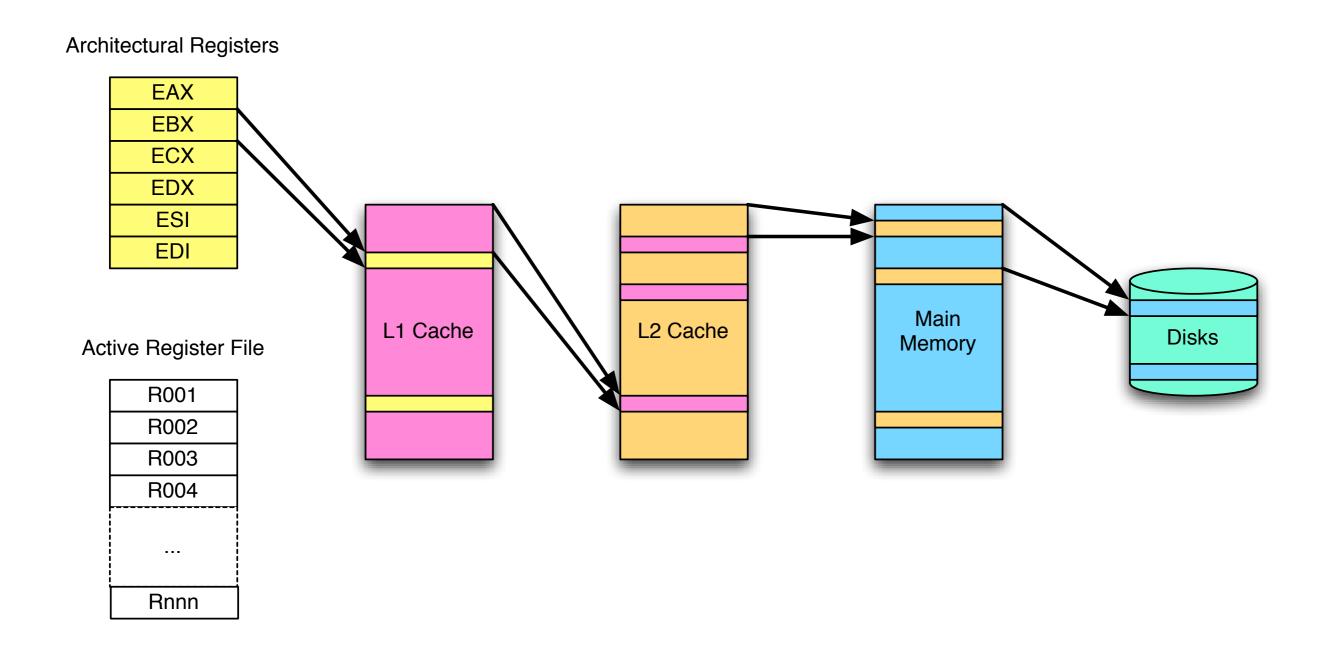
- ... but most data isn't used for evidence.
- If data is going to be used for evidence, there is an incentive to tamper with it.



## Memory Forensics

What was *really* happening on the subject's computer?

## Computer systems arrange memory in a hierarchy.



## There are many ways to get access to the memory.

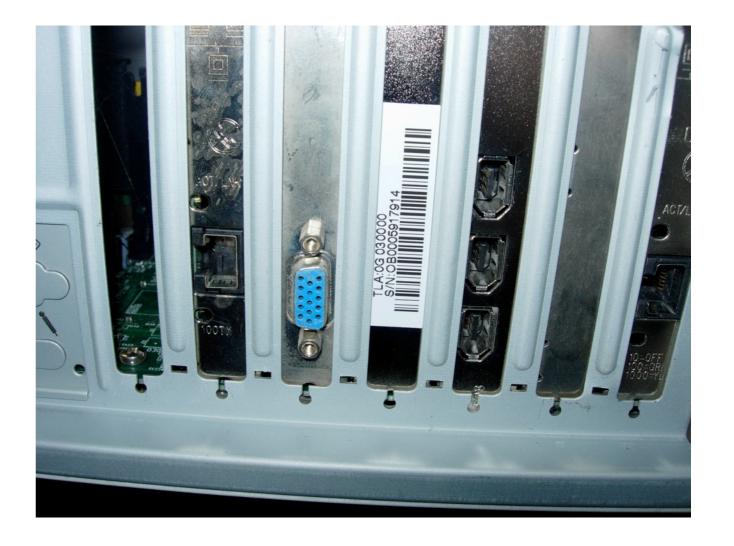
Unix/Linux: /dev/mem

Windows: Device Drivers

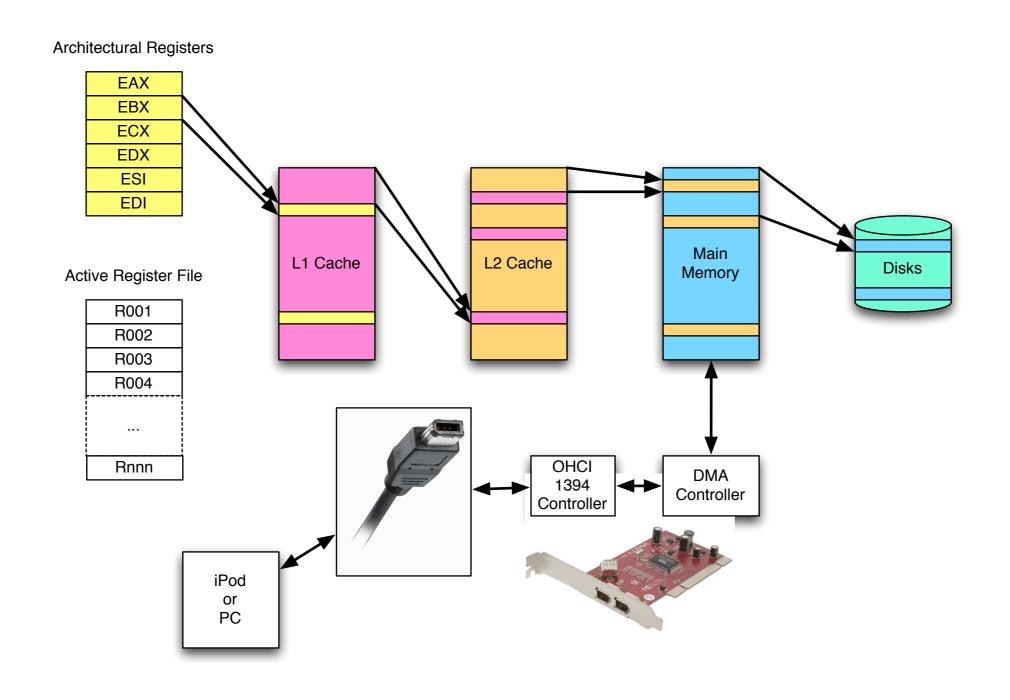
Hardware memory imagers

#### *Firewire*

- Firewire designed as a replacement for hard drives.
- ATA drives support DMA
- So Firewire supports DMA



## It's pretty easy to attack a system with an iPod



# Many different kinds of information can be retrieved from a computer's memory.

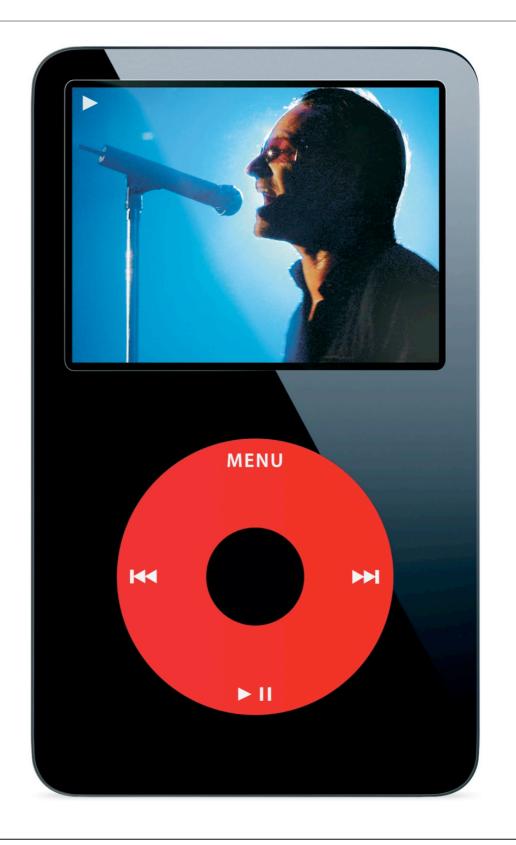
#### Reading:

- Current contents of the screen
- Cryptographic Keys
- Passwords (BIOS & programs)
- Programs
- All data

#### Writing:

- Patch programs on the fly
- Change security levels

DMA bypasses the operating system and the CPU.





## Cell Phone Forensics

Who did you call? Where have you been?

#### Cell Phone Forensics: What can be done

PHONE: Recovery of personal information (even after deletion)

- SMS messages
- Phone log
- Phone book

PHONE: Recovery of service information

• Cell sites passed, used

CELL SITES: Recovery of phone information

Phones in the area

## Cell phone forensics: Precautions

Turning on the phone can damage data!

• But sometimes you can't access the data any other way

## Paraben's tools for cell phone forensics

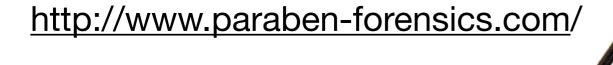
Paraben "Device Seizure" to image the phone's content.

- Acquires phone flash and some of GSM SIM card
- Understands some of the phone's internal databases
- Views some of the photos, messages, etc.
- Only covers specific phones

"Device Seizure Toolbox" has lots of different cables.

"StrongHold Box" prevents phone from calling home.

Project-a-Phone captures screens





## Cell Phone Forensics: References & Resources

Guidelines on Cell Phone Forensics (NIST SP 800-101)

- August 2006
- http://csrc.nist.gov/publications/drafts/Draft-SP800-101.pdf

Cell Phone Forensic Tools: An Overview and Analysis (NISTIR 7250)

http://csrc.nist.gov/publications/nistir/nistir-7250.pdf

PDA Forensic Tools: An Overview and Analysis (NISTIR 7100)

• http://csrc.nist.gov/publications/nistir/nistir-7100-PDAForensics.pdf

```
if(dirlist.size()==0){
    if(argc!=2){
        fprintf(stderr,"Please specify a directory or just two AFF files.\n\n");
        usage();
    }
    /* Must be copying from file1 to file2. Make sure file2 does not exist */
    if(access(argv[1],R_OK)==0){
        errx(1,"File exists: %s\n",argv[1]);
    }
    vector<string> outfiles;
    outfiles.push_back(argv[1]);
    return afcopy(argv[0],outfiles);
}
```

## Software Forensics

Who authored a program? Are two programs similar? How old is a program?



## Anti-Forensics: Techniques, Detection and Countermeasures

## 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:

- Avoiding detection
- Disrupting information collection
- Increasing the examiner's time
- Casting doubt on a forensic report or testimony (Liu and Brown, 2006)
- Forcing a tool to reveal its presence
- Subverting the tool using it to attack the examiner or organization
- Leaving no evidence that the AF tool has been run

# Physical destruction makes forensic recovery impossible.





# One traditional Anti-Forensic technique is to overwrite or otherwise destroy data.

Overwriting: Eliminate data or metadata (e.g. disk sanitizers, Microsoft Word metadata "washers," timestamp eliminators.)

Disk Sanitizers; Free Space Sanitizers; File Shredders

Microsoft Remove Hidden Data Tool; cipher.exe; ccleaner

Metadata Erasers

Example: timestomp

Hard problem: What should be overwritten?

# Anti-Forensic tools can hide data with cryptography or steganography.

Cryptographic File Systems (EFS, TrueCrypt)

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

Program Packers (PECompact, Burneye) & Rootkits

Steganography

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
- Host Protected Areas & Device Configuration Overlay

<sup>\*</sup>Onion routing also protects from traffic analysis

## 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
- Anonymous Identities and Storage

(don't worry; we have slides for each of these)

# 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

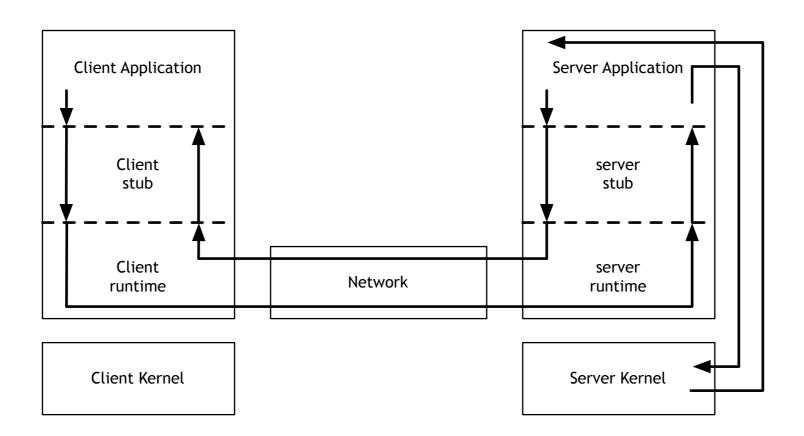
#### **Userland Execve**

- 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."

- 1. Announce BGP route
- 2. Conduct attack
- 3. Withdraw BGP address

Being used by spammers today

(<a href="http://www.nanog.org/mtg-0602/pdf/feamster.pdf">http://www.nanog.org/mtg-0602/pdf/feamster.pdf</a>)

# 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:

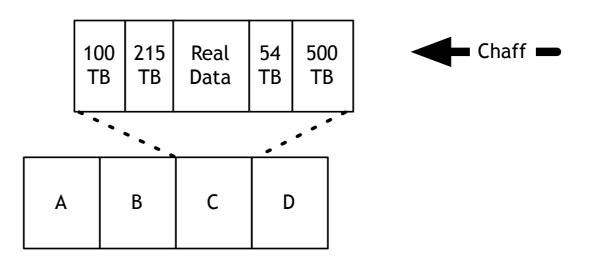
- → Ability to run code on the forensic appliance
- ➡ Erase collected evidence
- ➡ Break the investigative software
- → Leak information about the analyst or the investigation
- → Implicate the investigator

# 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.:



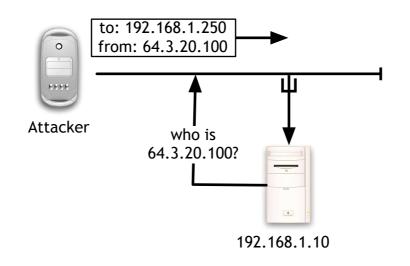
# Anti-Forensic Tools can detect Computer Forensic Tools: cat-and-mouse.

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

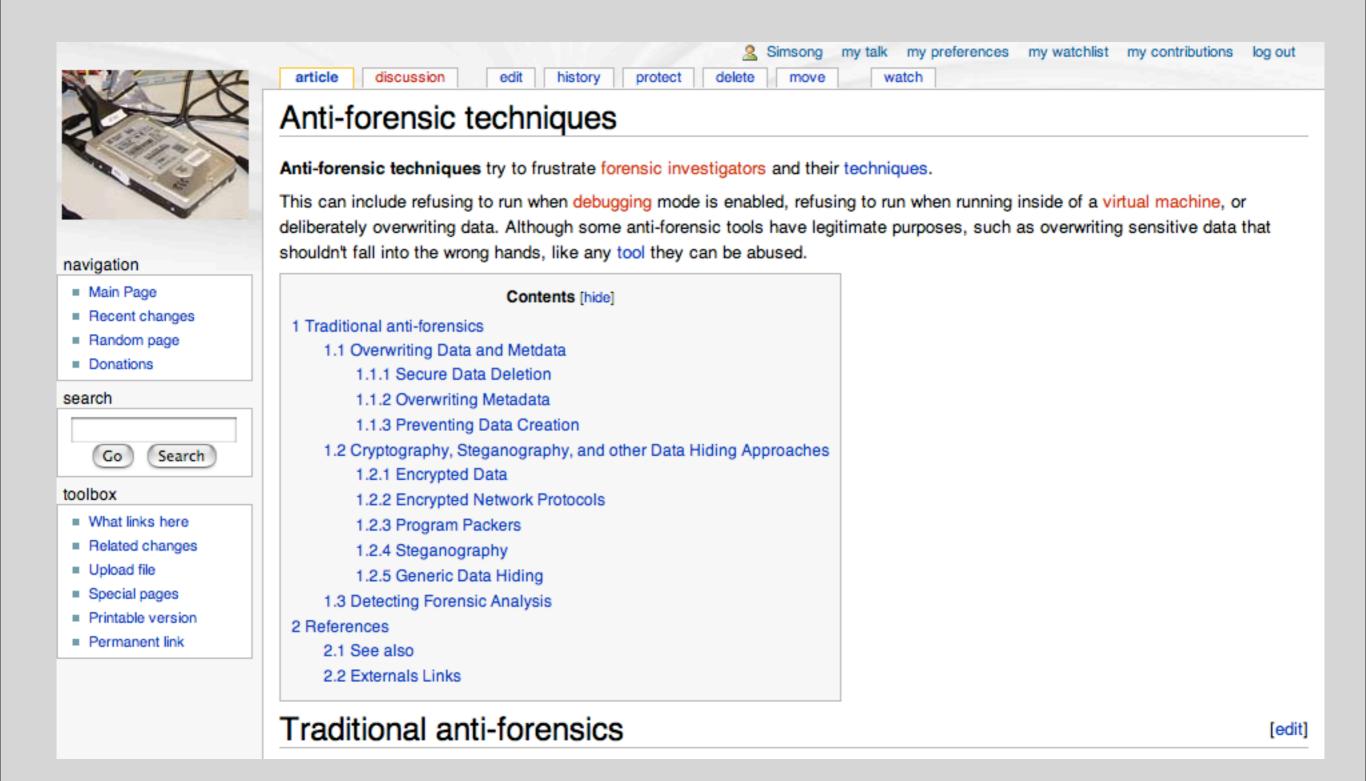
Improve the tools — many CFTs are poorly written.

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



Find out more at the Forensics Wiki:

http://www.forensicswiki.org/

### In Conclusion:

Many forensic techniques in use today can be circumvented

Circumvention tools are widely available

### Common approaches:

- Overwriting data
- Encrypting data
- Anonymous identities & resources
- Exploit bugs in computer forensic tools to hide

#### New approaches:

- Minimizing or eliminating memory footprints
- Virtual machines
- Direct attacks against computer forensic tools

http://www.simson.net/ref/2007/slides-ICIW.pdf

## Research directions in Computer Forensics

### **Environmental Data Survey Projects**

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

#### Theory and Algorithm Development:

- Brian Carrier 2006 PhD
- Cross-Drive Analysis
- Carving Fragmented Objects with Validation

### **Tool Development**

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

## Forensics, Conclusion

Forensic analysis is a growth area.

Being a practitioner is hard:

- Many skills
- Many tools
- In-depth knowledge of many different systems

What is the forensics research agenda?

## Other Resources

http://www.forensicswiki.org/

http://www/forensicwiki.com/

http://staff.washington.edu/dittrich/forensics.html

http://faculty.ncwc.edu/toconnor/426/426links.htm