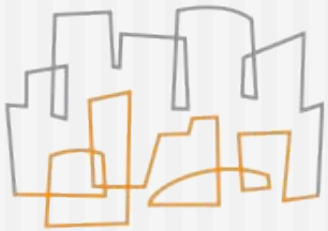
Remembrance of Data Passed: Used Disk Drives and Computer Forensics

Simson L. Garfinkel

Computer Science and Artificial Intelligence Laboratory



Acknowledgements

- Abhi Shelat (MIT)
- Ben Gleb (MIT)

© 2004 Simson L. Garfinkel

August 1998

- I purchased 10 used computers from a computer store…
- Mostly ‘386 and ‘486 machines…
- … for a project

© 2004 Simson L. Garfinkel

Computer #1

- Operational hard drive … It boot!
- File server from a law firm…
- Still had client documents…

© 2004 Simson L. Garfinkel

Computers #2-#5

- Server from a law firm
- Database of mental health patients
- Quicken files
- Draft manuscript of a novelist…

© 2004 Simson L. Garfinkel

Other Stories of Data Passed…

April 1997

- A woman in Pahrump, NV, purchases a used IBM PC and discovers records from 2000 patients who had prescriptions filled at Smitty's Supermarkets pharmacy in Tempe, AZ.

August 2001

- More than 100 computers from Viant with confidential client data sold at auction by Dovebid.

Spring 2002

- Pennsylvania state Department of Labor and Industry sells computers with "thousands of files of information about state employees."

August 2002

- Purdue student purchased used Macintosh computer at equipment exchange; computer contains FileMaker database with names and demographic information of 100 applicants to Entomology Department.

© 2004 Simson L. Garfinkel

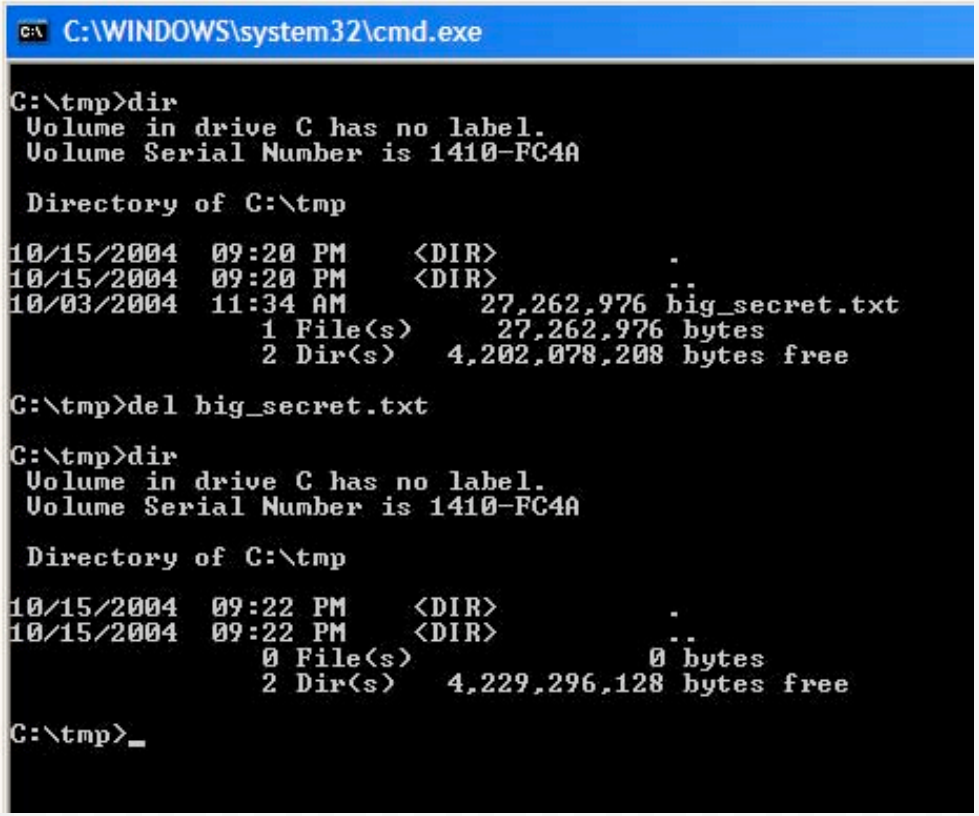
With so many used systems, why so few stories of actual data disclosure

- Hypothesis #1: Disclosure of “data passed” is exceedingly rare because most systems are properly sanitized.
- Hypothesis #2: Disclosures are so common that they are not newsworthy.
- Hypothesis #3: Systems aren’t properly sanitized, but few notice the data.

© 2004 Simson L. Garfinkel

How could people not notice the data?

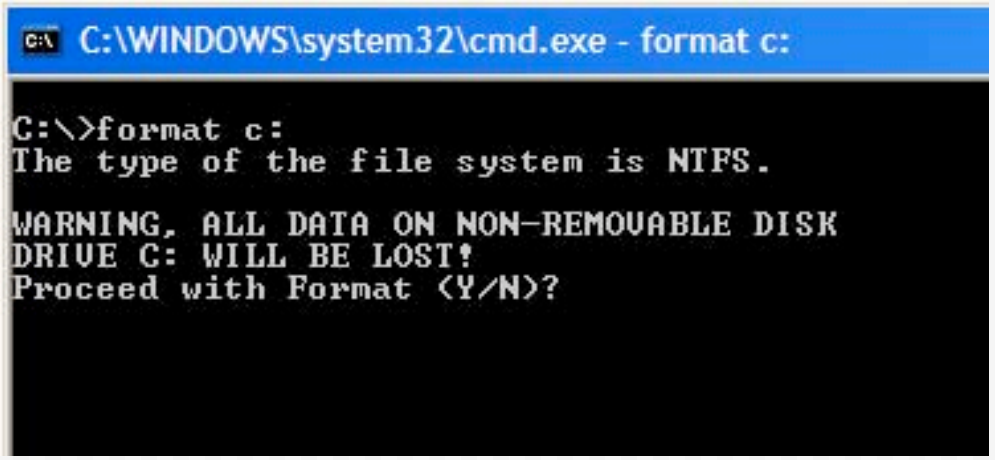
- DEL removes the file's name…
- … but doesn't delete the file's data



© 2004 Simson L. Garfinkel

How could people not notice the data?

- FORMAT C: writes a new root directory…



© 2004 Simson L. Garfinkel

FORMAT is misleading

```
A:\>format c:

WARNING, ALL DATA ON NON-REMOVABLE DISK
DRIVE C: WILL BE LOST!
proceed with Format (Y/N)?y

Formatting 1,007.96M
100 percent completed.
Writing out file allocation table
Complete.
```

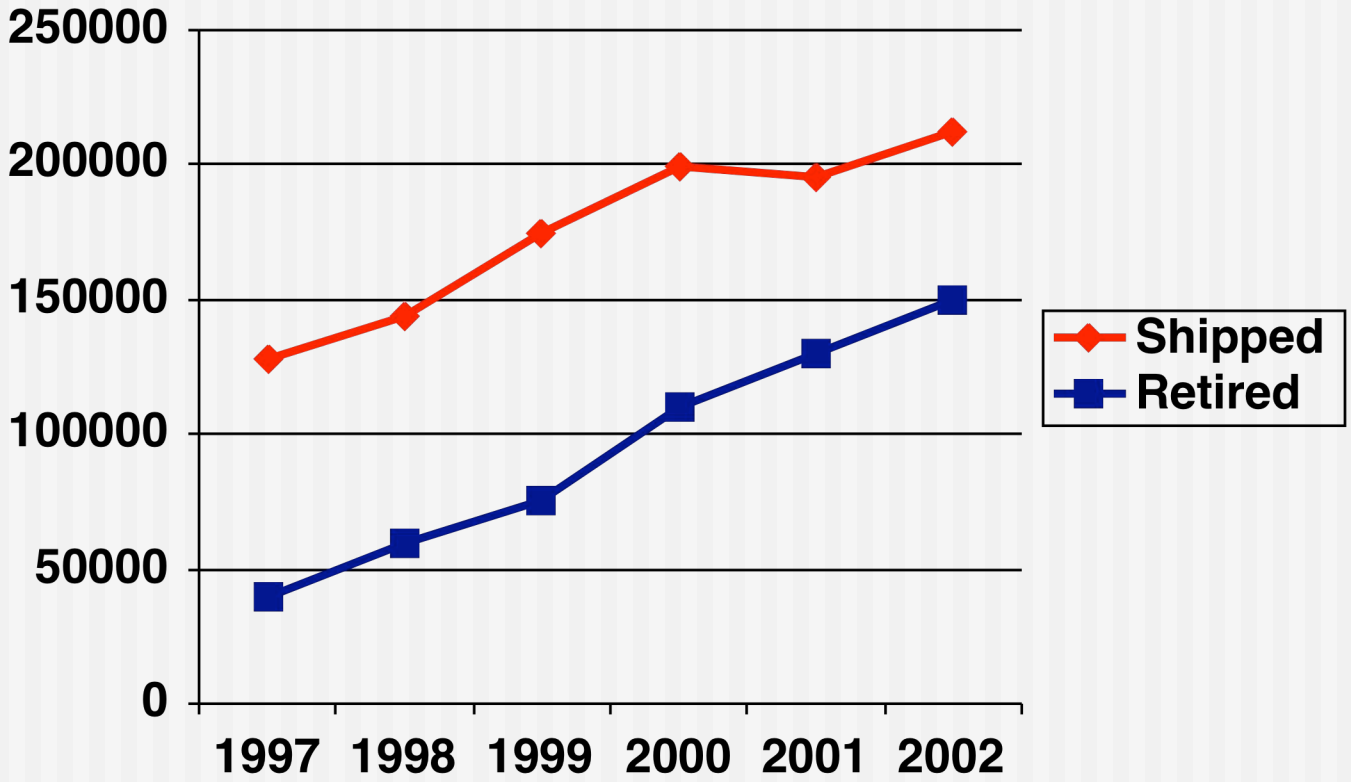

© 2004 Simson L. Garfinkel

Hard Drives Pose Special Problem For Computer Security

- Do not forget data when power is removed.
- Can contain data that is not immediately visible.
- Today's computers can read hard drives that are 15 years old!
  - Electrically compatible (IDE/ATA)
  - Logically compatible (FAT16/32 file systems)
  - Very different from tape systems
- Strong social bias against destroying a working drive

© 2004 Simson L. Garfinkel

149M Drives Retired in 2002!



© 2004 Simson L. Garfinkel

“Retire?”

Deckard (Harrison Ford) retires a replicant.

Blade Runner (1982)

© 2004 Simson L. Garfinkel

Throwing out a Hard Drive Feels Wrong

- Give to:
  - School
  - Church
  - Parents
- Send it to India
- Find somebody to "take it away."

© 2004 Simson L. Garfinkel

Many hard drives are "repurposed," not "retired"

- Re-used within an organization
- Given to charities
- Sold on eBay

All Categories Save this search

350 items found for **hard drives**

Sort by items: **ending first** | newly listed | lowest priced | highest priced

Picture hide	Item Title	Price	Bids	Time Left
	Lot of hard and floppy drives	$5.50	2	14n
	Lot of hard and floppy drives	$5.50	2	22n
	Lot of hard and floppy drives	$5.50	2	25n
	Lot of 2 hard drives IDE	$8.00	12	29n
	3.2 gig Hard Drives	$180.00	-	59n
	(5) 1.2 hard drives & (15) 10/100 network	$25.00	1	1h 00n
	Lot of 3 Quantum 9.1 gig SCSI Hard Drives	$26.00	6	1h 25n
	IDE HARD DRIVES (3)	$6.50	6	1h 46n
	LOT OF 5 Hard Drives! 3.2 Gig Western Digital	$120.00 $124.95 Buy It Now	-	1h 50n
	QTY 3....IDE Hard Drives 2.5 Gig	$20.50	5	2h 02n
	5 WESTERN DIGITAL 2.5 GIG HARD DRIVES	$30.00	4	2h 03n
	QTY 3....IDE Hard Drives 1.0 Gig	$9.99	1	2h 04n
	Western Digital 850 meg IDE Hard Drives dutch	$6.00	1	2h 57n
	WINDOWS	$6.00	-	3h 18n

© 2004 Simson L. Garfinkel

Modern systems use several techniques for assuring data privacy:

#1 - Physical security

#2 - Logical access controls (operating system)

#3 - Cryptography (disk & link)

© 2004 Simson L. Garfinkel

Data privacy techniques don't apply to repurposed disks

Techniques for assuring confidentiality:

- ~~#1 - Physical security~~
- ~~#2 - Logical access controls (operating system)~~
- #3 - Cryptography (disk & link)

… and most data isn't encrypted

© 2004 Simson L. Garfinkel

Weird Stuff, Sunnyvale California,January 1999

- 10 GB drive: $19 “tested”
- 500 MB drive: $3 “as is”

Q: “How do you sanitize them?”

A: “We FDISK them!”

© 2004 Simson L. Garfinkel

FDISK does not sanitize disks

- 10 GB drive: 20,044,160 sectors
- “FDISK”
  - Writes 2,563 sectors (0.01%)
- “FORMAT”
  - Writes 21,541 sectors (0.11%)
  - Erases the FAT
  - (complicates recovery of fragmented files.)

© 2004 Simson L. Garfinkel

The “Remembrance of Data Passed” Study

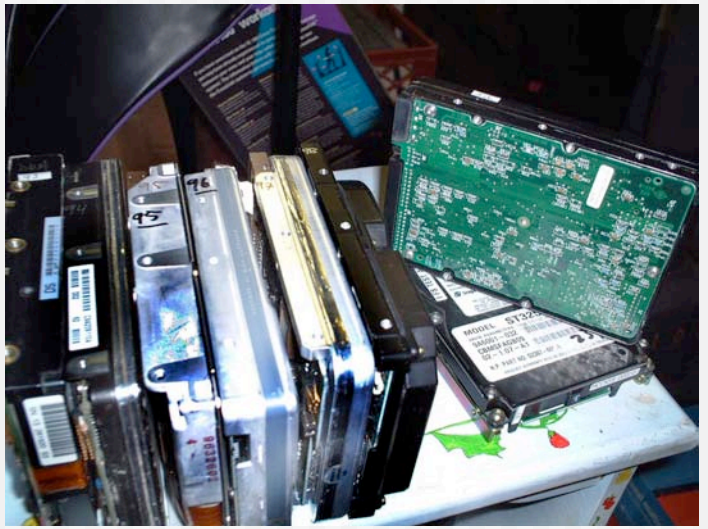
- I purchased 235 used hard drives between November 2000 and January 2003
  - eBay
  - Computer stores
  - Swap fests
  - No more than 20 from the same vendor
- Mounted the drives, copied off the data, looked at what I found.

© 2004 Simson L. Garfinkel

Drives arrived by UPS.

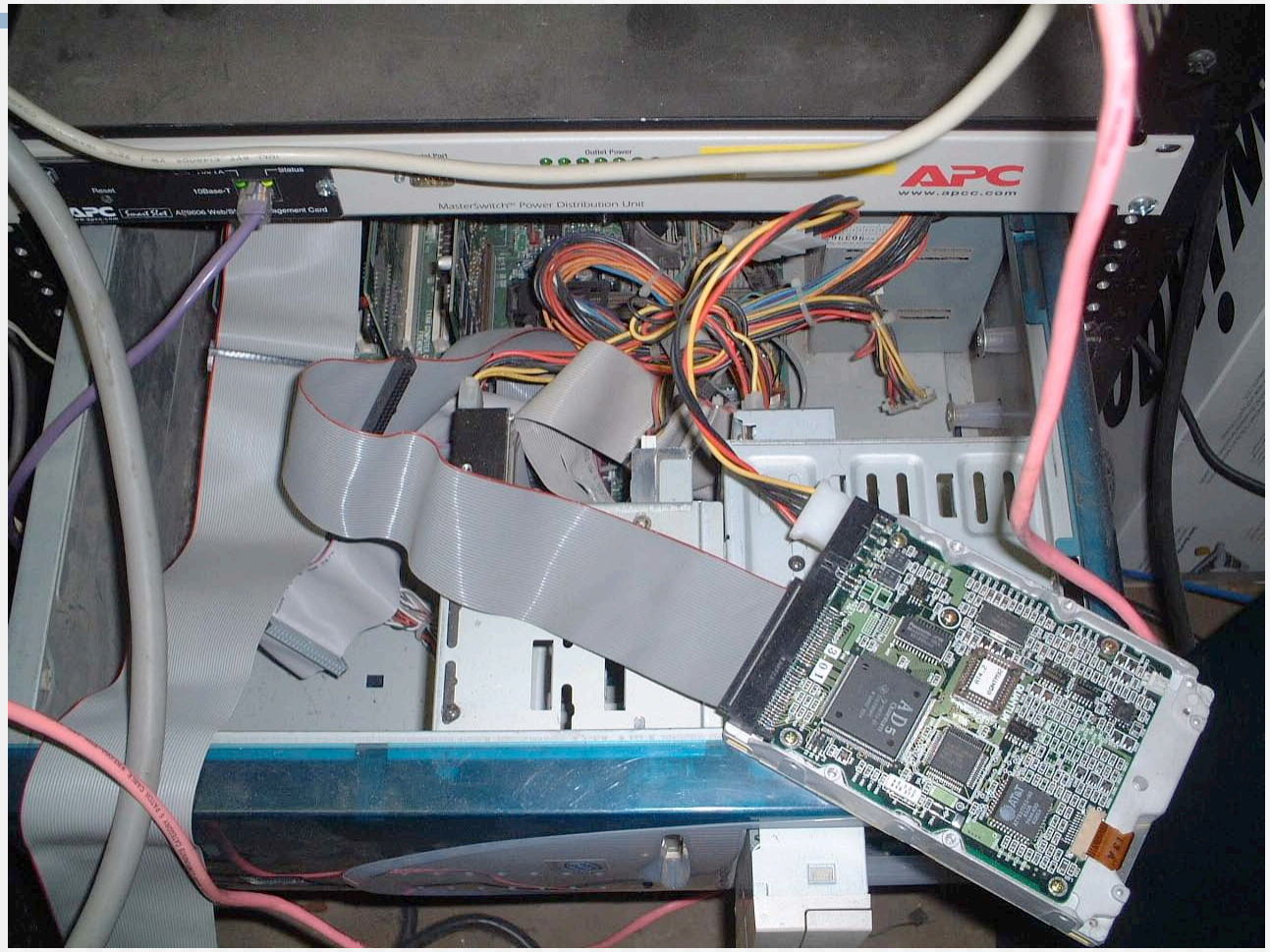
© 2004 Simson L. Garfinkel

Numbered and put on shelf



© 2004 Simson L. Garfinkel

Imaged using FreeBSD



© 2004 Simson L. Garfinkel

Stored images on RAID

© 2004 Simson L. Garfinkel

Lore

- Use atacontrol(8) to attach/deteach

```
dd if=/dev/ad1 of=/image conv=noerror,sync
```

- Use a modified “dd” that fills error blocks with a distinct pattern / code.

© 2004 Simson L. Garfinkel

Stored file metadata in MySQL

- Disk #
- Dir name
- File name
- Length
- mtime
- MD5 (Actually, md5id)

© 2004 Simson L. Garfinkel

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

- Purchased for $5 from a Mass retail store on eBay
- Copied the data off: 541MB
- Initial analysis:
  - 1,057,392 disk blocks
  - 67,878 blocks are all NULs (6%)

```
-r--r-----  1 root     project  541384704 Aug  9  2002 70.img
-rw-r-----  1 simsong  project     205892 Aug  9  2002 70.tar.gz
```


© 2004 Simson L. Garfinkel

#70 the disk partition report

```
******* Working on device /dev/ad2 *******
parameters extracted from in-core disklabel are:
cylinders=524 heads=32 sectors/track=63 (2016 blks/cyl)

parameters to be used for BIOS calculations are:
cylinders=524 heads=32 sectors/track=63 (2016 blks/cyl)

Media sector size is 512
Warning: BIOS sector numbering starts with sector 1
Information from DOS bootblock is:
The data for partition 1 is:
sysid 11,(DOS or Windows 95 with 32 bit FAT)
    start 63, size 1054305 (514 Meg), flag 80 (active)
        beg: cyl 0/ head 1/ sector 1;
        end: cyl 522/ head 31/ sector 63
The data for partition 2 is:
<UNUSED>
The data for partition 3 is:
<UNUSED>
The data for partition 4 is:
<UNUSED>
```


© 2004 Simson L. Garfinkel

70.tar.gz: Visible Files

```
% tar tfz images/tar.gz/70.tar.gz
./
IO.SYS
MSDOS.SYS
COMMAND.COM
%
```


© 2004 Simson L. Garfinkel

% strings 70.img | more

```
% strings img.70 | more
…
[.??
!ZY[
0123456789ABCDEFS
W0W0W090
W0W0
6,.h
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 long filenames.
To override this protection, see the LOCK /? command for more information.
The system has been halted.  Press Ctrl+Alt+Del to restart your computer.
You started your computer with a version of MS-DOS incompatible with this
version of Windows. Insert a Startup diskette matching this version of

OEMString = "NCR 14 inch Analog Color Display Enchanced SVGA, NCR Corporation"
      Graphics Mode: 640 x 480 at 72Hz vertical refresh.
      XResolution          = 640
      YResolution          = 480
      VerticalRefresh      = 72
…
```

56M of printable strings!

© 2004 Simson L. Garfinkel

70.img con’t

Wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwq
ling the Trial Edition

IBM AntiVirus Trial Edition is a full-function but time-limited
evaluation version of the IBM AntiVirus Desktop Edition product. You
may have received the Trial Edition on a promotional CD-ROM or as a
single-file installation program over a network. The Trial Edition
is available in seven national languages, and each language is
provided on a separate CC-ROM or as a separa
EAS.STCm
EET.STC
ELR.STCq
ELS.STC

© 2004 Simson L. Garfinkel

70.img ..

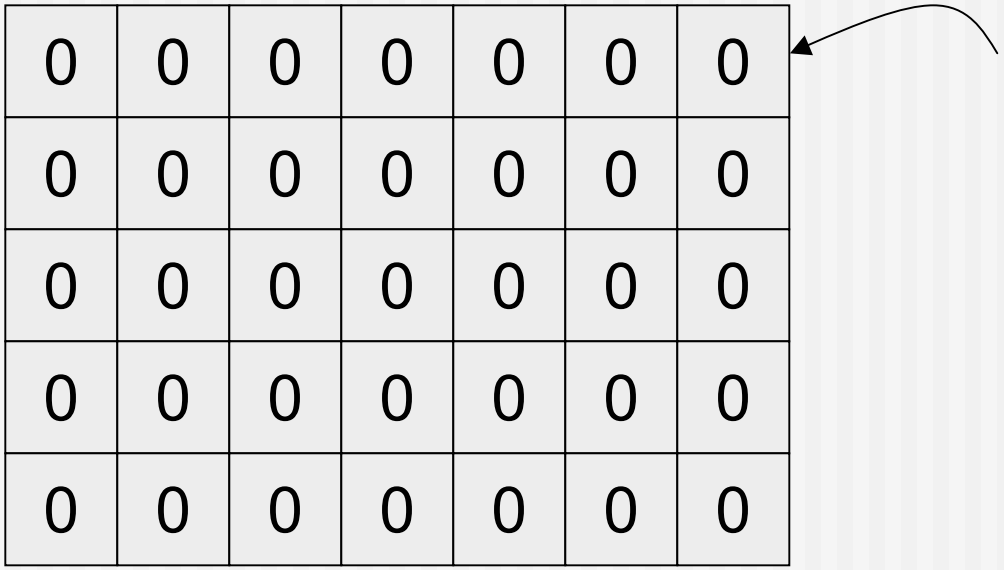
- Appears to have some kind of medical information on it.

MAB-DEDUCTIBLE
MAB-MOOP
MAB-MOOP-DED
METHIMAZOLE
INSULIN (HUMAN)
COUMARIN ANTICOAGULANTS
CARBAMATE DERIVATIVES
AMANTADINE
MANNITOL
MAPROTILINE
CARBAMAZEPINE
CHLORPHENESIN CARBAMATE
ETHINAMATE
FORMALDEHYDE
MAFENIDE ACETATE
s@ MALATHION
MAZINDOL
NOMIFENSINE MALEATE
PIPOBROMAN

© 2004 Simson L. Garfinkel

A typical hard disk

Factory-Fresh Hard disk: All Blank



Each block is 512 bytes

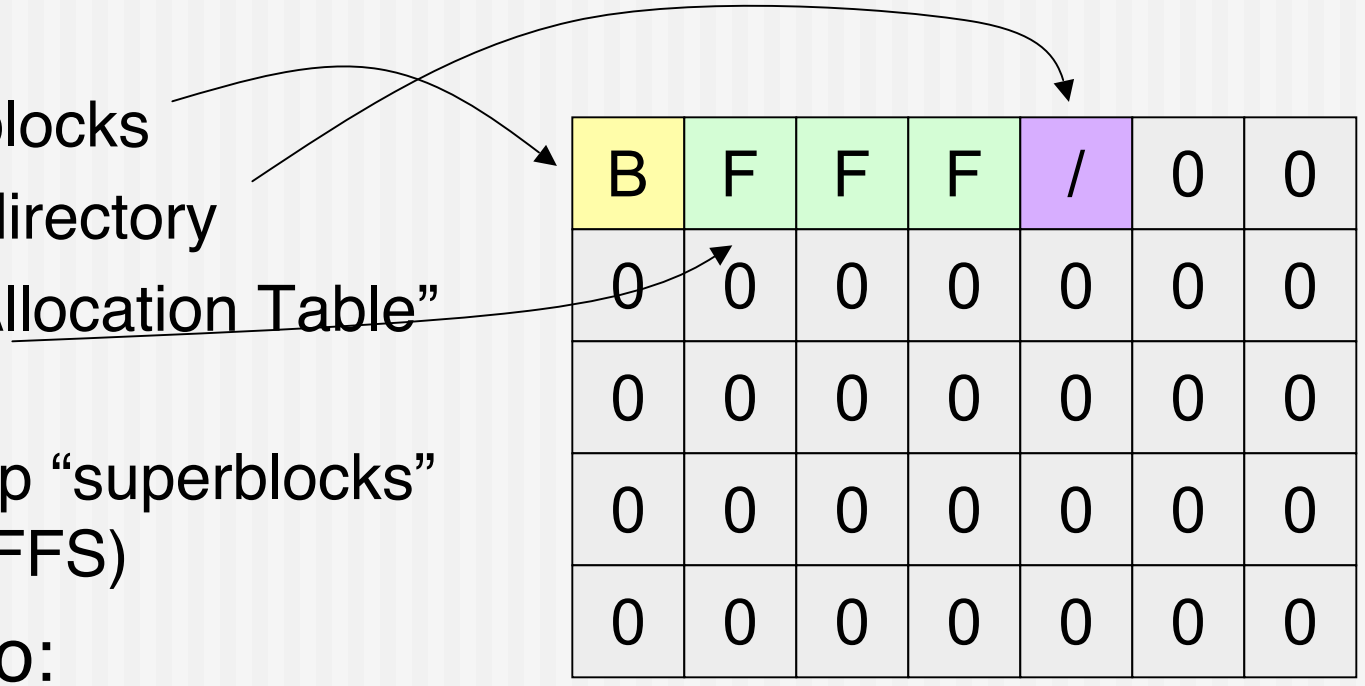
A 20G disk has 40M blocks.

Disk blocks (not to scale)

© 2004 Simson L. Garfinkel

% format C:*

- Writes:
  - Boot blocks
  - Root directory
  - “File Allocation Table” (FAT)
  - Backup “superblocks” (UFS/FFS)
- May also:
  - Validate surface

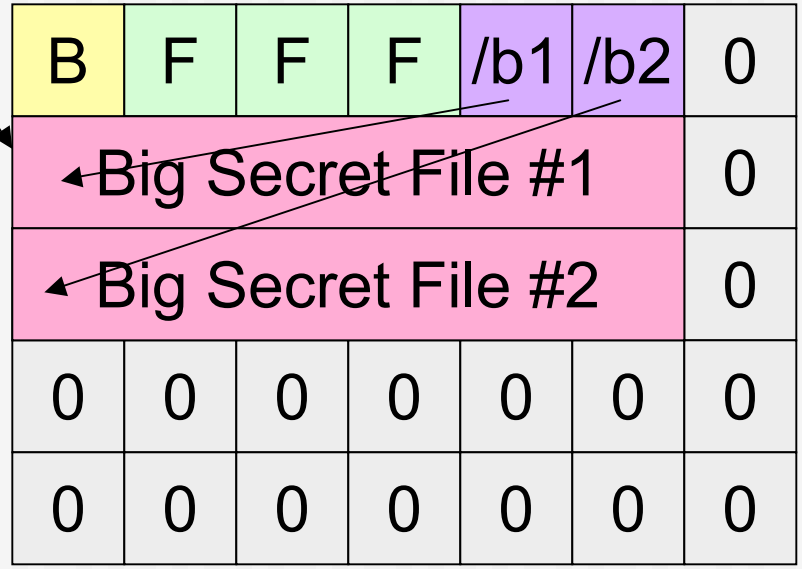


* Examples based on FAT32 running under Unix

© 2004 Simson L. Garfinkel

% cp b1 /mnt/b1
% cp b2 /mnt/b2

- Writes:
  - File Contents
  - File Directory Entry
  - Bookkeeping



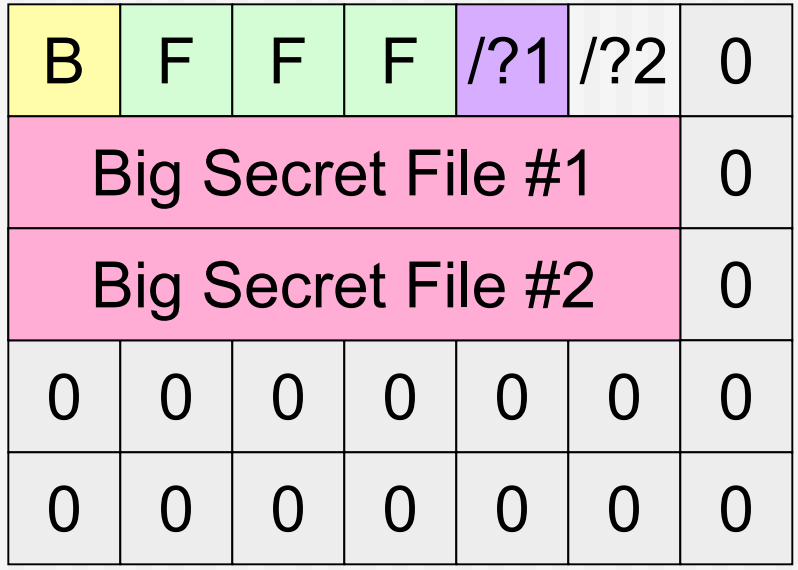
- root directory:

```
b1_____.___   jan 1 2004   block  7
b2_____.___   jan 1 2004   block 14
```


© 2004 Simson L. Garfinkel

% rm /mnt/b1
% rm /mnt/b2

- Writes:
  - New root directory
  - Bookkeeping



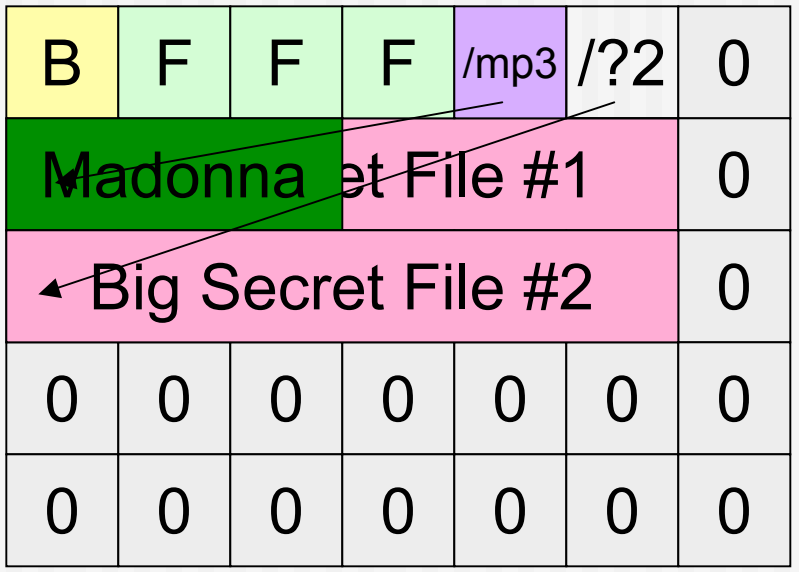
- new root directory:

```
?1_____.___   jan 1 2004   block  7
?2_____.___   jan 1 2004   block 14
```


© 2004 Simson L. Garfinkel

% cp Madonna.mp3 /mnt/mp3

- Writes:
  - New root directory
  - madonna.mp3
  - Bookkeeping



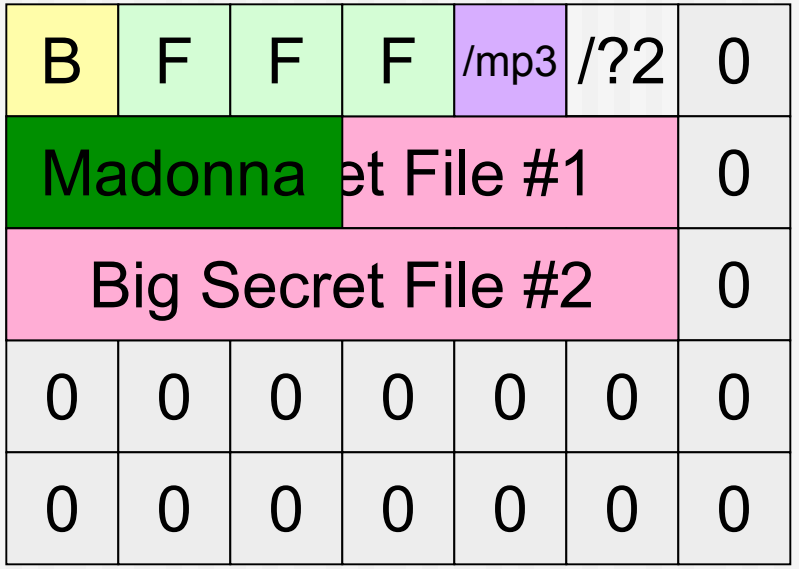
- new root directory:

```
Madonna_.mp3   jan 2 2004   block  7
?2_____.___    jan 1 2004   block 14
```


© 2004 Simson L. Garfinkel

What's on the disk?

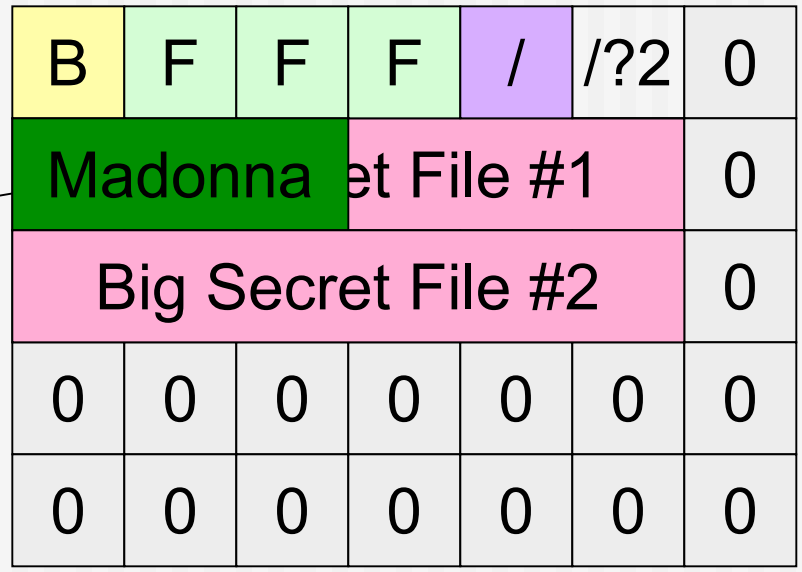
- Madonna.mp3
- Madonna.mp3's directory entry
- All of B2
- Most of B2's directory entry
- Part of B1



© 2004 Simson L. Garfinkel

% format C: Again!

- Writes:
  - Boot blocks
  - Root directory
  - “File Allocation Table” (FAT)
  - Backup “superblocks” (UFS/FFS)
- May also:
  - Validate surface



© 2004 Simson L. Garfinkel

Taxonomy of hard disk data

Level 0	Files in file system
Level 1	Temp files (/tmp, /windows/tmp, etc)
Level 2	Recoverable deleted files
Level 3	Partially over-written files

© 2004 Simson L. Garfinkel

Digital Forensics

- "Forensics" has two meanings:
  - The art or study of formal debate
  - The use of science and technology to investigate and establish facts in criminal or civil courts of law
- Digital Forensics:
  - Disk drive forensics
  - Network forensics
  - Software forensics

© 2004 Simson L. Garfinkel

Hard Disk Forensics

- Consumer Tools:
  - Disk sector editors
  - Norton Disk Doctor
- Professional Tools:
  - Access Data's Forensic Tool Kit (FTK)
  - Guidance Software's EnCase
- Open-Source Tools:
  - SleuthKit

© 2004 Simson L. Garfinkel

Capabilities of Forensic Tools

- All tools:
  - Undelete files (level 2 data)
  - Search for text (level 3 data)
- Professional Tools:
  - Display contents of Outlook .PST files
  - Search for files by MD5 or SHA-1
  - Create report of operator's actions
  - Create "timeline" of disk's activity

© 2004 Simson L. Garfinkel

The Forensics Challenge

- Most forensic tools are designed to spend a lot of time with one drive.
- I had a lot of drives and a little bit of time
- Tools that I used/created:
  - strings(1)
  - fatdump - a “forensic file system”
  - blockstats - forensics based on statistical analysis
  - level0 - Cataloging of existing files with MD5 factoring.

© 2004 Simson L. Garfinkel

“Automated Forensics:” Automatically find the good stuff

- Automatic searching for credit-card numbers
- Most common email address
- Searching for medical terms
- Combined timeline of all disks

© 2004 Simson L. Garfinkel

Email stop list: addresses to ignore!

111	c2le@mz.um	not an e-mail address
76	keywitness@keywitness.ca	something SSL related
71	cps-requests@verisign.com	""
70	server-certs@thawte.com	""
70	premium-server@thawte.com	""
56	1bf@g.ec	not an e-mail address
55	enews@microsoft.nwnet.com	
54	personal-premium@thawte.com	something SSL related
53	inet@microsoft.com	
52	personal-freemail@thawte.com	THAWTE personal freemail CA
52	personal-basic@thawte.com	THAWTE personal basic CA
51	mazrob@panix.com	Authors of Utopia sound scheme for Windows 95,
41	java-security@java.sun.com	Java stuff
41	java-io@java.sun.com	""
38	me@mycompany.com	Word Templete, "Elegant Fax.dot"
37	mbenson@msn.com	included in Word Template "Professional Resume.dot"
37	dgreer@mycompany.com	included in Word Template "Contemporary Resume.dot"

© 2004 Simson L. Garfinkel

Findings… Reloaded

- Jan 2002: 150 drives
- Jan 2004: 235 drives

- Drives DOA: 59
- Drives Imaged: 176
- Total files: 168,459
- Total data: 125G

Open-Source Security

Remembrance of Data Passed: A Study of Disk Sanitization Practices

Many discarded hard drives contain information that is both confidential and recoverable, as the authors' own experiment shows. The availability of this information is little publicized, but awareness of it will surely spread.

SIMSON L. GARFINKEL AND ABHI SHELAT
Massachusetts Institute of Technology

A fundamental goal of information security is to design computer systems that prevent the unauthorized disclosure of confidential information. There are many ways to assure this information privacy. One of the oldest and most common techniques is physical isolation: keeping confidential data on computers that only authorized individuals can access. Most single-user personal computers, for example, contain information that is confidential to that user.

Computer systems used by people with varying authorization levels typically employ authentication, access control lists, and a privileged operating system to maintain information privacy. Much of information security research over the past 30 years has centered on improving authentication techniques and developing methods to assure that computer systems properly implement these access control rules.

Cryptography is another tool that can assure information privacy. Users can encrypt data as it is sent and decrypt it at the intended destination, using, for example, the secure sockets layer (SSL) encryption protocol. They can also encrypt information stored on a computer's disk so that the information is accessible only to those with the appropriate decryption key. Cryptographic file systems[1–3] ask for a password or key on startup, after which they automatically encrypt data as it's written to a disk and decrypt the data as it's read; if the disk is stolen, the data will be inaccessible to the thief. Yet despite the availability of cryptographic file systems, the general public rarely seems to use them.

Absent a cryptographic file system, confidential information is readily accessible when owners improperly retire their disk drives. In August 2002, for example, the United States Veterans Administration Medical Center in Indianapolis retired 139 computers. Some of these systems were sold on the open market, and at least three ended up in a thrift shop where a journalist purchased them. Unfortunately, the VA neglected to *sanitize* the computer's hard drives—that is, it failed to remove the drives' confidential information. Many of the computers were later found to contain sensitive medical information, including the names of veterans with AIDS and mental health problems. The new owners also found 44 credit card numbers that the Indianapolis facility used.[4]

The VA fiasco is just one of many celebrated cases in which an organization entrusted with confidential information neglected to properly sanitize hard disks before disposing of computers. Other cases include:

- In the spring of 2002, the Pennsylvania Department of Labor and Industry sold a collection of computers to local resellers. The computers contained "thousands of files of information about state employees" that the department had failed to remove.[5]
- In August 2001, Dovebid auctioned off more than 100 computers from the San Francisco office of the Viant consulting firm. The hard drives contained confidential client information that Viant had failed to remove.[6]
- A Purdue University student purchased a used Macintosh computer at the school's surplus equipment exchange facility, only to discover that the computer's hard drive contained a FileMaker database containing the names and demographic information for more than 100 applicants to the school's Entomology Department.
- In August 1998, one of the authors purchased 10 used computer systems from a local computer store. The computers, most of which were three to five years old,

PUBLISHED BY THE IEEE COMPUTER SOCIETY ■ 1540-7993/03/$17.00 © 2003 IEEE ■ IEEE SECURITY & PRIVACY

© 2004 Simson L. Garfinkel

Zeroed drives (all 0s)

- 11 drives were zeroed
- Other drives from same vendors were not sanitized

Zeroed Drives	Vendor	# other working drives from vendor
#2	Driveguys.com	3*
#34	WeirdStuff	30
#72	eBay / PCSurplus	0
#82, 83, 84, 85, 86, 87, 88, 91	eBay / TSLi	3*

1 had just an OS

Purchased later…

© 2004 Simson L. Garfinkel

“Formatted Drives”

- Clean formatted
  - all 0s except for FAT and root directory
- Clean formatted with OS
  - FAT, root, & DOS or Windows install
- Dirty formatted
  - Lots of data, but with a clean FAT and root.

© 2004 Simson L. Garfinkel

Clean Formatted

- Easily identified with SQL:
  - img_blocks>0
    and img_blocks!=img_zblocks
    and img_blocks*0.01 > img_zblocks
- 22 drives were “clean-formatted.”
  - 1 from Driveguys (but other 2 had lots of data)
  - 18 from pcjunkyard (out of 25; 1 had parish data)
  - 1 from Mr. M. who sold his 2GB drive on eBay.
  - 1 from a VA reseller (1 DOA; 3 dirty formats)
  - 1 from unknown source (1 DOA; 1 dirty format)

© 2004 Simson L. Garfinkel

Clean format with OS

- Easily identified with SQL:
  - # blocks - # blocks in files where the MD5 is seen in more than one file

© 2004 Simson L. Garfinkel

MD5 factoring

- Register every found md5 in a database
- Allows quick determination of:
  - Unique files
  - Operating system files
  - Most common files
  - See: Garfinkel, S., A Web Service for File Fingerprints: The Goods, the Bads, and the Unknowns, January 2003.
- Coming soon: Factor blocks!
  - A 60GB file would have 3.6GB of MD5 codes…
  - Specialized database…

© 2004 Simson L. Garfinkel

Unique Files

- 783 Microsoft Word Files (!)
- 184 Microsoft Excel Files
- 30 Microsoft PowerPoint files
- 11 Outlook PST files!
- 977 audio files

- Notes:
  - This is a rapid way to find the good stuff!
  - Why so few unique files?

© 2004 Simson L. Garfinkel

Most common level 0 files

- “” (3235 copies)
- /Program Files/Internet Explorer/Connection Manager/00000001.tmp (2899 copies)
  - “204 No download Necessary”
- /WINDOWS/TEMP/~DFE014.TMP (143 copies)
- /WINDOWS/Temporary Internet Files/desktop.ini (104 copies)
- /WINDOWS/CURSORS/ARROW_IL.CUR (96 copies)
- /WINDOWS/Java/Packages/Data/TZ3P7BVN.DAT (82 copies)
- /WINDOWS/Temporary Internet Files/../space.gif (81 copies)
- …
- /msdos.sys (40 copies)
- /WINDOWS/SYSTEM/OLE2NLS.DLL (38 copies)

© 2004 Simson L. Garfinkel

More Data…

- Level 1 Files:
  - Web caches
    - Hotmail
    - Purchases
    - Pornography
  - Cookies
    - Authentication cookies

© 2004 Simson L. Garfinkel

More data…

- Level 3 data:
  - Credit card numbers
    - “comb” by A. Shelat
  - Email addresses

© 2004 Simson L. Garfinkel

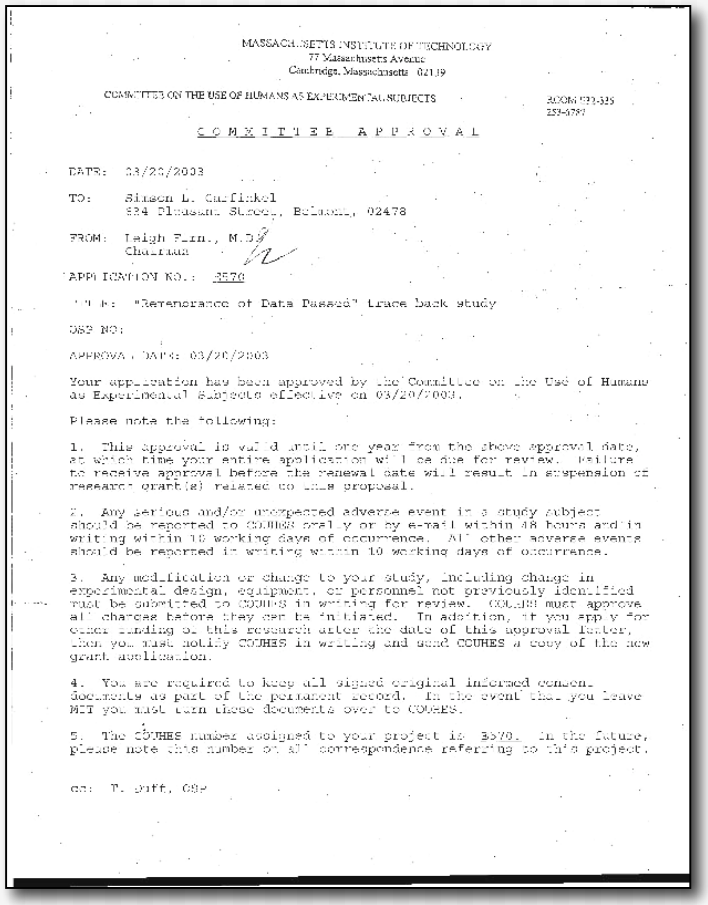
Confidential information found

- Medical records
- Short stories
- Personal correspondence
- HR correspondence
- Loan repayment schedules

© 2004 Simson L. Garfinkel

Trace back Study

- Started April 2003
- Required approve of MIT “Committee for of Humans as Experimental Subjects” (IRB)



MASSACHUSETTS INSTITUTE OF TECHNOLOGY
77 Massachusetts Avenue
Cambridge, Massachusetts 02139

COMMITTEE ON THE USE OF HUMANS AS EXPERIMENTAL SUBJECTS

ROOM 923-525
253-6787

COMMITTEE APPROVAL

DATE: 03/20/2003

TO: Simson L. Garfinkel
824 Pleasant Street, Belmont, 02478

FROM: Leigh Firn., M.D.
Chairman

APPLICATION NO.: 3570

TITLE: "Reverenance of Data Passed" trace back study

OSP NO:

APPROVAL DATE: 03/20/2003

Your application has been approved by the Committee on the Use of Humans as Experimental Subjects effective on 03/20/2003.

Please note the following:

1. This approval is valid until one year from the above approval date, at which time your entire application will be due for review. Failure to receive approval before the renewal date will result in suspension of research grant(s) related to this proposal.

2. Any serious and/or unexpected adverse event in a study subject should be reported to COUHES orally or by e-mail within 48 hours and in writing within 10 working days of occurrence. All other adverse events should be reported in writing within 10 working days of occurrence.

3. Any modification or change to your study, including change in experimental design, equipment, or personnel not previously identified must be submitted to COUHES in writing for review. COUHES must approve all changes before they can be initiated. In addition, if you apply for other funding of this research after the date of this approval letter, then you must notify COUHES in writing and send COUHES a copy of the new grant application.

4. You are required to keep all signed original informed consent documents as part of the permanent record. In the event that you leave MIT you must turn these documents over to COUHES.

5. The COUHES number assigned to your project is 3570. In the future, please note this number on all correspondence referring to this project.

cc: T. Piff, OSP

© 2004 Simson L. Garfinkel

Disk #6: Biotech Startup

- Memos & Documents from 1996
- Acquired Nov. 2000
- Company shut down; PCs disposed of without thought to contents.

© 2004 Simson L. Garfinkel

Disk #7: Major Electronic Manufacturer

- Company had a policy to clear data
- Policy apparently implemented with the FORMAT command
- New policy specifies DoD standard

© 2004 Simson L. Garfinkel

Disk #44

- Bay Area Computer Magazine
- Personal email and internal documents
- Many machines stripped and sold after a 70% reduction in force in summer 2000.
- No formal policy in place for sanitizing disks

© 2004 Simson L. Garfinkel

Disk #54

- Woman in Kirkland
- Personal correspondence, financial records, *Last Will and Testament*
- Computer had been taken to PC Recycle in Belleview by woman's son.
- PC Recycle charged $10 to "recycle" drive and sold it to me for $5.

© 2004 Simson L. Garfinkel

Disks #73, #74, #75, #77

- Community College (WA)
- Exams, student grades, correspondence, etc.
- Protect information under Family Educational Rights and Privacy Act!
- School did not have a procedure in place for wiping information from systems before sale, “but we have one now!”

© 2004 Simson L. Garfinkel

Disk #134

- Chicago bank
- Drive removed from an ATM machine.
- One year's worth of transactions; 3000+ card numbers
- Bank had hired contractor to upgrade machines; contractor had hired a sub-contractor.
- Bank and contractor assumed disks would be properly sanitized, but procedures were not specified in the contract.

© 2004 Simson L. Garfinkel

Main Sources of Failure:

- Failing or Defunct Companies
  - Nobody charged with data destruction
- Trade-ins and PC upgrades
  - Owner assumed that service provider would sanitize
- Failure to supervise contract employees
  - Sanitization was never verified

© 2004 Simson L. Garfinkel

USB Drives & Digital Cameras

- Everything about hard drives applies to other storage media that is treated as a “hard disk.”
- Most are formatted with FAT32

© 2004 Simson L. Garfinkel

Example: Digital Photography

- Many police have forced photographers to "delete" images they didn't want taken.
  - Ground Zero, post-9/11. Unnammed photographer forced by police to delete photos. Was able to recover with help from slashdot.
  - College student Mohammed Budeir, Philadelphia, Sept. 4, 2002, taking photographs of police cars. http://www.copcar.com/mo0902.htm
  - Airlines.net photographer Daniel Wojdylo, forced to delete photos photographed at BUF in April 2002.
- Google for:
  - officer made me delete pictures in my digital camera

© 2004 Simson L. Garfinkel

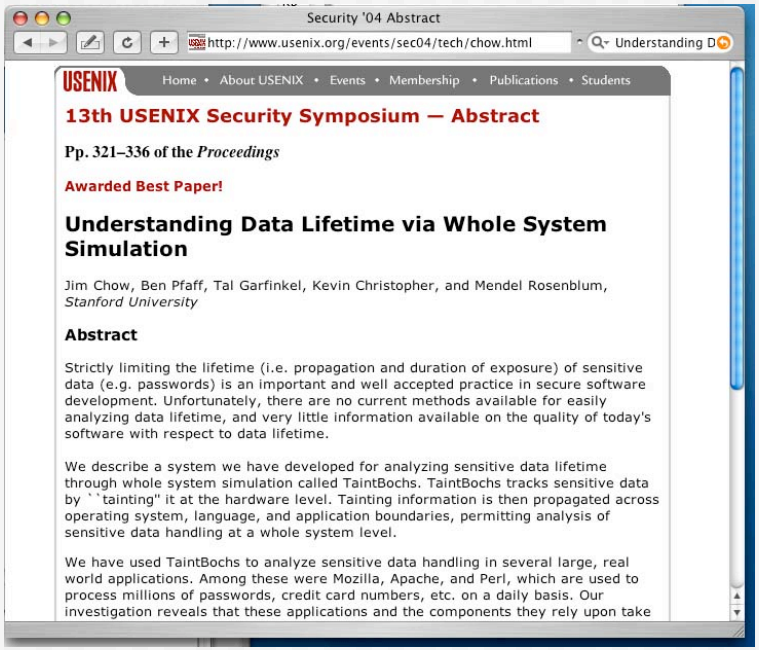
PalmOS 3.5.2 and others

- @Stake Security Advisory 3/01/2001
- Debugging back door:
  - dm - displays memory
  - saveimage - saves a memory image
- All Databases (including private entries), & delete information in memory!
- http://www.atstake.com/research/advisories/2001/a030101-1.txt

© 2004 Simson L. Garfinkel

“Virtually no limit to the lifetime of sensitive data.”

- “Understand Data Lifetime via Whole System Simulation,” Jim,Crow, Ben Pfaff, Tal Garfinkel, Kevin Christopher, Mendel Rosenblum,
- Best Paper, Usenix Security 2004



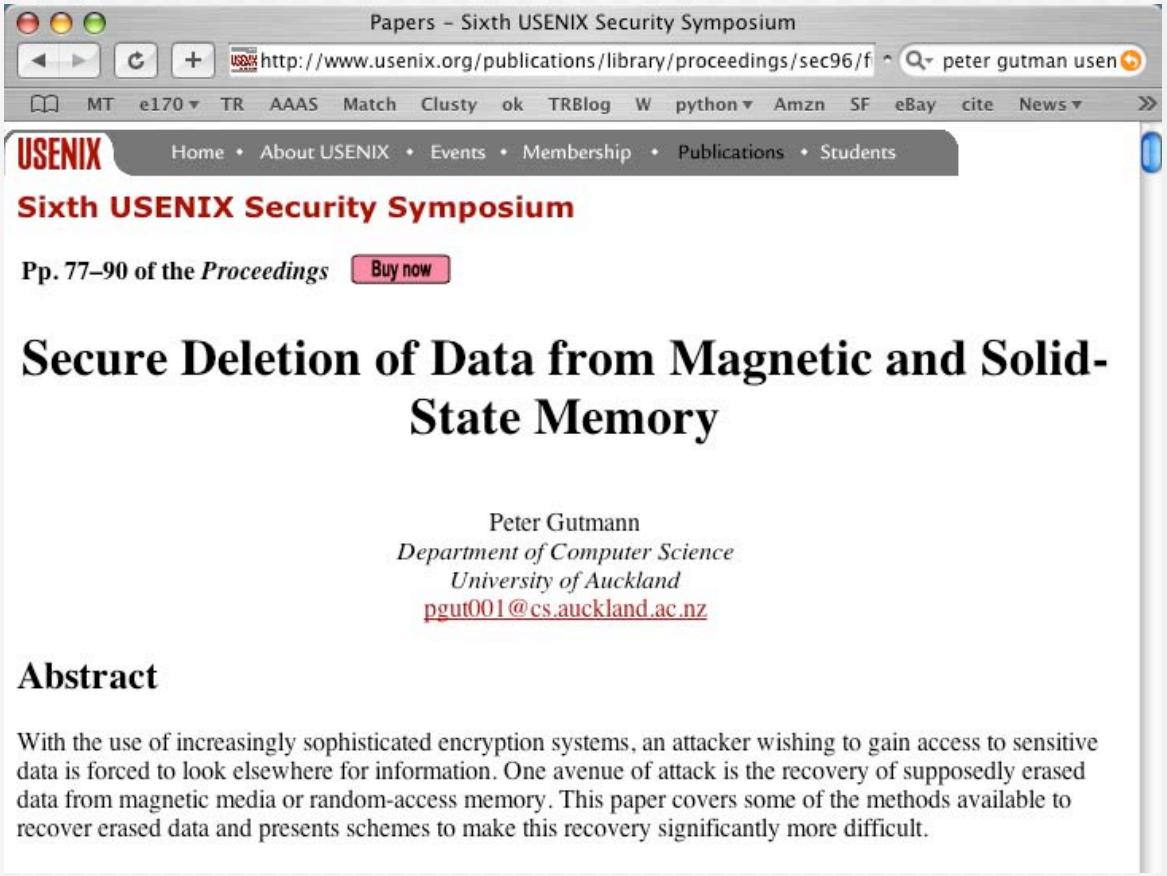
© 2004 Simson L. Garfinkel

What's the threat?

- Many people ask about recovering data that has been over-written

© 2004 Simson L. Garfinkel

Gutmann '96



© 2004 Simson L. Garfinkel

Gutmann Epilogue

- http://www.cs.auckland.ac.nz/~pgut001/pubs/secure_del.html
  - “some people have treated the 35-pass overwrite technique described in it more as a kind of voodoo incantation to banish evil spirits …”
  - “…performing the full 35-pass overwrite is pointless”
  - “For any modern PRML/EPRML drive, a few passes of random scrubbing is the best you can do.”
  - “This was true in 1996, and is still true now.”

© 2004 Simson L. Garfinkel

Overwritten Data…

- People from secret government agencies with advanced technology might be able to recover overwritten data…
- … but nobody else can.

© 2004 Simson L. Garfinkel

Threat Models: What are you afraid of?

- For most threats…
  - Snoop in the office
  - Data recovered from a discarded disk.
  - Disk seized by cops; data recovered.
- writing new data over old data should be sufficient…

© 2004 Simson L. Garfinkel

DOD 5220.22-M — standard for sanitizing media with *non-classified data*.

- "Degauss with a Type I degausser"
- "Degauss with a Type II degausser"
- "Overwrite all locations with:
  - a character,
  - it's complement,
  - then a random character
  - and verify"
- "Destroy, Disintegrate, incinerate, pulverize, shred, or melt."

© 2004 Simson L. Garfinkel

Tools for overwriting…

- dd if=/dev/zero of=/dev/ad2
- Stand alone tools:
  - AutoClave
    - http://staff.washington.edu/jdlarious/autoclave
  - DataGone
    - Now part of Symantec's professional offering.
  - SecureClean
    - http://www.bluesquirrel.com/so/secureclean/
  - DBAN: Darik's Boot and Nuke
    - http://dban.sourceforge.net/
- Suites
  - Norton Disk Doctor has a "wipe" feature.
- Missing: tools for *verifying* something is sanitized.

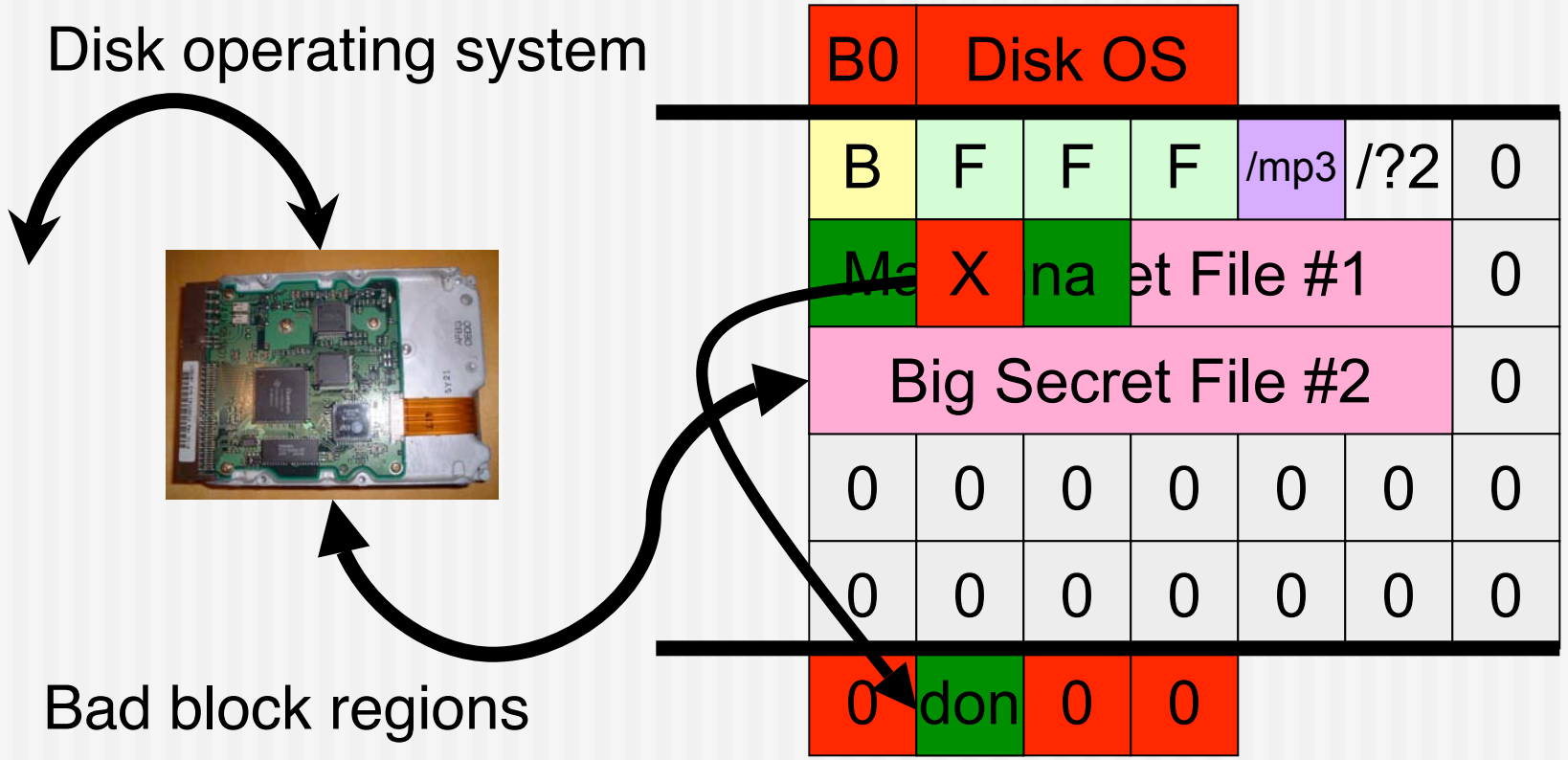
© 2004 Simson L. Garfinkel

Exotic Threat #2: Hostile Hard Disk

- "I'm bad; send me back for service."
- Scopes out data on *other* hard drives
- Lies when you try to try to sanitize it.

© 2004 Simson L. Garfinkel

Level 4 Data: Vendor Area



© 2004 Simson L. Garfinkel

Solutions for hostile hard drives

Approach #1:

- Write the entire disk with non-repeating data.
- Read the entire disk to make sure that the data is accurate.

Approach #2:

- Never write plaintext to the drive
- (This works for all cases…)

- Approach #3:
  - Never send hard drives back for service

© 2004 Simson L. Garfinkel

DOD 5220.22-M — standard for sanitizing media with *non-classified data*.

- "Degauss with a Type I degausser"
- "Degauss with a Type II degausser"
- "Overwrite all locations with:
  - a character,
  - it's complement,
  - then a random character
  - and verify"
- "Destroy, Disintegrate, incinerate, pulverize, shred, or melt."

© 2004 Simson L. Garfinkel

Type 1 Degausser

- Model HD-2000
- 73 seconds cycle time
- 260 lbs
- $13,995
- Monthly rental $1,400

- Note:
  - Your hard disk won't work after it's been degaussed (why not?)

http://www.datadev.com/v90.html

© 2004 Simson L. Garfinkel

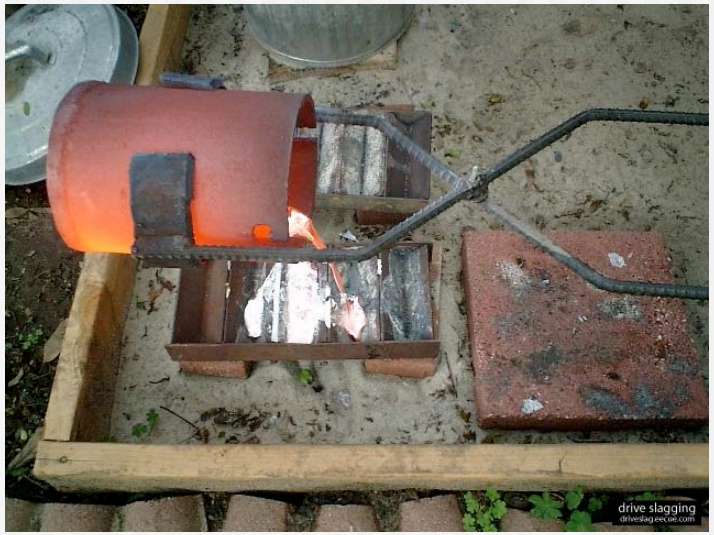
Drive Slagging: Melting the drive works just fine!

- Dave Bullock, John Norman, & CHS

http://driveslag.eecue.com/

© 2004 Simson L. Garfinkel

“Good luck removing data from this.”



“Our prognosis: drive slagging is a fool-proof method to prevent data recovery.”

© 2004 Simson L. Garfinkel

The Bad News:

- Most people aren't using these techniques…
- Most people are using "del" and format.
- This is an issue that *must* be addressed by OS vendors *in the kernel.*
  - Add-on software doesn't work
  - Even programs like CIPHER.EXE don't work

© 2004 Simson L. Garfinkel

Thoughts…

- Do we really want computers to give us "strong delete?"
- In legal "discovery," is the opposing side entitled to:
  - All of the files on your hard drive?
  - An image of your hard drive?
- If you delete a file, can you still be legally liable for having it?

© 2004 Simson L. Garfinkel

Bruce Mirken, 1999

- Gay journalist, advocate for rights of gay teenagers.
- Police man posing as a gay 14-year-old send Mirken child pornography
- Mirken deletes photographs.
- Police raid Mirken's apartment, use forensic software to recover deleted files.

- Case eventually dismissed ($50K in legal bills)

- http://www.journalism.sfsu.edu/flux/bayCurrents/mirken.html
- http://gaytoday.badpuppy.com/garchive/events/051799ev.htm
- **July 8, 1999, Page 3B, San Jose Mercury News**

© 2004 Simson L. Garfinkel

Michelle Theer

- Husband Air Force Capt. Marty Theer shot by Army Staff Sergeant John Diamond on Dec. 17, 2000
- Examination of computer's hard drive found:
  - 21,000 documents, mostly deleted.
  - Personal ads that Theer had written in 1999 and responses to the advt.
  - Theer active in swinger's clubs in winter & spring 2000
  - Affair between Diamond and Theer started in Spring 2000

© 2004 Simson L. Garfinkel

Final thoughts…

- Spending less than $1000 and working part time, I was able to collect:
  - Thousands of credit card numbers
  - Detailed financial records on hundreds of people
  - Confidential corporate files
- Who else is doing this?

© 2004 Simson L. Garfinkel