New Directions in Disk Forensics



Simson L. Garfinkel January 15, 2006, 3:15pm

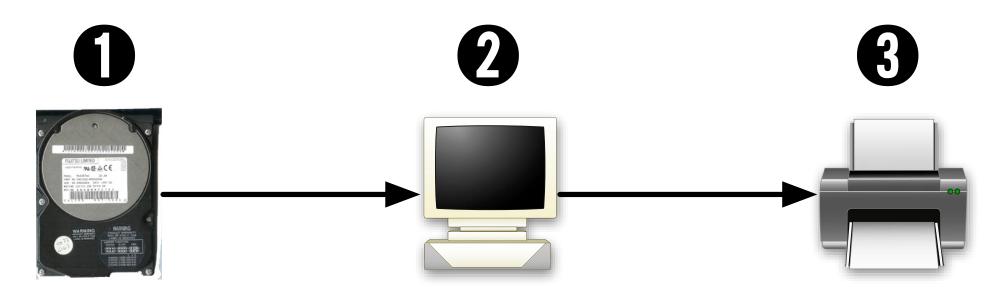
Postdoctoral Fellow, Center for Research on Computation and Society Harvard University Consulting Scientist, Basis Technology Corp.

Here are 200 hard drives



Which contain the email address "simsong@media.mit.edu"?

Today's forensic tools and file formats are designed to analyze a single drive at a time.



These tools are not adequate for today's forensic challenge.

This talk presents new tools and techniques for performing forensic analysis on a large number of disk drives.

	5 X	100		rry 🎯 Tirreline 🧟				
	of Stintory #4+		Nation 3 Are you being naugo -	Pron Secto Class	To and a later of the later of	Dubject.	Text.	Febale Repairs
Hone () Atachine	its		3 Got your meccage	Ober Warfare	verfake@addphia.ref			Return-Paths co
S-CE (End		123		Give Pake	unfoleDatebla.cet		10/24/03 10:4712791	
B-CC C PST Solute			A Heating this weekend	Marcard Ferzusio	verhile@advictia.cm	Pleasing this reveland	20/24/22 10 42:29/91	Return-Paths on
9-07 🙆 Outoe			3 Need lead help	Teddy Bear	verficke@addphia.ref	Need least help	20/25/02 12 14:05491	Return-Path:
	1 A T		For Are you availab.		Willow Pole		10/24/03 10:55:58PM	Estury Paths into
-OT Colleged Items			Welcome to Yahoo!		vehicepadephare			Repair-Path ch
-001 00 2000	4	0.0	A Melcane to Tabast	Tabool Member Serv	verfiele@addphia.nd	Walcone to Yahool Plu	20/00/02 09:22:144/1	Return-Paths on
-CCI 🎦 Nota								
-CC ն Sent								
L-CC C Tool		۰.						3
Diffed [] Her III !	store @Pepot 2	Console	ig tetals 🗖 tek (1/1779/7				
From	Santa Claus							<u> </u>
To	wmfiskei2ed	de bion	107					
Subject	Are you bein:							
Sent	10/25/03 12	1:30A	M.					
Folder	Inbox							
EntryPath	FisheACADoci	ments	and Settings/Will	am/Local Settin	gs\Application			
	Data/Microso	(tiOutic	iol@Outlook.pstP1	ST Volume/Roo	folder/Top of Pe	rsonal		
	Folders/Inbox	lAre yo	u being naughty?					
64								
Bil,								
Bill, I am watching you	ı.							- 11

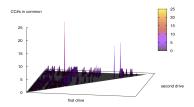
The state of the art



The drives Project



The Traceback Study



Cross Drive Forensics and AFF

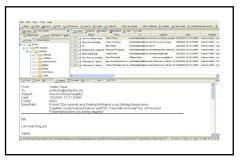
Digital forensics have opened up a whole new world for law enforcement and intelligence.

- Recovery of "deleted" files and email
- Automatic identification of "child pornography"
- Rapid searching for target names and email addresses



Tools of today's forensic trade:

GUI-based programs





Mirrored disks



Write blockers

Today's forensic work products:



Printed Reports



Courtroom Testimony

Today's forensic techniques don't scale.

Process is labor intensive.

Disk drives are getting bigger.

Law enforcement seizes more drives every year.

I am developing a different approach based on a different set of requirements.

Purchased used from a computer store in August 1998:



Computer #1: 486-class machine with 32MB of RAM

A law firm's file server...

...with client documents!

Computers #2 through #10 had:

- Mental health records
- Home finances
- Draft of a novel...

Was this a chance accident or common occurrence?



Hard drives pose special problem for computer security

Do not forget data when power is removed.

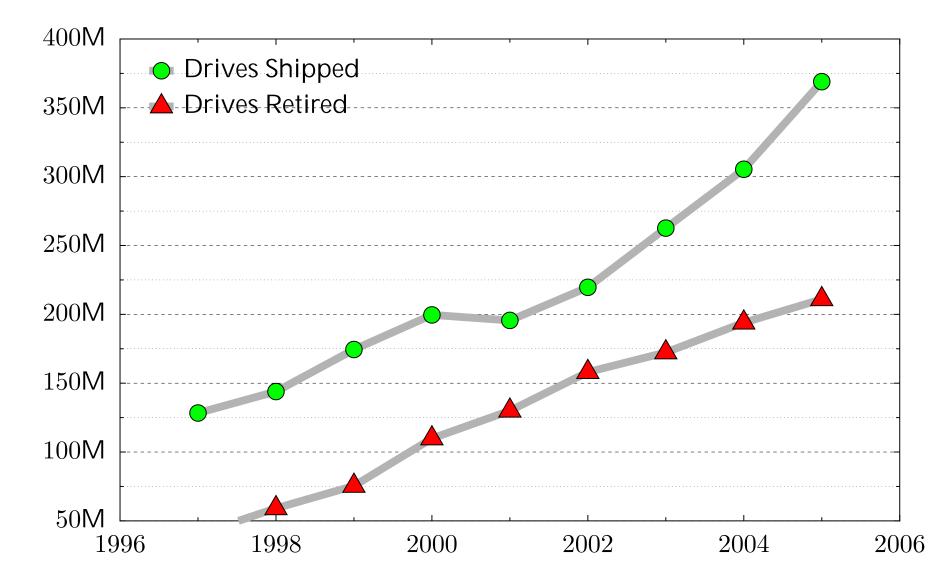
Contain data that is not immediately visible.

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



Scale of the problem: huge!



210 million drives will be retired this year.

Physical destruction will remove the information...







...but many "retired" drives are not physically destroyed.

There is a significant secondary market for used disk drives.

Retired drives are:

- Re-used within organizations
- Given to charities
- Sold at auction

350 items f	Second for hard drives not ending first peekly listed sweat priced highest priced hig	iced		Save this search	
Picture	Item Tide	Price	Bids	Time Left	
đ	Lot of hard and floppy drives	\$5.50	2	14n	
đ	Lot of hard and floppy drives	\$5.50	2	22n	
đ	Lot of hard and floppy drives	\$5.50	2	256	
đ	Lot of 2 hard drives IDE	\$8.00	12	29n	
	3.2 gig Hard Drives	\$180.00		59n	
đ	(5) 1.2 hard drives & (15) 10/100 network	\$25.00	1	1h 00p	
	Lot of 3 Quantum 9.1 gig SCSI Hard Drives	\$26.00	6	1h 25p	
	IDE HARD DRIVES (3)	\$6.50	6	1h 46n	
đ	LOT OF 5 Hard Drives 3.2 Oig Western Digital	\$120.00 \$124.95 #	Ny E Nor	1h 50p	
	OTY 3 IDE Hard Drives 2.5 Gig	\$20.50	5	2h 02n	
đ	5 WESTERN DIGITAL 2.5 GIG HARD DRIVES	\$30.00	4	2h 03a	
	QTY 3 IDE Hard Drives 1.0 Gig	\$9.99	1	2h 04p	
	Western Digital 850 meg IDE Hard Drives dutch	\$6.00	1	2h 57n	
-	WINDOWS		6.00	- 3h 18n	

About 1000 used drives/day sold on eBay.

There are roughly a dozen documented cases of people purchasing old PCs and finding sensitive data.

- A woman in Pahrump, NV bought a used PC with pharmacy records [Markoff 97]
- Pennsylvania sold PCs with "thousands of files" on state employees [Villano 02]
- Paul McCartney's bank records sold by his bank [Leyden 04]
- O&O Software GmbH 100 drives.[O&O 04]
- O&O Software GmbH 200 drives.[O&O 05]



None of these are scientifically rigorous studies.

I purchase hard drives on the secondary market.



Drives arrive by UPS and USPS



Some drives are purchased in person

10GB drive: \$19 "tested" 500 MB drive: \$3 "as is" Q: "How do you sanitize them?" A: "We FDISK them!"



Weird Stuff, Sunnyvale California, January 1999

Data on drives "imaged" using FreeBSD and Almage



Images stored on external firewire drives



This is 900GB of storage.

Note: I am not considering exotic recovery techniques.

I assume that writing a sector destroys its previous contents.

Some people claim that secret government agencies with advanced technology can recover overwritten data.

This technology has never been publicly demonstrated.



Even without the Men In Black, a lot of data can be recovered!

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:

drwxrwxrwx	0 root	0 Dec 31	1979 ./
-r-xr-xr-x	0 root	222390 May 11	1998 ID.SYS
-r-xr-xr-x	0 root	9 May 11	1998 MSDOS.SYS
-rwxrwxrwx	0 root	93880 May 11	1998 COMMAND.COM

Clearly, this disk was FORMATed...

C:\WINDOWS\system32\cmd.exe - format c:

```
C:\>format c:
The type of the file system is NTFS.
WARNING, ALL DATA ON NON-REMOVABLE DISK
DRIVE C: WILL BE LOST!
Proceed with Format (Y/N)?
```

FORMAT and FDISK overwrite very few disk sectors.

10 GB drive: 20,044,160 sectors

	Sectors	
Command	Written	%
FORMAT	21,541	0.11%
FDISK	2,563	0.01%

FORMAT erases the FAT, complicating the recovery of fragmented files.

UNIX "strings" reveals the disk's previous contents...

% 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 incompa version of Windows. Insert a Startup diskette matching thi

OEMString = "NCR 14 inch Analog Color Display Enchanced SV Graphics Mode: 640 x 480 at 72Hz vertical refresh. XResolution = 640 YResolution = 480

70.img con't...

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-RC single-file installation program over a network. The Tria is available in seven national languages, and each languag provided on a separate CC-ROM or as a separa EAS.STCm

EET.STC

ELR.STCq

ELS.STC

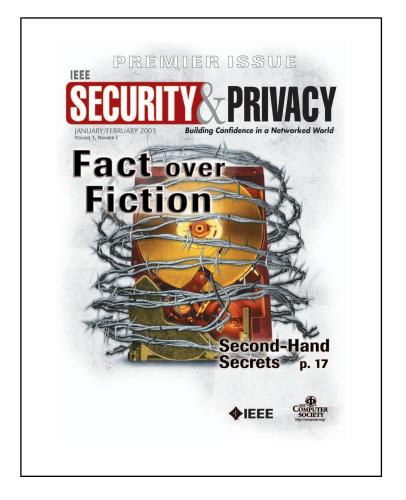
70.img con't...

MAB-DEDUCTIBLE MAB-MOOP MAB-MOOP-DED **METHIMAZOLE** INSULIN (HUMAN) COUMARIN ANTICOAGULANTS CARBAMATE DERIVATIVES AMANTADINE MANNITOL MAPROTILINE CARBAMAZEPINE CHLORPHENESIN CARBAMATE ETHINAMATE FORMALDEHYDE MAFENIDE ACETATE

[Garfinkel & Shelat 03] established the scale of the problem.

We found:

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

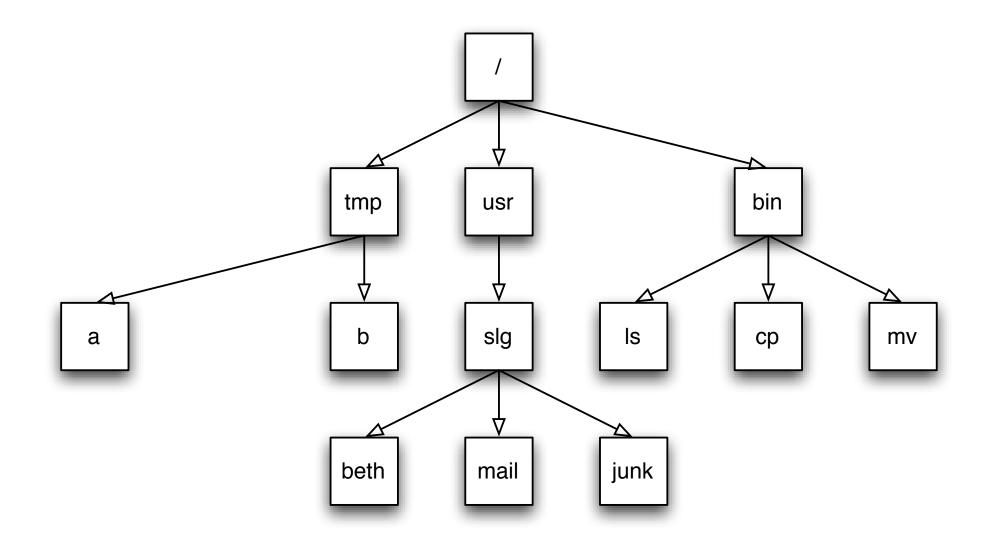


We did not determine why the data had been left behind.

Why don't we hear more stories?

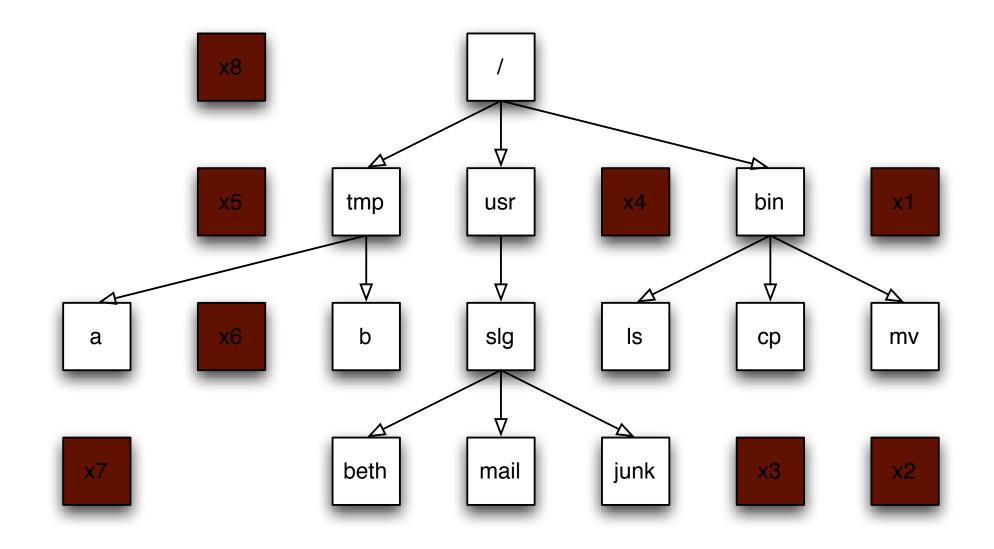
- Hypothesis #1: Disclosure of "data passed" is exceedingly rare because most systems are properly cleared.
- Hypothesis #2: Disclosures are so common that they are not newsworthy.
- Hypothesis #3: Systems aren't properly cleared, but few people notice the data.

Data on a hard drive is arranged in sectors.



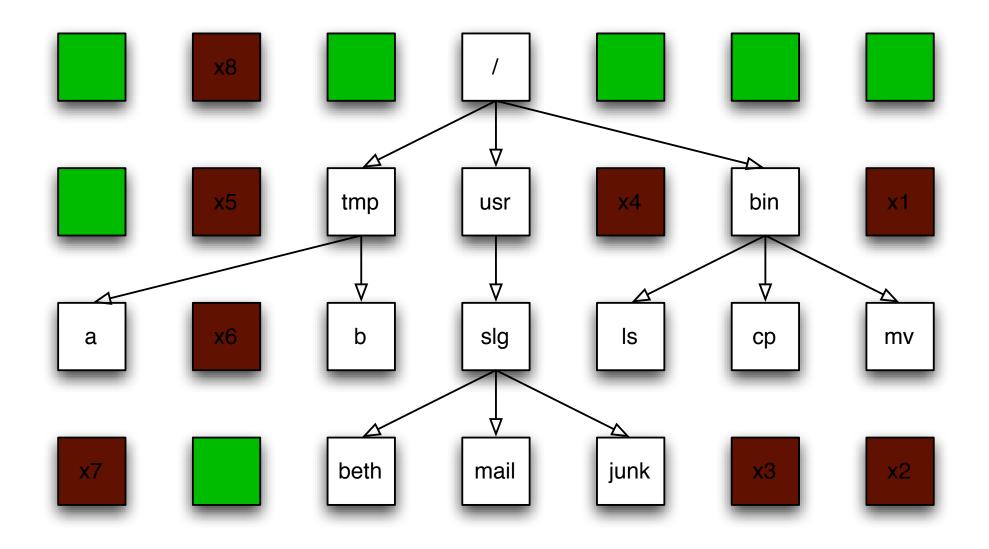
The white sectors indicate directories and files that are visible to the user.

Data on a hard drive is arranged in sectors.



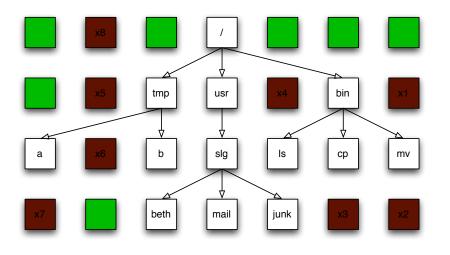
The brown sectors indicate files that were deleted.

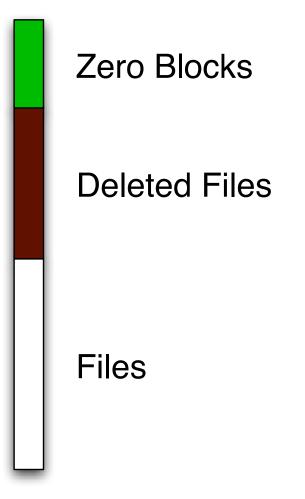
Data on a hard drive is arranged in sectors.



The green sectors indicate sectors that were never used (or that were wiped clean).

Stack the disk sectors:





•

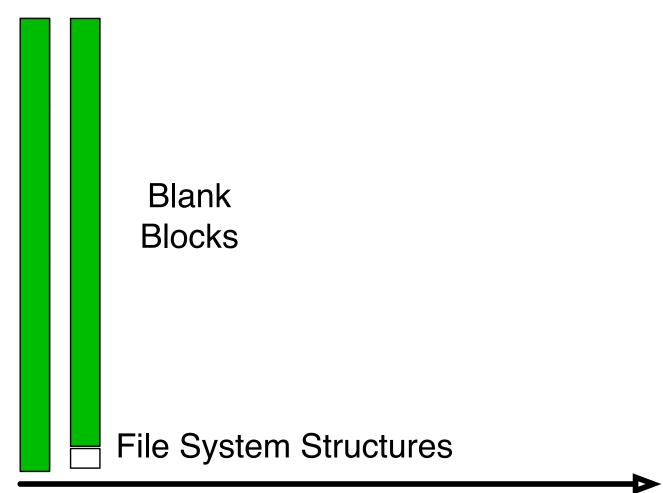
NO DATA: The disk is factory fresh.

All Blocks are Zero

.

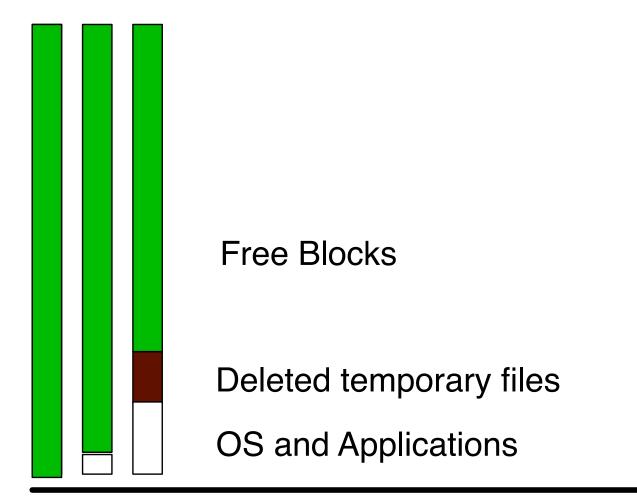
time

FORMATTED: The disk has an empty file system



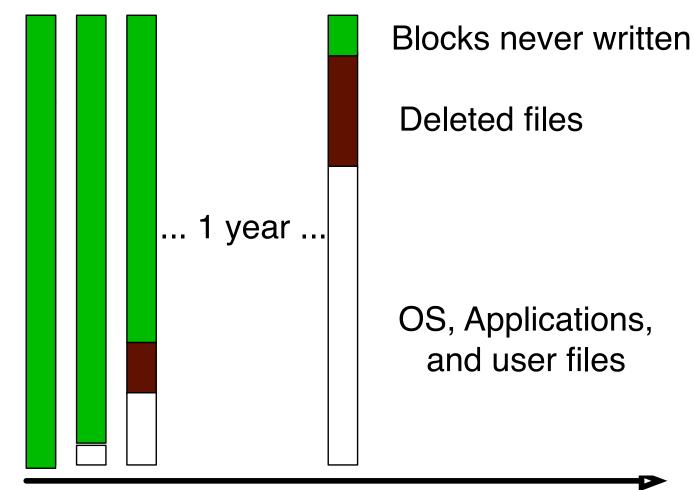
time

AFTER OS INSTALL: Temp. files have been deleted



time

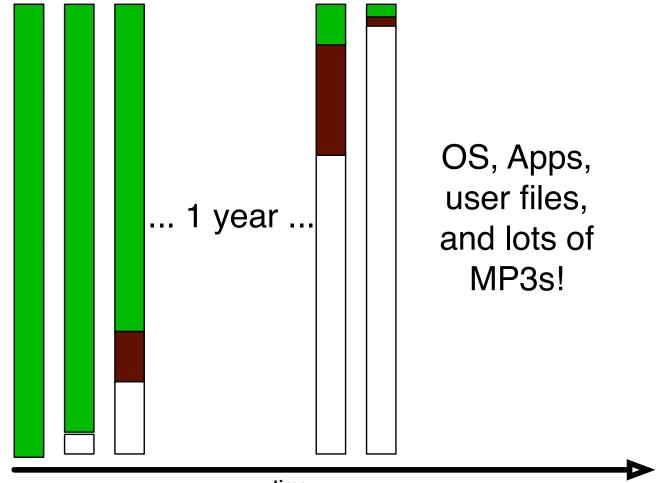
AFTER A YEAR OF SERVICE



time

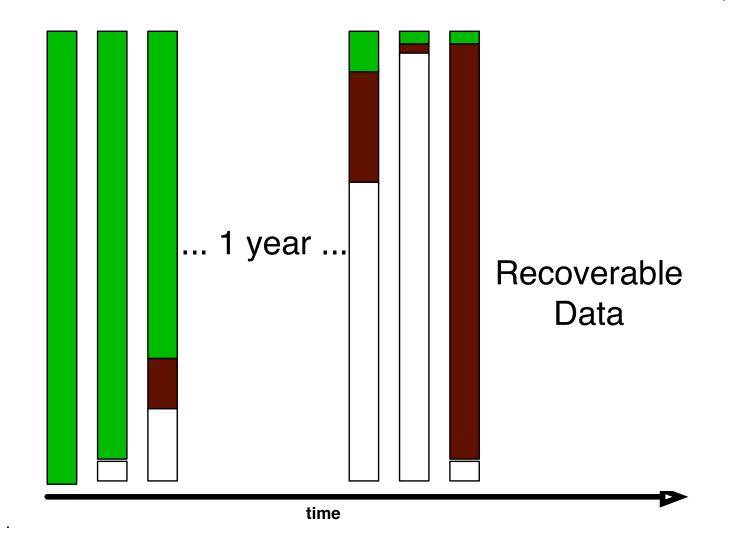
DISK NEARLY FULL!

.

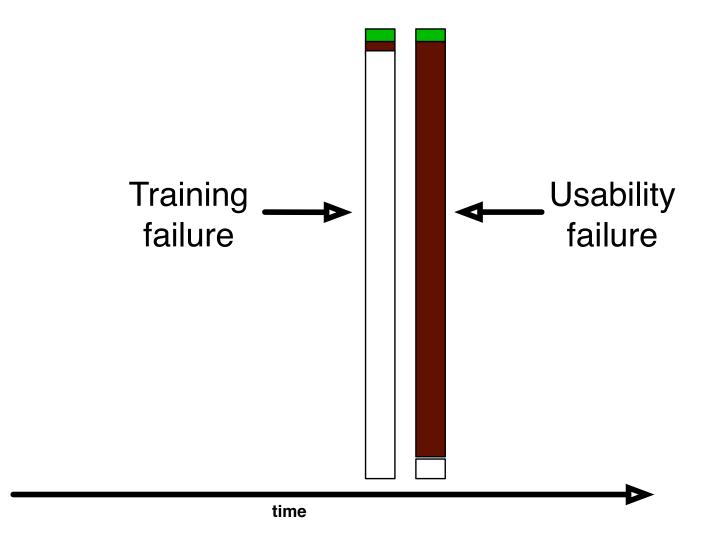


time

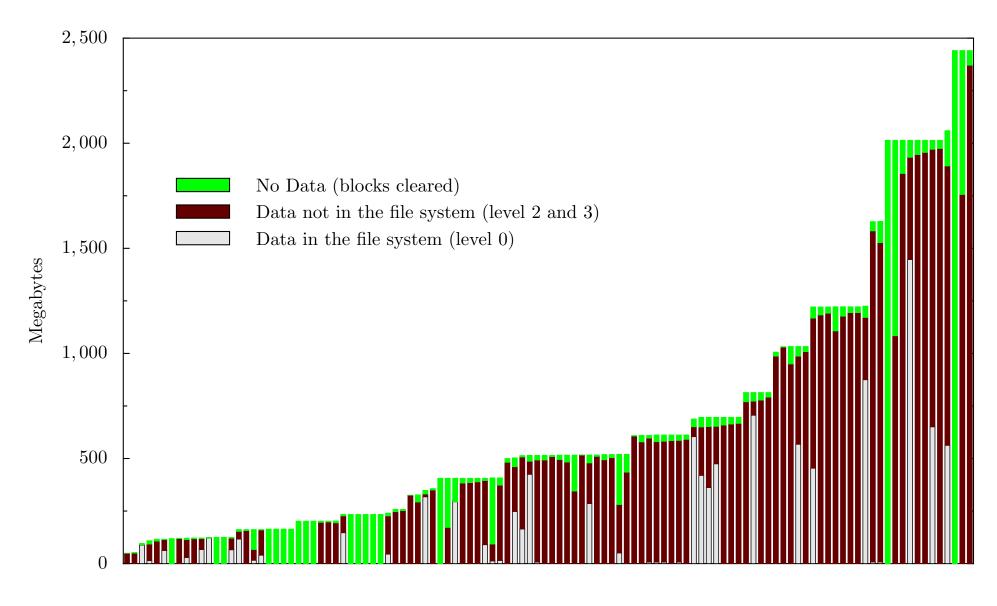
FORMAT C:\ (to sell the computer.)



We can use forensics to reconstruct motivations:



Drives 1–236 are dominated by failed sanitization attempts.



..but training failures are also important.

Overall numbers for the June 2005 report:

Total data:

Drives Acquired:	236
Drives DOA:	60
Drives Images:	176
Drives Zeroed:	11
Drives "Clean Formatted:"	22
Total files:	168,459

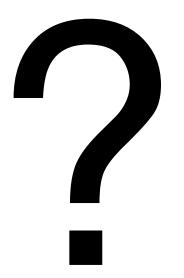
125G

Only 33 out of 176 working drives were properly cleared!

- 1 from Driveguys but 2 others had lots of data.
- 18 from pcjunkyard but 7 others had data.
- 1 from a VA reseller 1 DOA; 3 dirty formats.
- 1 from an unknown source 1 DOA, 1 dirty format.
- 1 from Mr. M. who sold his 2GB drive on eBay.

There is no consistency on which organizations deliver cleared drives.

But what *really* happened?



I needed to contact the original drive owners.

The *Remembrance of Data Passed Traceback Study.* [Garfinkel 05]

- 1. Find data on hard drive
- 2. Determine the owner
- 3. Get contact information for organization
- 4. Find the right person *inside* the organization
- 5. Set up interviews
- 6. Follow guidelines for human subjects work

06/19/1999 /:dir216/Four H Resume.doc
03/31/1999 /:dir216/U.M. Markets & Society.doc
08/27/1999 /:dir270/Resume-Deb.doc
03/31/1999 /:dir270/Deb-Marymount Letter.doc
03/31/1999 /:dir270/Links App. Ltrdoc
08/27/1999 /:dir270/Resume=Marymount Udoc
03/31/1999 /:dir270/NCR App. Ltrdoc
03/31/1999 /:dir270/Admissions counselor, NCR.doc
08/27/1999 /:dir270/Resume, Deb.doc
03/31/1999 /:dir270/UMUC App. Ltrdoc
03/31/1999 /:dir270/Ed. Coordinator Ltrdoc
03/31/1999 /:dir270/American Collegedoc
04/01/1999 /:dir270/Am. U. Admin. Dirdoc
04/05/1999 /:dir270/IR Unknown Lab.doc
04/06/1999 /:dir270/Admit Slip for Modernism.doc
04/07/1999 /:dir270/Your Honor.doc

This was a lot harder than I thought it would be.

Ultimately, I contacted 20 organizations between April 2003 and April 2005.



The leading cause: betrayed trust.

Trust Failure: 5 cases

- ✓ Home computer; woman's son took to "PC Recycle"
- Community college; no procedures in place
- Church in South Dakota; administrator "kind of crazy"
- Auto dealership; consultant sold drives he "upgraded"
- ✓ Home computer, financial records; same consultant

This specific failure wasn't considered in [GS 03]; it was the most common failure.

Second leading cause: Poor training and supervision

Trust Failure: 5 cases

Lack of Training: 3 cases

- California electronic manufacturer
- Supermarket credit-card processing terminal
- ✓ ATM machine from a Chicago bank

Alignment between the interface and the underlying representation would overcome this problem.

Sometimes the data custodians just don't care.

Trust Failure: 5 cases Lack of Training: 3 cases

Lack of Concern: 2 cases

- Bankrupt Internet software developer
- ✓ Layoffs at a computer magazine

Regulation on resellers might have prevented these cases.

In seven cases, no cause could be determined.

Trust Failure: 5 cases Lack of Training: 3 cases Lack of Concern: 2 cases

Unknown Reason: 7 cases

- ✗ Bankrupt biotech startup
- X Another major electronics manufacturer
- **X** Primary school principal's office
- ✗ Mail order pharmacy
- ✗ Major telecommunications provider
- X Minnesota food company
- **X** State Corporation Commission

Regulation might have helped here, too.

The techniques developed for [Garfinkel '05] are different than traditional forensics techniques.

Traditional forensics tools:

- Interactive user interface.
- Recovery of "deleted" files.
- Generation of "investigative reports" for courtroom use.
- Focus on one or a few disks.

Search Hits in the	a Anna Anna	-	Name	Erom	To	Subject	Sent	Header
-	mail 🕼 History 🚦		Are you being naug	Santa Claus	wmfiske@adelohia.net	Are you being paught:	10/25/03 12:31:30AM	Return-Path: <s< td=""></s<>
📑 Home 👔 Attachn	ients		Got your message	Cyber Warfare	wmfiske@adelphia.net	Got your message	10/08/03 02:26:57PM	Return-Path: <c< td=""></c<>
-O 🖂 Email			ан	Alice Fiske	wmfiske@adelphia.net		10/24/03 10:43:12PM	Return-Path: <a< td=""></a<>
B-D- BST Volu		4	Meeting this weekend	Maynard Fergusen	wmfiske@adelphia.net	Meeting this weekend	10/24/03 10:42:26PM	Return-Path: <m< td=""></m<>
		5	Need legal help	Teddy Bear	wmfiske@adelphia.net	Need legal help	10/25/03 12:14:09AM	Return-Path: <b< td=""></b<>
		6	📄 Re: Are you availab	Mac McDonald	William Fiske	Are you available???	10/24/03 10:55:58PM	Return-Path: <m< td=""></m<>
		7	Welcome to Yahoo!	geo-civics@yahoo-inc	wmfiske@adelphia.net	Welcome to Yahoo! Ge	10/08/03 09:40:28AM	Return-Path: <h< td=""></h<>
-011 🔂 In		8	Welcome to Yahoo!	Yahoo! Member Servic	wmfiske@adelphia.net	Welcome to Yahoo! Pl	10/08/03 09:23:14AM	Return-Path: <m< td=""></m<>
From: To: Subject Sent: Folder: EntryPath:		naug 1:30. ment	ahty?			rsonal		
	Folders\Inbox	Ares	/ou being naughty?					

In [Garfinkel '05], there were *hundreds* of disks to analyze.

Today's tools choke when confronted with thousands of disks.

- Has this drive been previously imaged?
- Which drives belong to my target?
- Do any drives belong to my target's associates?
- Where should I start?



Today's tools are for criminal investiations. Increasingly, we need tools for intelligence analysis.

Intelligence objectives can be furthered by correlating information from multiple drives.

- Where any drives were used by the same organization?
- What names/places/email addresses are in common?
- Which drives were used in a place or at a time of interest?



Example problem: Who owned this disk drive?

Approach #1: Look for Microsoft Word files and try to determine the owner.

- Needs forensic skill.
- Requires complete documents.

Approach #2: Compute a histogram of all email addresses.

- Works with any file system.
- Works with incomplete data.

The email histogram works even if you can't find any files.

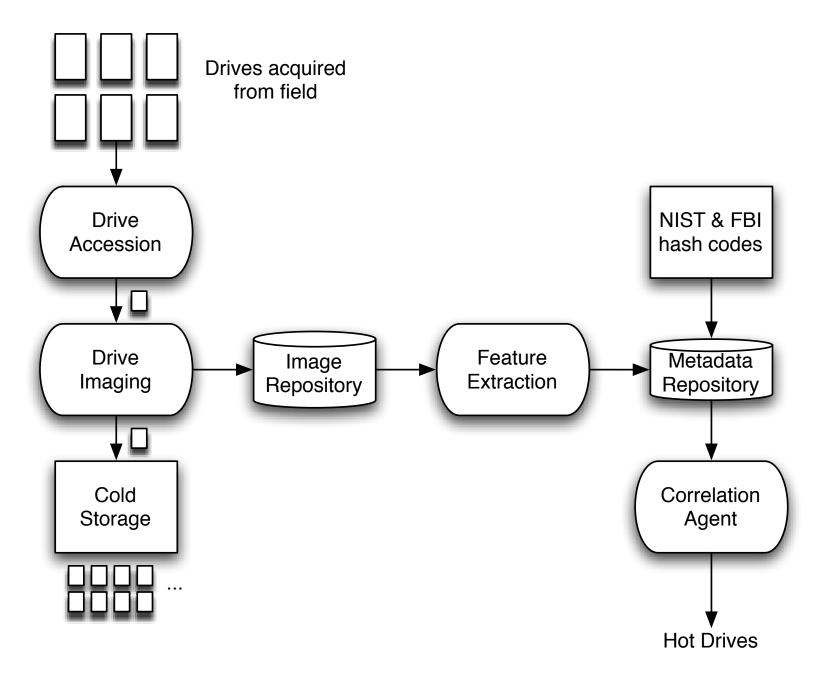
The email histogram approach works quite well.

Drive #51: Top email addresses (sanitized)

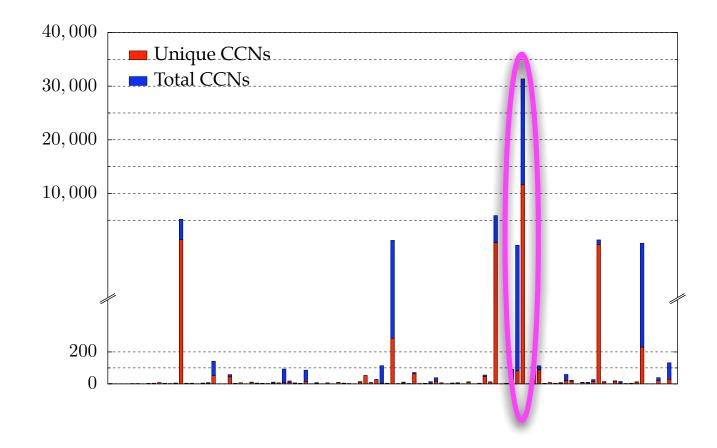
Count Address(es)

- 8133 ALICE@DOMAIN1.com
- 3504 BOB@DOMAIN1.com
- 2956 ALICE@mail.adhost.com
- 2108 JobInfo@alumni-gsb.stanford.edu
- 1579 CLARE@aol.com
- 1206 DON317@earthlink.net
- 1118 ERIC@DOMAIN1.com
- 1030 GABBY10@aol.com
 - 989 HAROLD@HAROLD.com
 - 960 ISHMAEL@JACK.wolfe.net
 - 947 KIM@prodigy.net
 - 845 ISHMAEL-list@rcia.com
 - 802 JACK@nwlink.com
 - 790 LEN@wolfenet.com
 - 763 natcom-list@rcia.com

Cross-Drive Forensics systematizes this approach.



"First Order Cross-Drive Forensics" analyzes each drive with a filter.



Drives with high response warrant further attention.

Example: The Credit Card Number Detector.

The CCN detector scans bulk data for ASCII patterns that look like credit card numbers.

- CCNs are found in certain typographical patterns.
 - (e.g. XXXX-XXXX-XXXX-XXXX
 - or XXXX XXXX XXXX XXXX
- CCNs are issued with well-known prefixes.
- CCNs follow the Credit Card Validation algorithm.
- Certain numeric patterns are unlikely. (e.g. 4454-4766-7667-6672)

CCN detector: written in flex and C++

Scan of disk #105: (642MB)

Test	# pass
typographic pattern	3857
known prefixes	90
CCV1	43
numeric histogram	38

Sample output:

'CHASE NA 5422-4128-3008-3685	pos=13152133
'DISCOVER 6011-0052-8056-4504	pos=13152440
.'GE CARD 4055-9000-0378-1959	pos=13152589
BANK ONE 4332-2213-0038-0832	pos=13152740
.'NORWEST 4829-0000-4102-9233	pos=13153182
'SNB CARD 5419-7213-0101-3624	pos=13153332

Even with the tests, there are occasional false positives.

CCN scan of Disk #115: (772MB)

Test	# pass
pattern	9196
known prefixes	898
CCV1	29
patterns	27
histogram	13

@: 4444486666108 :<@<74444:@@@<<44	pos=82473275
#"&'&&' 445447667667667 050014&'4"1"&'.	pos=86493675
221267241667& 454676676654450 &566746566726322.	pos=86507818
330210212676677 30232676630232 .1001.01	pos=86516059
"&#&&`&41&&`645445& 454454672676632 .3	pos=86523223
".#""#"&' 445467667227023	pos=87540819
D#9?.32400.,,+14%?B 499745255278101 *02)46+;<17756669	pos=118912826
.GGJJB>.JJGGG 3534554333511116 6	pos=197711868
$5 \dots } \} \} \} \dots \ 44444322233345 \dots \} \} \} \} \dots$	pos=228610295
)6"!) .&*%,,%-0)07. 373484553420378 <67<038+.5(+0+.3.	pos=638491849
)6"!) .&*%,,%-0)07. 373484553420378 <67<038+.5(+0+.3.	pos=645913801

Results of scanning 2003 corpus with CCN scanner:

Total number of image files:178Number of CCNs found:47,771Total number of distinct cards:15,613Most popular CCN6404 6521 6029 6650

(Seen 34 times on 30 drives)

Context analysis shows this is not a valid CCN:

[6]	6213 1 6758	6367	6404	6521	6029	6650	v	6025	6646 1 -138
[7]	6213 1 6758	6367	6404	6521	6029	6650	v	6025	6646 1 -138
[8]	6213 1 6758	6367	6404	6521	6029	6650	v	6025	6646 l -138
[10]	6213 1 6758	6367	6404	6521	6029	6650	v	6025	6646 l -138
[11]	6213 1 6758	6367	6404	6521	6029	6650	v	6025	6646 l -138
[11]	6213 1 6758	6367	6404	6521	6029	6650	v	6025	6646 l -138
[15]	6213 1 6758	6367	6404	6521	6029	6650	v	6025	6646 l -138
[18]	6213 1 6758	6367	6404	6521	6029	6650	v	6025	6646 l -138
[18]	6213 1 6758	6367	6404	6521	6029	6650	v	6025	6646 1 -138
[24]	6213 1 6758	6367	6404	6521	6029	6650	v	6025	6646 1 -138
[25]	6213 1 6758	6367	6404	6521	6029	6650	v	6025	6646 l -138

A "stop list" can be used for these common number.

Ignore "6404 6521 6029 6650' and we repeat the experiment:

Total number of image files: Number of CCNs found: Total number of distinct cards 15,612 (was 15,613) New "most popular CCN"

```
178
47,737 (was 47,771)
5501 8501 3501 3705
(Seen 35 times on 27 drives)
```

Once again, this does not appear to be a valid CCN:

[14]	3201	4901	:	5501	8501	3501	3705	5102yes.%d\Off
[112]	3201	4901	:	5501	8501	3501	3705	5102yes.%d\Off
[121]	3201	4901	:	5501	8501	3501	3705	5102yes.%d\Off
[128]	3201	4901	:	5501	8501	3501	3705	5102yes.%d\Off
[133]	3201	4901	:	5501	8501	3501	3705	5102yes.%d\Off
[181]	3201	4901	:	5501	8501	3501	3705	5102yes.%d\Off
[182]	3201	4901	:	5501	8501	3501	3705	5102 13505yes.
[184]	3201	4901	:	5501	8501	3501	3705	5102 13505yes.
[186]	3201	4901	•	5501	8501	3501	3705	5102 13505yes.

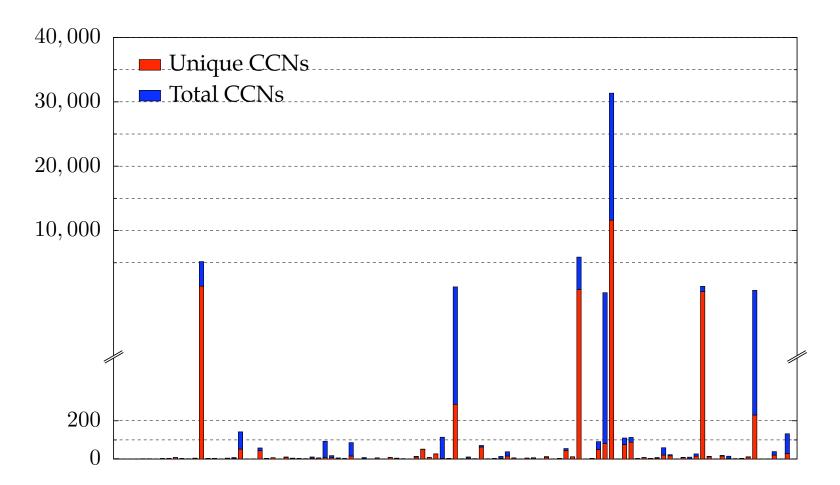
There are several problems with the "stop list" approach:

The list must be:

- Constructed.
- Maintained.
- Tuned for different applications.

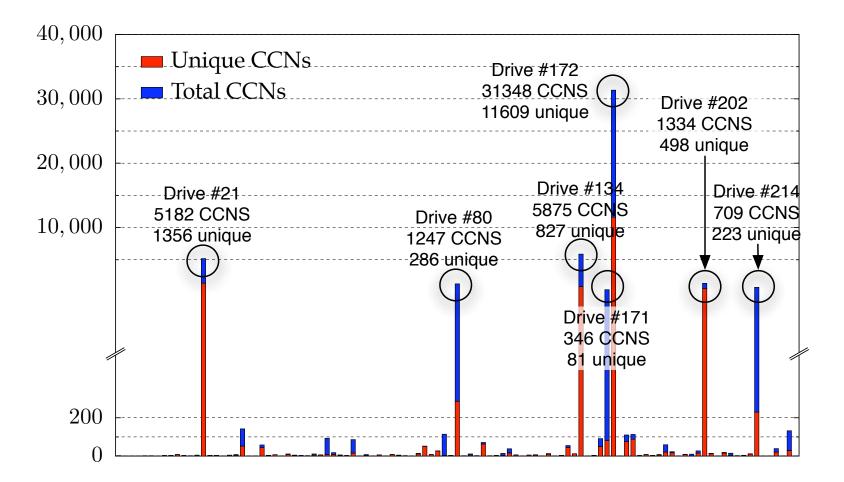
Building a "stop list" requires judgement and patience.

An alternative is to assume that "false positives" are rare and focus on those drives with high response.



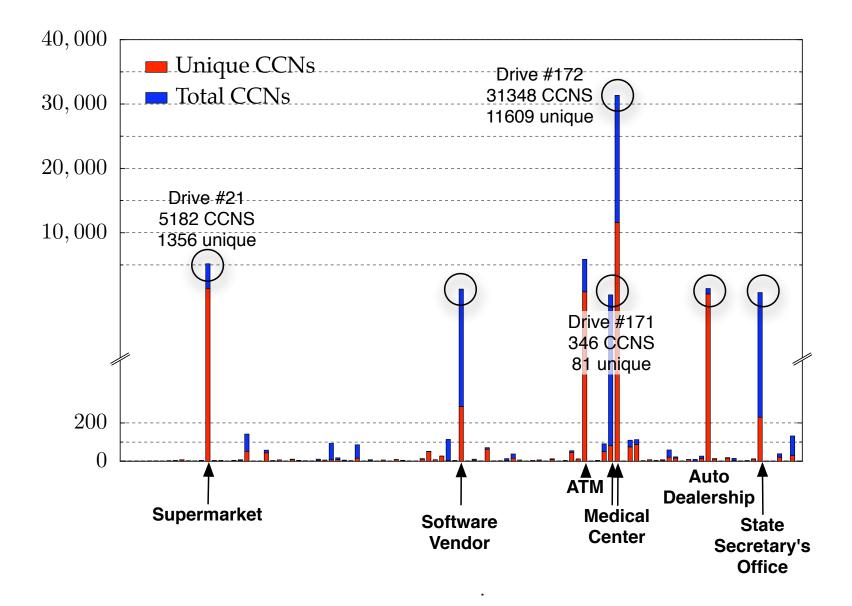
By definition, no drive should contain a large number of CCNs, so these drives are all interesting.

An alternative is to assume that "false positives" are rare and focus on those drives with high response.



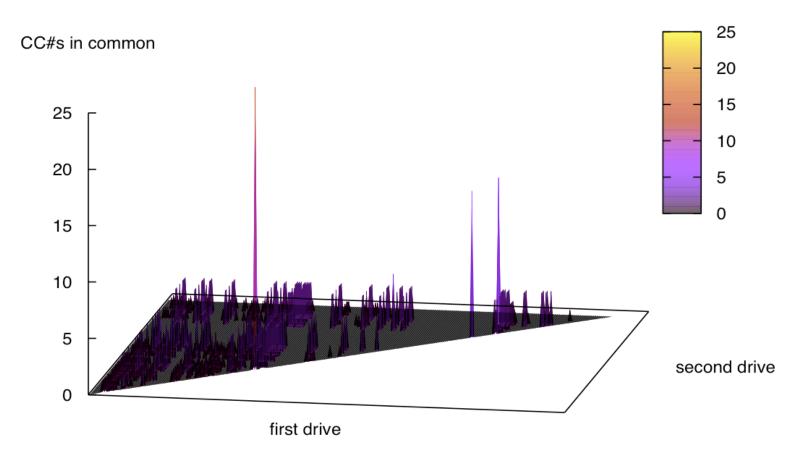
Only 7 drives had more than 300 credit card numbers.

These drives were traced back to their original owners.



Second-order analysis uses correlation techniques to identify drives of interest.

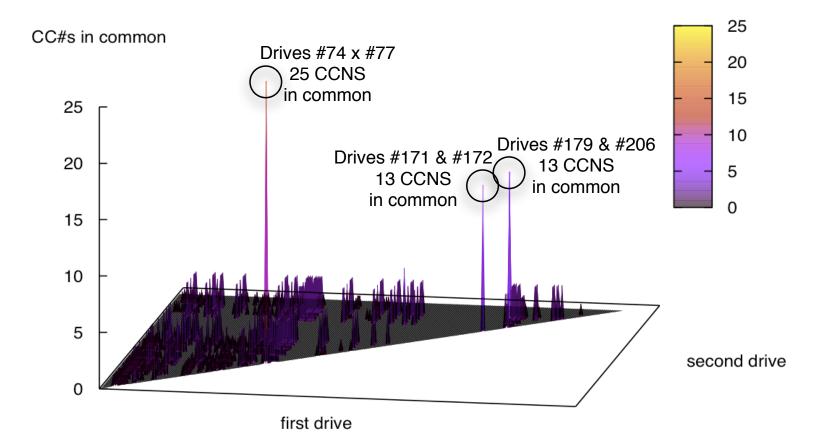
Cross Drive Correlation



In this example, three pairs of drive appear to be correlated.

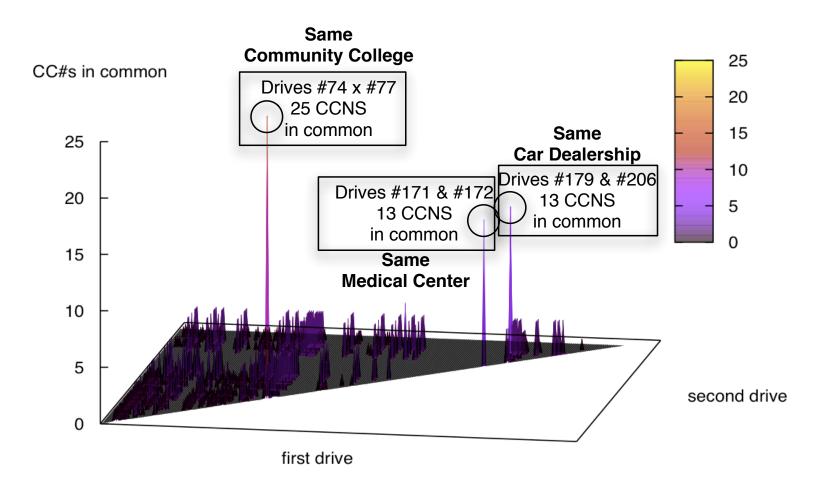
Second-order analysis uses correlation techniques to identify drives of interest.

Cross Drive Correlation



Manual analysis of on-drive data reveals that these drives are from the same organization.

Cross Drive Correlation



Second-order applications:

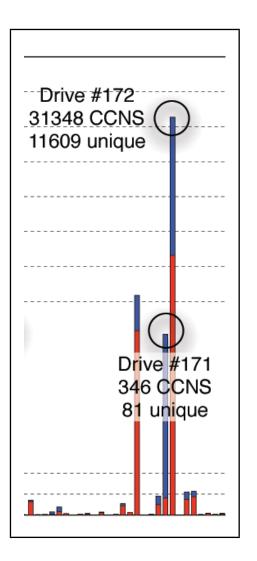
Possible Identifiers:

- CCNs
- Email addresses
- Message-IDs
- MD5 of disk sectors

Possible Uses:

- Identifying new social networks
- Testing for inclusion in an existing network.
- Measuring dissemination of information

Let's look at drives #171 and #172 again.



Cross-drive analysis tells us that #171 and #172 are from the same medical center. Drive #171: Development drive

• Has source code.

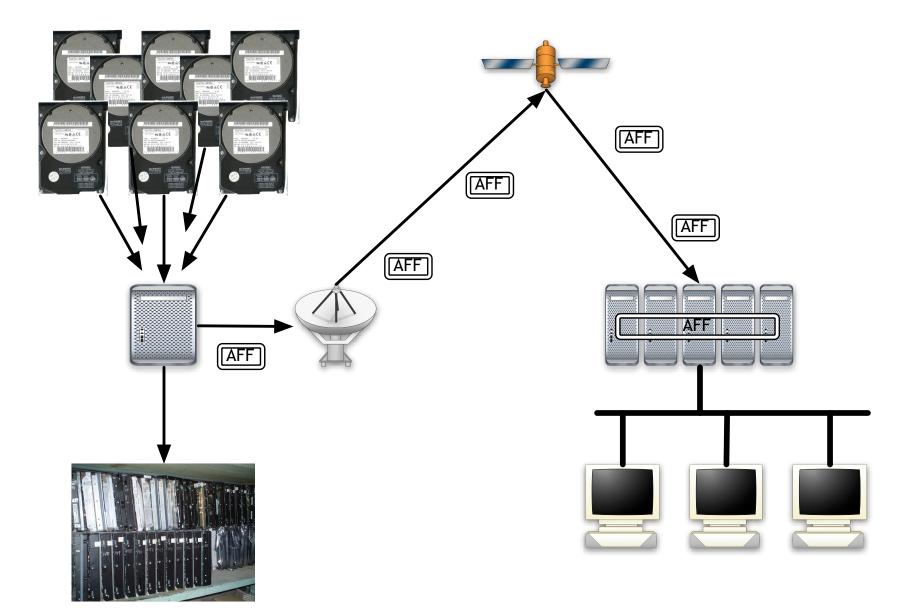
• 346 CCNS; 81 unique.

Drive #172: Production system.

- 31,348 CCNS; 11,609 unique
- Oracle database (hard to reconstruct).

The programmers used live data to test their system.

AFF is a simple, compact, self-describing, and open way to move around disk images.



AFF is superior to existing disk imaging formats.

Raw:

- No compression
- No metadata

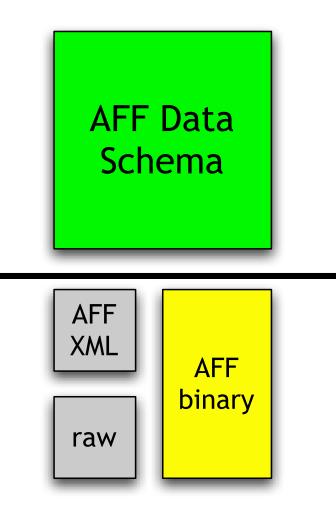
E01:

- Proprietary
- Limited storage of metadata
- 2GB file size limit

AFF:

- No filesize limits (64-bit clean)
- Stores arbitrary metadata
- Compressed or raw
- Remembers "bad" sectors
- Extensible
- Smaller images than EnCase

The AFF specification is divided into two layers.



The AFF defines a series of name/value pairs.

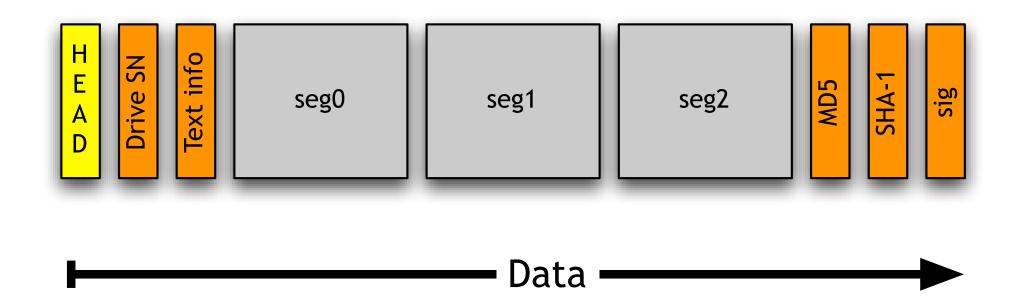
imagesize	# bytes in image
seg0	First data segment
seg1	Second data segment
segsize	data segment size
md5	MD5 of uncompressed disk image
sha1	SHA1 of uncompressed disk image
badflag	pattern for "bad" sectors
case_num	Case number
image_gid	Unique 128-bit identifier
device_sn	Drive serial number
imaging_date	Date imaged
imaging_notes	Notes on imaging

New segment names can be added as the need arises.

. . .

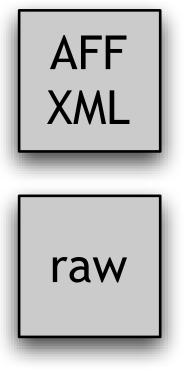
. . .

AFF stores disk images and metadata in named segments.



Segments can contain up to 2^{32} data bytes.

AFFXML can be used to annotate raw (dd) image files using the AFF schema.



<affinfo name='/project/affs/551.aff'> <segsize arg='16777216' /> <imagesize arg='2'>4290600960</imagesize> <md5>5M/87t7f6N4N5koMKb4QPA==</md5> <sha1>2XzW2A+kfKIGtmDloxiovqXjWcg=</sha1> <badsectors>AAAAAAAAAA=</badsectors> <blanksectors>AASGTQAAAAA=</blanksectors> </affinfo>

AFF TOOLS make it easy to work with AFF files.

afcat	Pipe AFF to <i>stdout</i>
afcompare	Compares two files
afconvert	Convert AFF↔raw
afinfo	Print information about an AFF
aftest	Validate the AFF Library
afxml	Output AFF metadata as XML

aimage The Advanced Disk Imager.

These tools work under FreeBSD, Linux, and Windows*.

AFFLIB makes it easy to add AFF support to existing programs.

```
AFILE *af;
af = af_open(filename,O_RDONLY,0666);
af_seek(af,offset,whence);
af_read(af,buf,count);
af_write(af,buf,count);
af_close(af);
```

26 functions supporting streams & name/value access.

Support has been added to Brian Carrier's Sleuth Kit.

Almage is the Advanced Disk Imager.

00	Terminal -	– ssh — 80x24
Source device: Model #: firmware: S/N:	QUANTUM FIREBALL ST3.24	Thu Nov 10 10:53:27 2005 AFF Output: /project/junk.aff Sector size: 512 bytes Total sectors:6,306,048
	ing sector: 97,792 (512 ctors read: 98,304 (1.5	· · · · · · · · · · · · · · · · · · ·
Time spe	nt reading: 00:00:05 bytes read: 50,331,648	Estimated total time left: 00:21:34
Time spent co Overall compres	es written: 25,735,396 ompressing: 00:00:09 sion ratio: 48.87% (0% 92.168.1.1:/project: 68,	3 is none; 100% is perfect) 937 MB (12.44%)

Almage features.

Writes AFF and/or raw (dd)

Automatically reads drive SN and other drive metadata

Compress during imaging or afterwards

MACs & signatures for segments and/or image

Intelligent error recovery

$\Theta \Theta \Theta$	Termina	al — ssh — 80x24
Source devica Model #: firmware: S/N:	QUANTUM FIREBALL ST3 A0F.0800	AFF Output: /project/junk.aff
	nding sector: 97,792 (5 Sectors read: 98,304 (
Time s	pent reading: 00:00:05 l bytes read: 50,331,64	Estimated total time left: 00:21:34
	i bytes feuu: 50,331,64	•

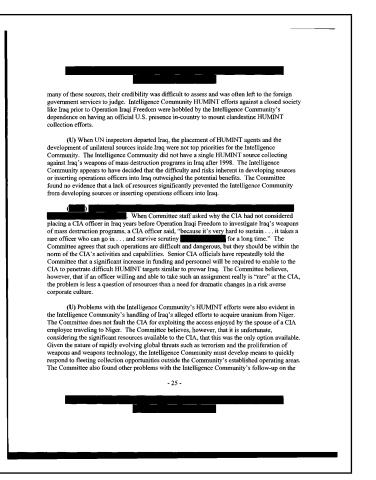
Advanced Forensics Format 1.0 in summary:

- AFF Specification: Simple, self-describing, robust
- AFF Library: open source AFF implementation
- AFF Tools: For working with AFF files
- Aimage: easy-to-use imager

Download from http://www.afflib.org

Future Work: Deploying Compete Delete

- Make FORMAT actually erase the disk.
- Make "Empty Trash" actually overwrite data.
- Integrate this functionality with web browsers, word processors, operating systems.
- Address usability dangers of clean delete.
- Analysis of "one big file" technique.



Future Work: 2500 Drive Corpus

- Automated construction of stop-lists.
- Detailed analysis of false positives/negatives in CCN test.
- Explore identifiers other than CCNs.
- Support for languages other than English.

Future Work: AFF Toolkit

- Improved imaging, storage and backup.
- Web-based database of hash codes.

	IMAGIN		nu Nov 10 10:53:27 2005
Source devic		AFF Output: /pro	oject/junk.aff
Model #: firmware:	· · · · · · · · · · · · · · · · · · ·	.ZA Sector size:	512 hutas
	153718340531	Total sectors:6,	,
3			
	ading sector: 97,792 (5 Sectors read: 98,304 (:		026
	pent reading: 00:00:05 l bytes read: 50,331,64		time left: 00:21:34
Time spent Dverall compr	ytes written: 25,735,39 compressing: 00:00:09 ession ratio: 48.87%	(0% is none; 100% is	: perfect)
Free space on	192.168.1.1:/project: (68,937 MB (12.44%)	

Future Work: Economics and Society

- Who is buying used hard drives and why?
- Compliance with FACT-A

All Categories Save this search 350 items found for hard drives Sort by items: ending first <u>revely listed lowest priced</u> <u>highest priced</u>				
Picture	item Title	Price	Bids	Time Left
đ	Lot of hard and floppy drives	\$5.50	2	14n
	Lot of hard and floppy drives	\$5.50	2	22n
đ	Lot of hard and floppy drives	\$5.50	2	25e
	Lot of 2 hard drives IDE	\$8.00	12	29n
	3.2 gig Hard Drives	\$180.00		59n
1	(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
đ	LOT OF 5 Hard Drives 13.2 Gig Western Digital	\$120.00 \$124.95 [#]	Bışt iku r	1h 50p
	QTY 3 IDE Hard Drives 2.5 Gg	\$20.50	5	2h 02n
đ	5 WESTERN DIGITAL 2.5 GIG HARD DRIVES	\$30.00	4	2h 03p
	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

Summary

A lot of information is left on used drives.

Working with these drives gives insights for improving forensic practice.

Cross drive forensics and AFF are two tangible benefits to date.

There is a lot more work to do.



Questions?

References

[Garfinkel & Shelat 03] Garfinkel, S. and Shelat, A., "Remembrance of Data Passed: A Study of Disk Sanitization Practices," *IEEE Security and Privacy*, January/February 2003. http://www.simson.net/clips/academic/2003.IEEE. DiskDriveForensics.pdf

[Markoff 97] John Markoff, "Patient Files Turn Up in Used Computer," *The New York Times*, April 1997.

[Villano 02] Matt Villano, "Hard-Drive Magic: Making Data Disappear Forever," *The New York Times*, May 2002.