Digital Forensics

Frank Adelstein

November 18, 2014

Guest Lecture for CS 5434
Defending Computer Networks

Overview

- Background
- Early forensics
- Modern forensics/trends
- Future

Background: Me

- Senior software engineer at Cayuga Networks focusing on evaluations (and other things)
- Background in distributed systems, network protocols, security, digital forensics, etc.
- Assorted hands-on hacking at various levels (OS, library, application, UI, etc.)
- Been involved in digital forensics R&D for 10+ years including organizer for Digital Forensics Research Workshop (DFRWS) since 2005

Background: What is it?

Digital Forensics:

“Tools and techniques to recover, preserve, and examine digital evidence on or transmitted by digital devices.”

Digital evidence:

Computers, disks, and cell phones.

But also a **lot** more potential sources…

SMART HUB
zip 100
SONY

Where is Evidence

- Media/disks
  - File system metadata
  - File data
  - Unallocated space (deleted files)
- Network
  - Flow logs, full packet content, passwords
  - Attacker traceback
- Memory
  - Process structures, RAM-only programs, passwords

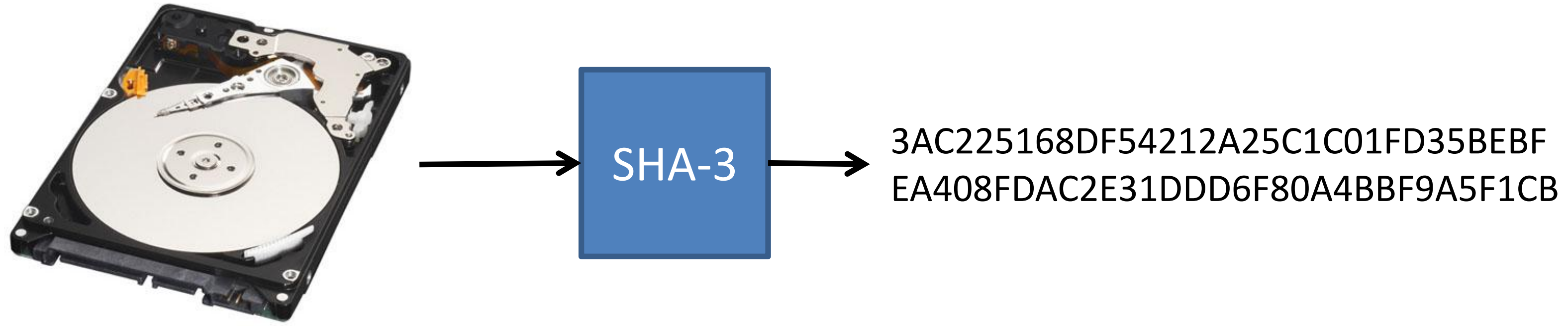
- Files
  - Configuration
  - Logs
  - Media (video, audio)
  - Structured (databases, registry, binary, browser caches)
  - Temp files
  - Swap files

Background: Goals

- Find out details
  - What happened, who did it, when, and why (**intent**)
  - Operate under rules (legal, policy, etc.)
  - How do we fix it, what did *they* do, what did *they* get? (IR)
  - Impartial: looking for answers, not just cherry-picking results
- Preserve information
  - Make sure “nothing changed” (at least as much as possible)
    - Chain of custody
    - Use cryptographically secure hashes to preserve evidence
- Present the results (report, deposition, court testimony, …)
  - Must be factual and clear
  - May be used in a court of law, depending on context

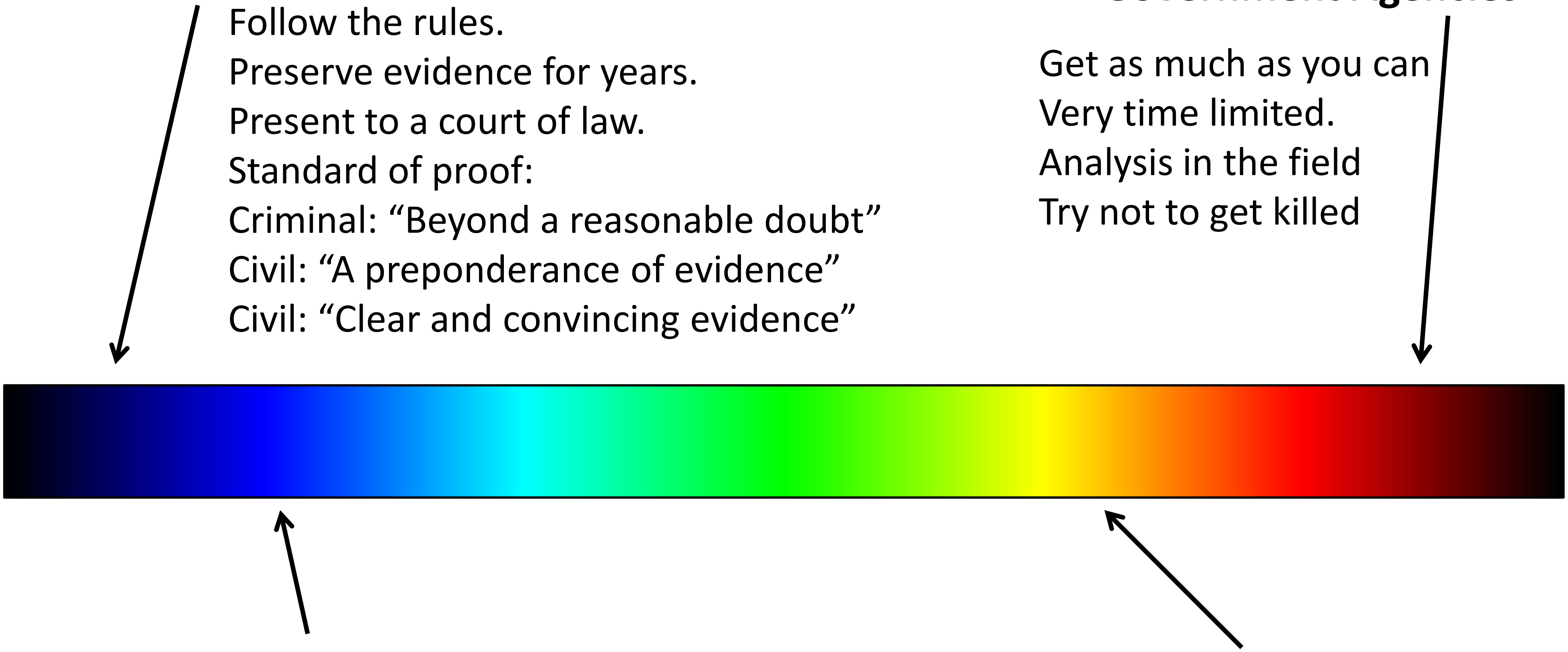
Background: Hashes

- Cryptographically secure hash function
  - Map arbitrary size input into fixed size output (hash)
  - Cannot reproduce original input given output (1-way)
  - Given output, cannot create input that produces output that matches (pre-image resistant)
  - Cannot generate two inputs that produce the same hash (collision resistant)
  - Small change in input (e.g., 1 bit) produces large change in output (e.g., half the bits)
  - Examples: md5, SHA-1, SHA-3
- “Used to preserve evidence”
  - Show that evidence has not changed between time hash was taken and when presented to court (beyond a reasonable doubt



3AC225168DF54212A25C1C01FD35BEBF
EA408FDAC2E31DDD6F80A4BBF9A5F1CB

Background: DF Users and Goals



Law Enforcement

Follow the rules.
Preserve evidence for years.
Present to a court of law.
Standard of proof:
Criminal: “Beyond a reasonable doubt”
Civil: “A preponderance of evidence”
Civil: “Clear and convincing evidence”

Government Agencies

Get as much as you can
Very time limited.
Analysis in the field
Try not to get killed

Archivists

Preserve as much as possible “until the end of the republic.”
“The program’s on a cassette tape?”
“What’s a cassette tape?!?”
“Shhhhhhh!”

Companies (I.R.)

Have a standard process to follow
Fix it NOW! Find out what you can.
Try to preserve what you can

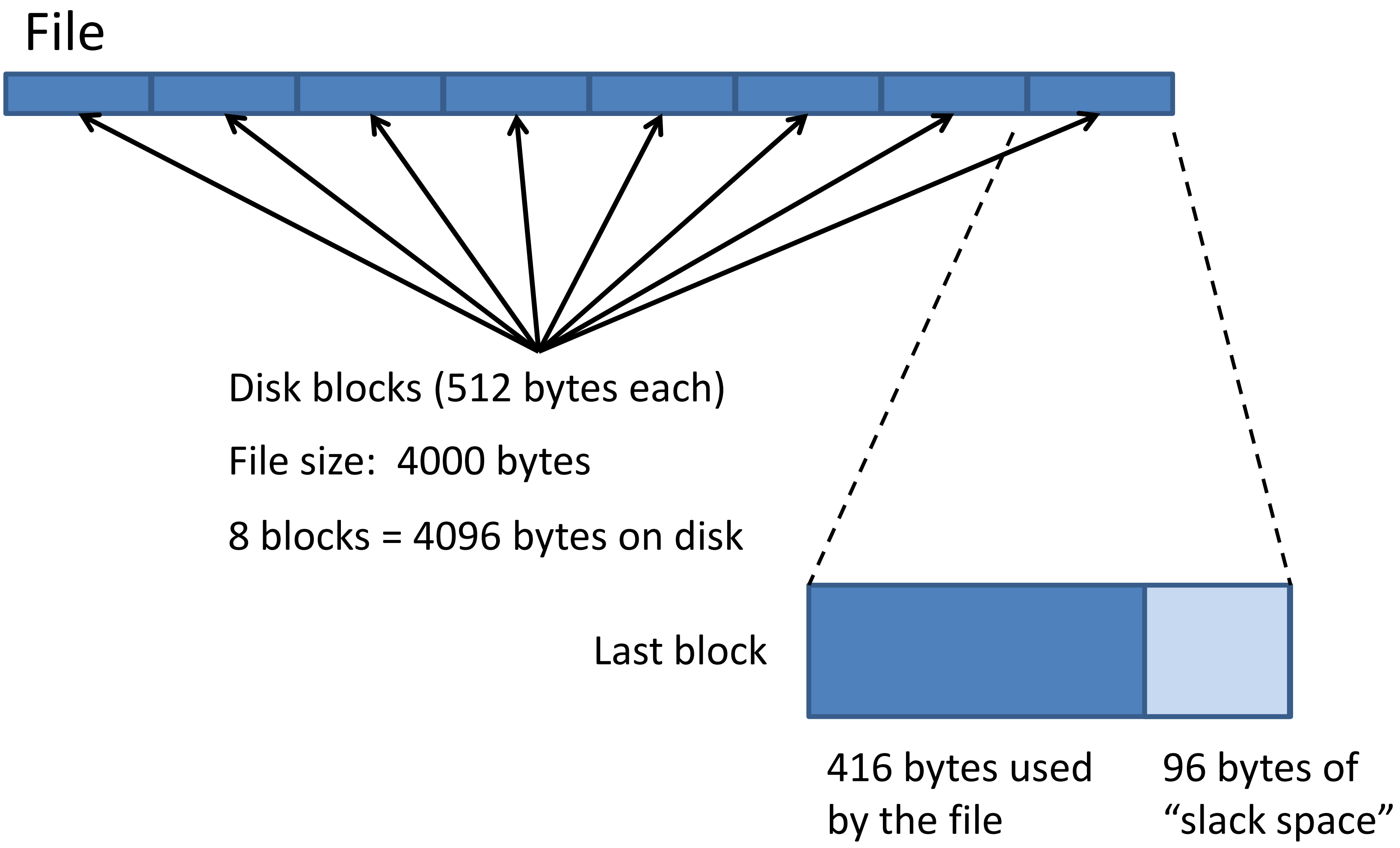
Analysis I: Old School
Examine Everything on the Disk

- Extract files by extension
- Extract files by types
- Extract deleted files (unallocated space)
- Slack space (stegonography)
- Can analyze a disk with a hex editor or even vi

Quick Delve Into…

- Slack space
- Basic file structures
- Deleting (and undeleting files)
- MAC Times and Other Metadata

Slack Space



File

Disk blocks (512 bytes each)

File size: 4000 bytes

8 blocks = 4096 bytes on disk

Last block

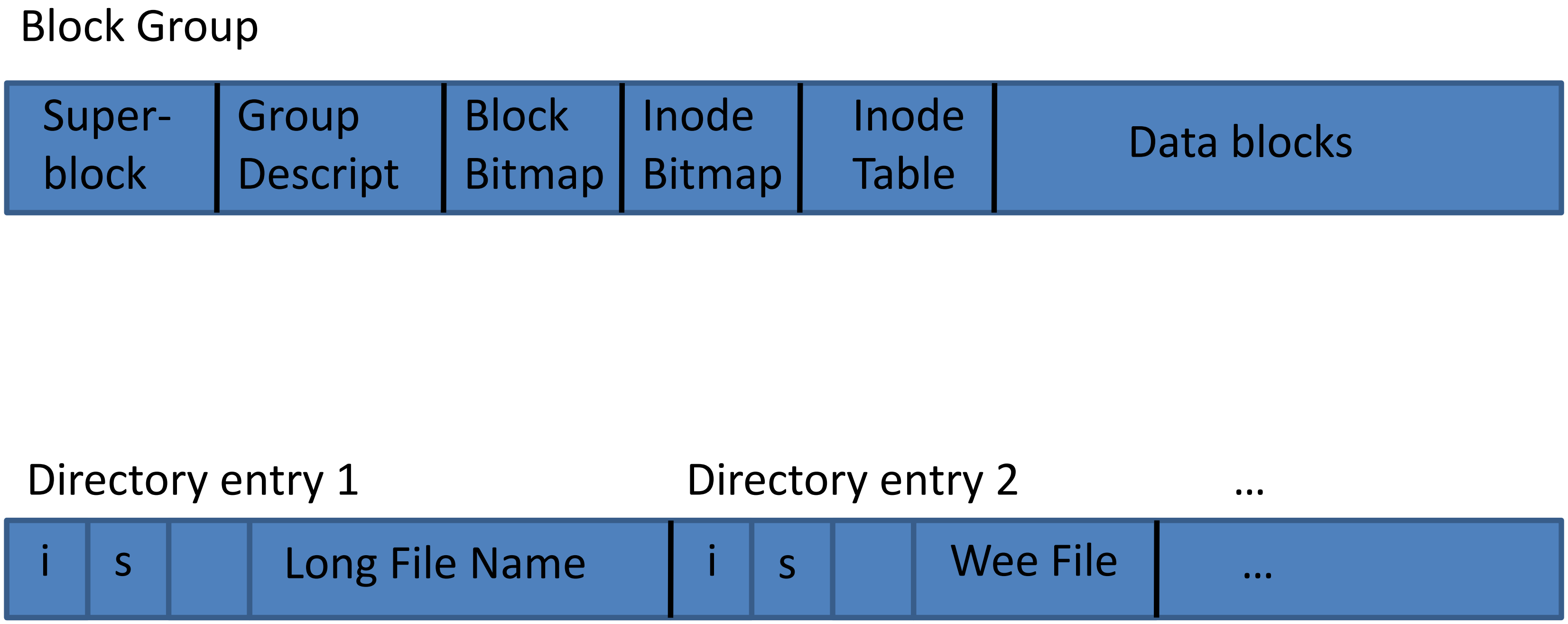
416 bytes used by the file

96 bytes of "slack space"

Slack Space…

- If there was not enough data in the buffer to fill a block then just use what was already in the buffer (typically when it was read)
- Old data in the last block of a file would not be overwritten and could be retrieved
- Space could be used to hide secret data
- Modern OSs zero the buffers before writing, leaving no artifacts in the slack space

Brief File Systems Overview



Block Group

Super- block	Group Descript	Block Bitmap	Inode Bitmap	Inode Table	Data blocks

Directory entry 1 | Directory entry 2 | …

i	s		Long File Name	i	s		Wee File	…

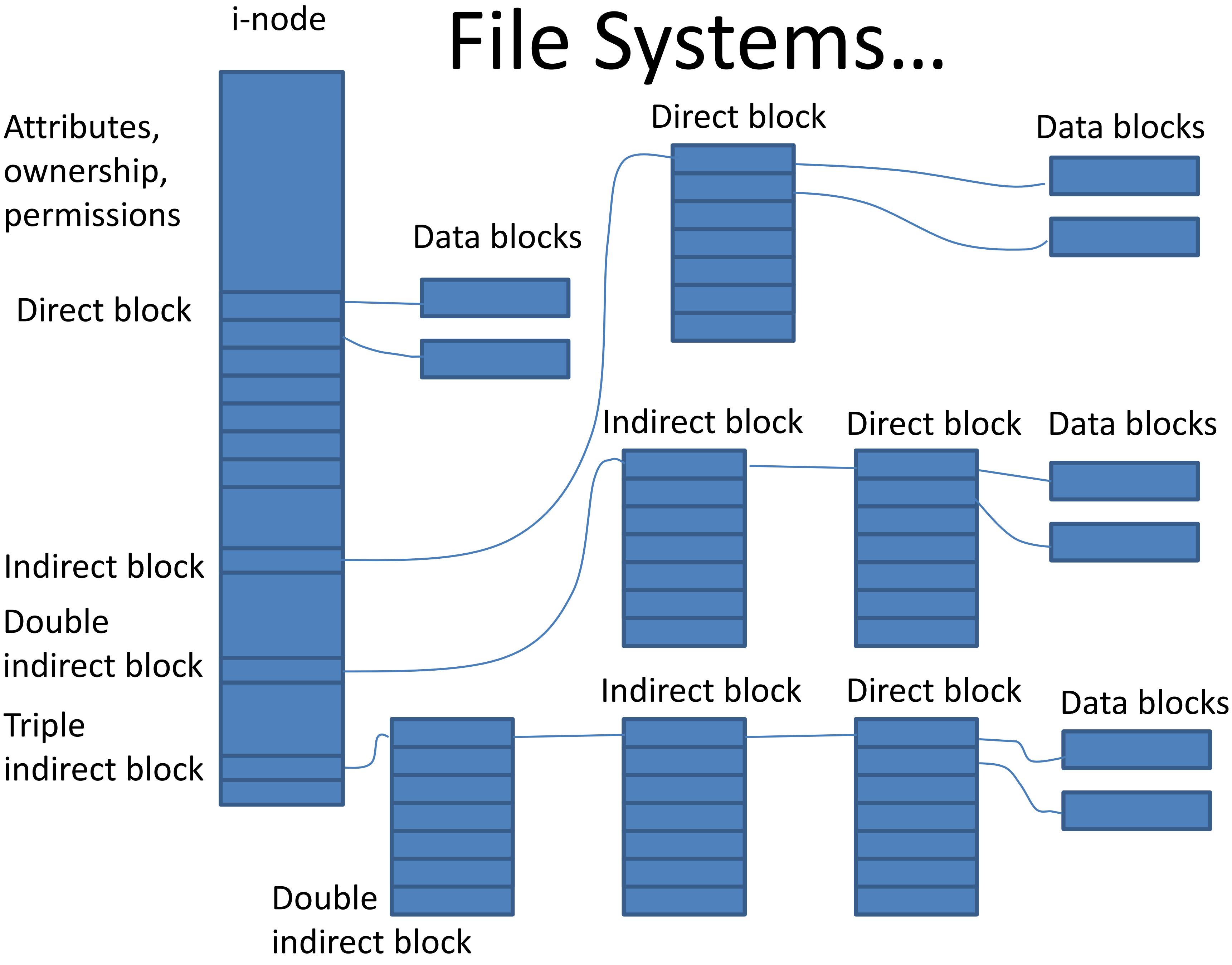
i = inode number

s = file name size

Unix File Systems

- Directories contain files, each entry has
  - the file name
  - pointer to an inode
- Inodes (index nodes) contain information about the files
  - Owner and group
  - Permissions
  - File size
  - Pointer to blocks of data (can be direct, indirect, double indirect, triple indirect)

File Systems…



Recovering Deleted Files: FAT, ext2

- FAT: OS deletes files by setting first byte in directory entry and marking data sectors on disk as free
- ext2: entry removed from directory, inode and data blocks marked as free
- Undelete by reading all directory entries or inodes and getting data blocks from disk (if they haven’t been reused)

File Metadata

- Timestamps on files:
  - Modify, Access, inode Change, Delete, Create (last two optional)
- Time resolution
  - FAT: 10 msec (creation), 2 sec (modify) – 1 day (access)
  - NTFS: updated 1 hour (access), 100 nsec precision
  - Linux: ext3: 1 sec, ext4: 1 nsec
- High precision, not necessarily high accuracy
- Timestamp time zone
  - System clock on local time or UTC?
  - Is clock correct (or synchronizes with a source)?
  - Where is investigator?
  - Is or was it daylight savings time but not anymore?
  - Synchronizing with multiple computers?

File Metadata…

- When did an event occur
  - Editing a file
  - Copying files
  - Sitting in front of the computer
  - Downloading pictures
- Permissions
  - Who else could get at or alter the data?
  - Is this a potential possession or distribution crime?

Application Metadata

- In addition to when and who, it may provide provenance data
  - Where did that document come from, where did that *phrase* you cut and pasted in Word come from
- Examples
  - Browser history
  - Document history
  - Email (to, from, subject, time, message-id, IP address)
  - Printing

2000s Era: Automation

- Simple keyword searches
- Search sophistication: less than grep
- Parse file systems without using the OS
  - The Coroner’s ToolKit (TCT) and later The Sleuth Kit (TSK), plus commercial tools
- Find files, deleted files
- Build primitive time lines
- Still manually intensive

Modern File Systems

- Basic file structures
  - Similar but a bit more complex
- Deleting (and undeleting files)
  - NTFS deletes entry from dir but leaves MFT entry (inode) intact with pointer back to parent, can reconstruct file path, name, and contents.
  - Ext3 wipes inodes, harder to undelete without carving
- MAC Times and Other Metadata
  - atime generally no longer set
- Privacy and OS efficiency modifications can make some information more difficult to get

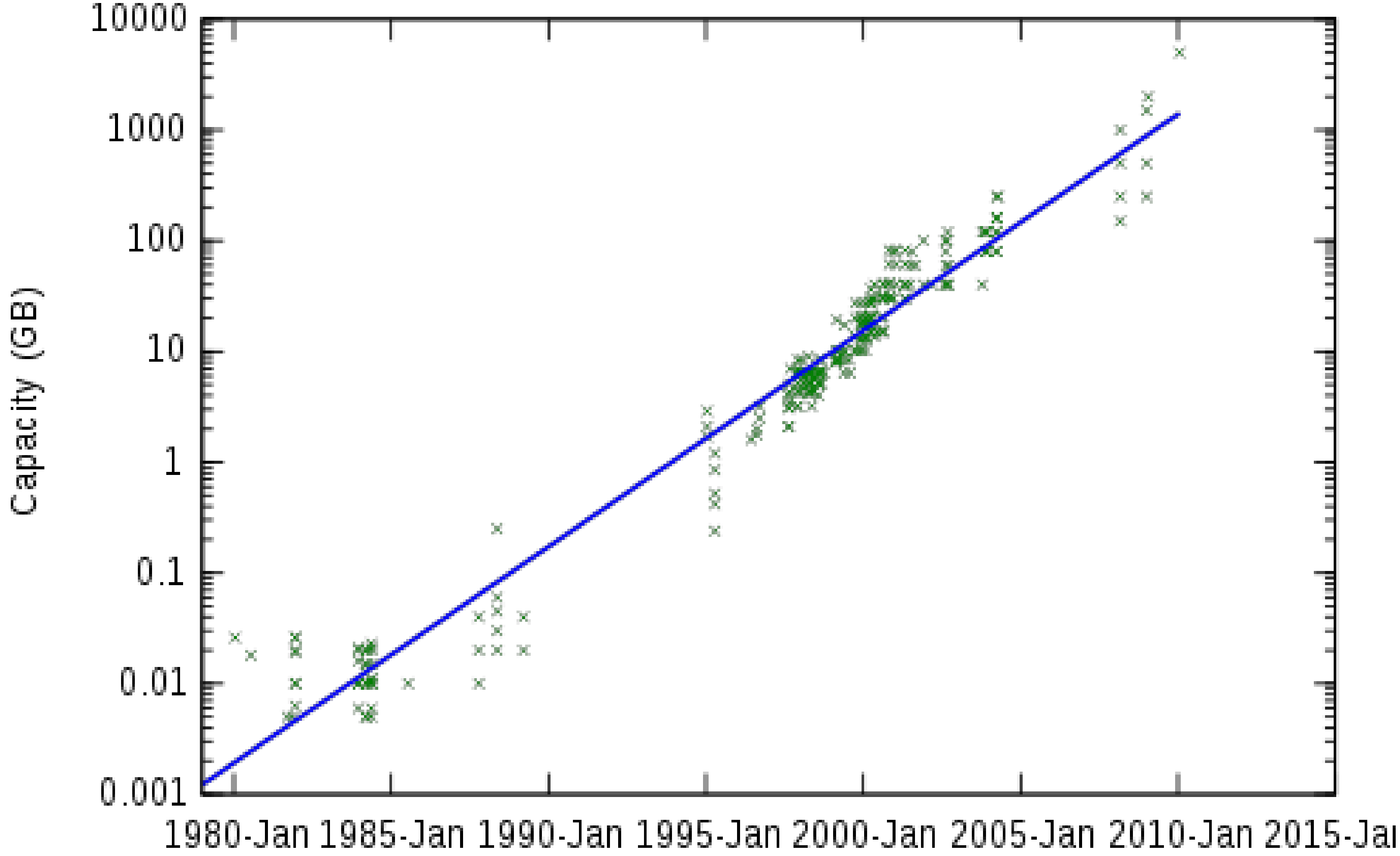
File Carving

- Deleted files still exist as data runs of blocks on the disk in unallocated space.
- Challenge: Can these blocks be reassembled?
- Answer: **File Carving** (e.g., Scalpel)
- Carvers have deep knowledge of various file formats (jpeg, gif, png, mp3, pdf, etc.)
- Look for start and end markers
- Look for ways to link the end of one span to the beginning of another one

Modern Era Problems

- Disk capacity (base system is >=1TB)
  - 3TB @ $100
- SSDs
  - Much faster than mechanical disks (yay)
  - *Wear leveling* creates many copies of data (yay)
  - *Trim* command permanently deletes data fast (boo)
- Encryption (file, whole disk, and network)
  - Without a password, it’s all opaque
  - But, passwords are often resident in memory

Disks Keep Getting Bigger



Source: Wikipedia.org

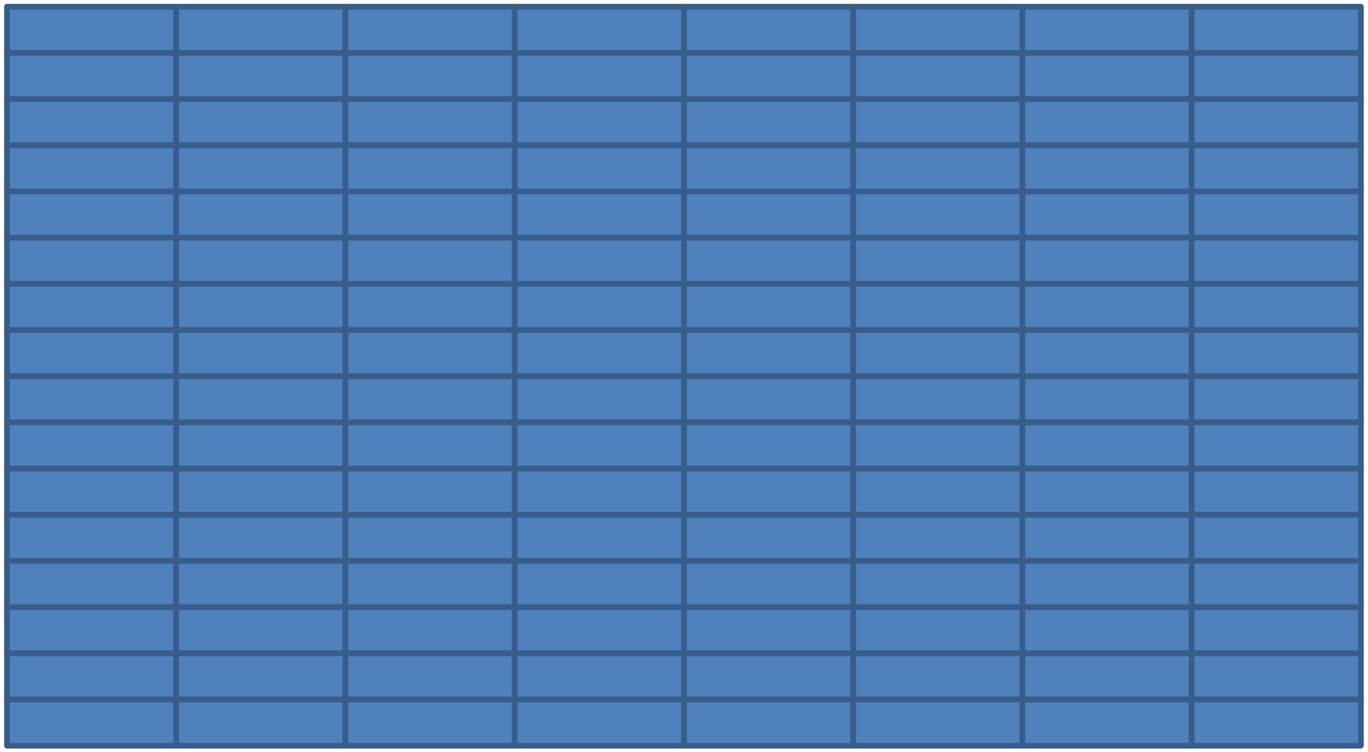
Big Data, Big Problems

- Imaging a 1TB disk can take a few hours (depending on too many factors)
- Indexing the data can take a long time
  - 128 Gigabytes of memory is still way less than 1TB
- Distributed computing, GPUs, and storage (many spindles) can help, but …
- Most forensic investigators (short of large government agencies) don’t have those kind of resources to allocate *per case*.

Solid State Drives

SSDs: Pages and Blocks

Page: 4KB or 16KB

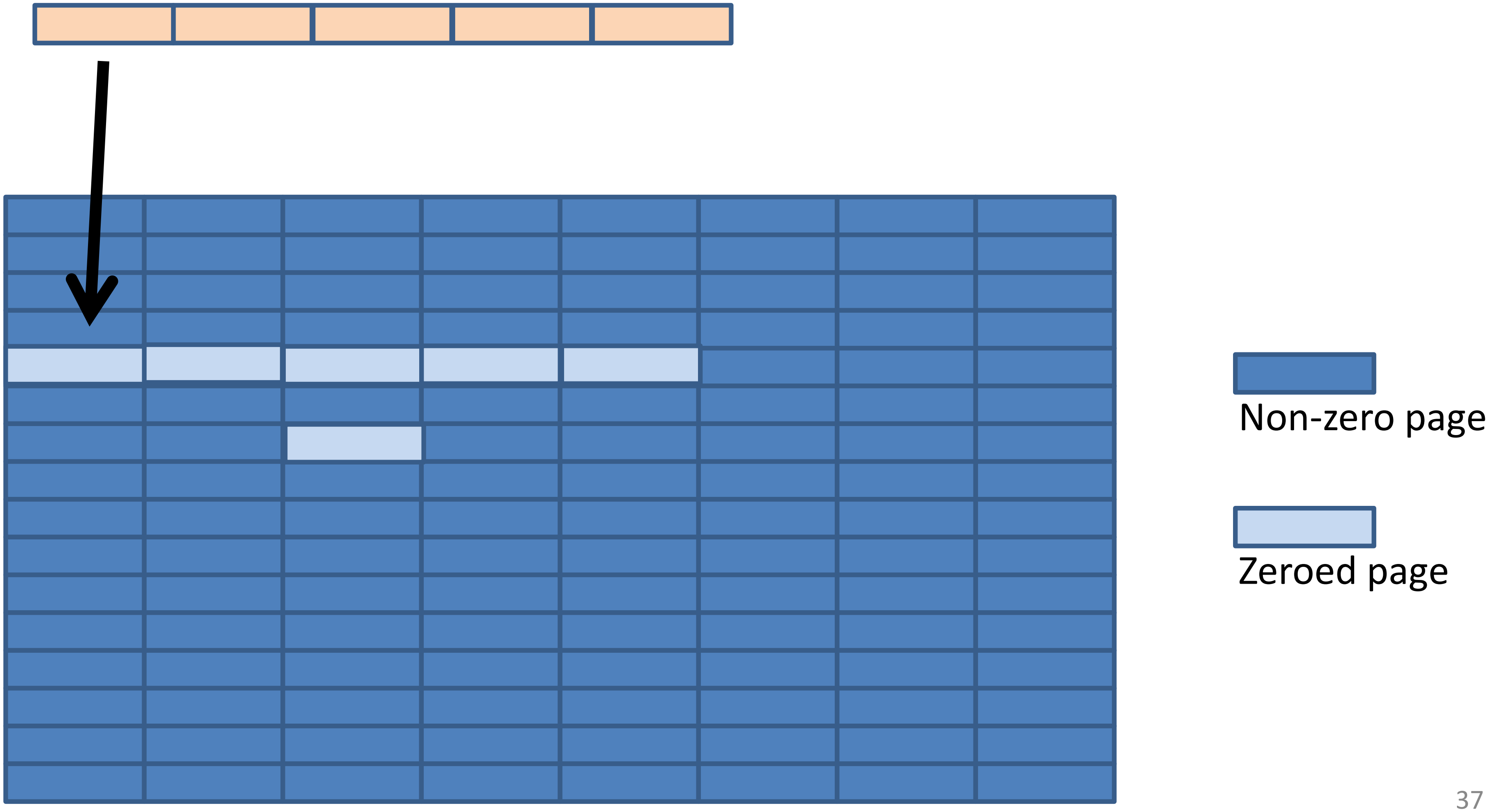


Block: 128 - 512 pages (128 4KB blocks in this example, 512KB total)

A 16GB disk might have 3200 blocks (not shown).

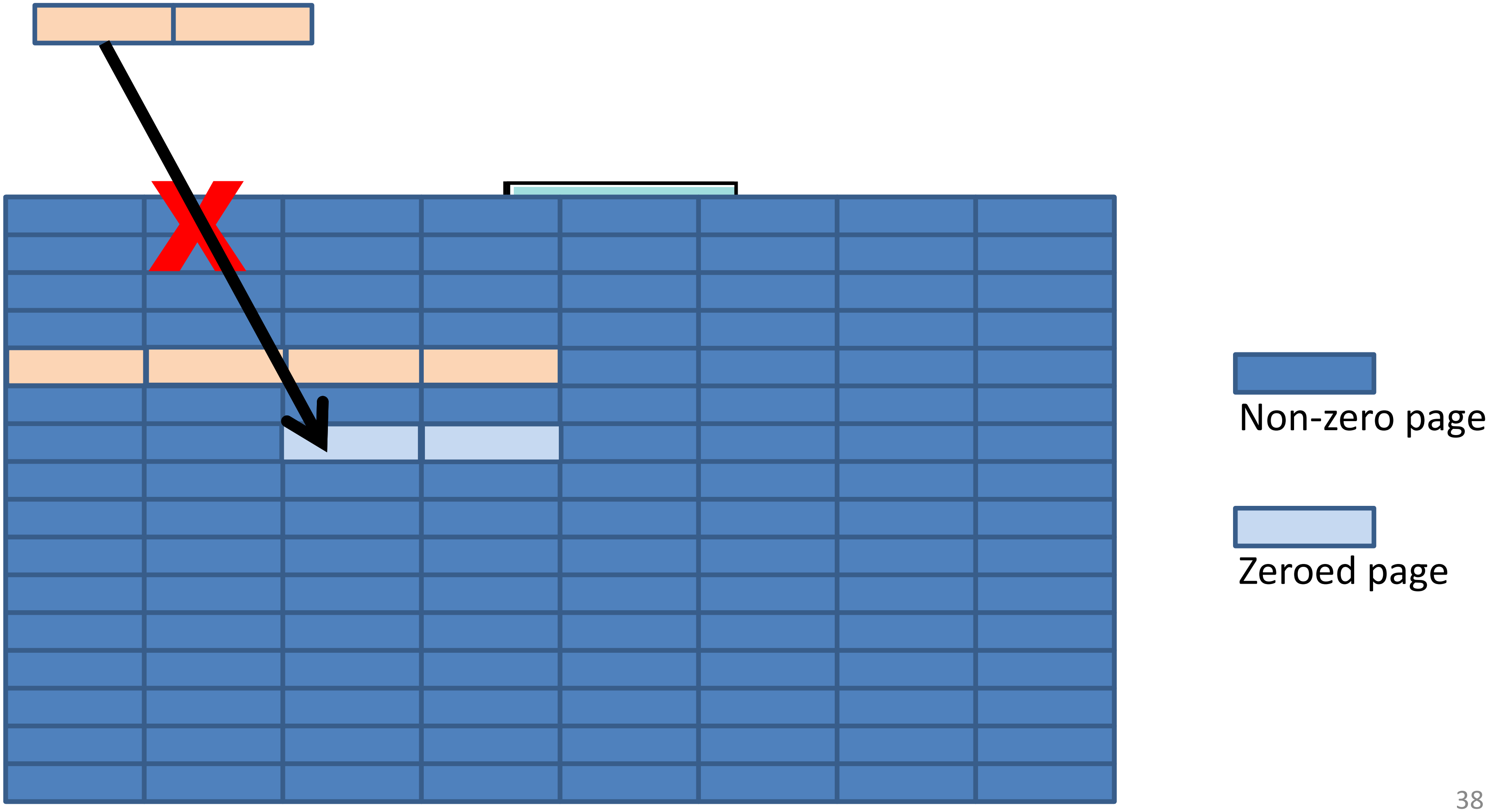
SSDs: Writing

SSD can write a page at a time, if it's an empty page.



SSDs: Erasing

SSD can only erase an entire block at a time, which is slow.



SSD Trim Command

- Previously, an OS updates its own free block bitmap when it deletes files (doesn’t tell the disk)
- The disk doesn’t know if pages are free, so it must preserve all pages.
- When the SSD erases a block, it must copy all pages in block which hurts performance. (SSDs have more space than they advertise to support this.)
- Now, the OS sends **Trim** command to SSD to indicate a block is no longer in use. SSD will **not** preserve (copy) trimmed blocks.

Trim Command…

- Blocks have a limited lifetime (100K or 1M writes)
- Blocks are constantly being erased in an SSD.
  - *Wear leveling* used for writes to prevent a block from being written too many times (dynamic)
  - Or *too few times* (static) to ensure all blocks are used about the same amount.
- SSDs perform wear leveling *anytime* when they have power, i.e., whenever plugged into a USB even if OS isn’t running.
- Trimmed blocks have a lifespan typically measured in seconds and *cannot* be recovered
- Limits effectiveness of file carving
- Good for privacy, bad for forensics

Analysis II: The Next Generation

- Tools provide a deep understanding of specific data formats
- Tools help with:
  - File carving, web history, timeline analysis, *peer-to-peer/file sharing* analysis, email (text messages for phones), *registry*, and more
- Live foreniscs/memory interpretation provides insight into
  - Running and terminated processes
  - Kernel and application data structures
  - Memory resident programs (malicious)

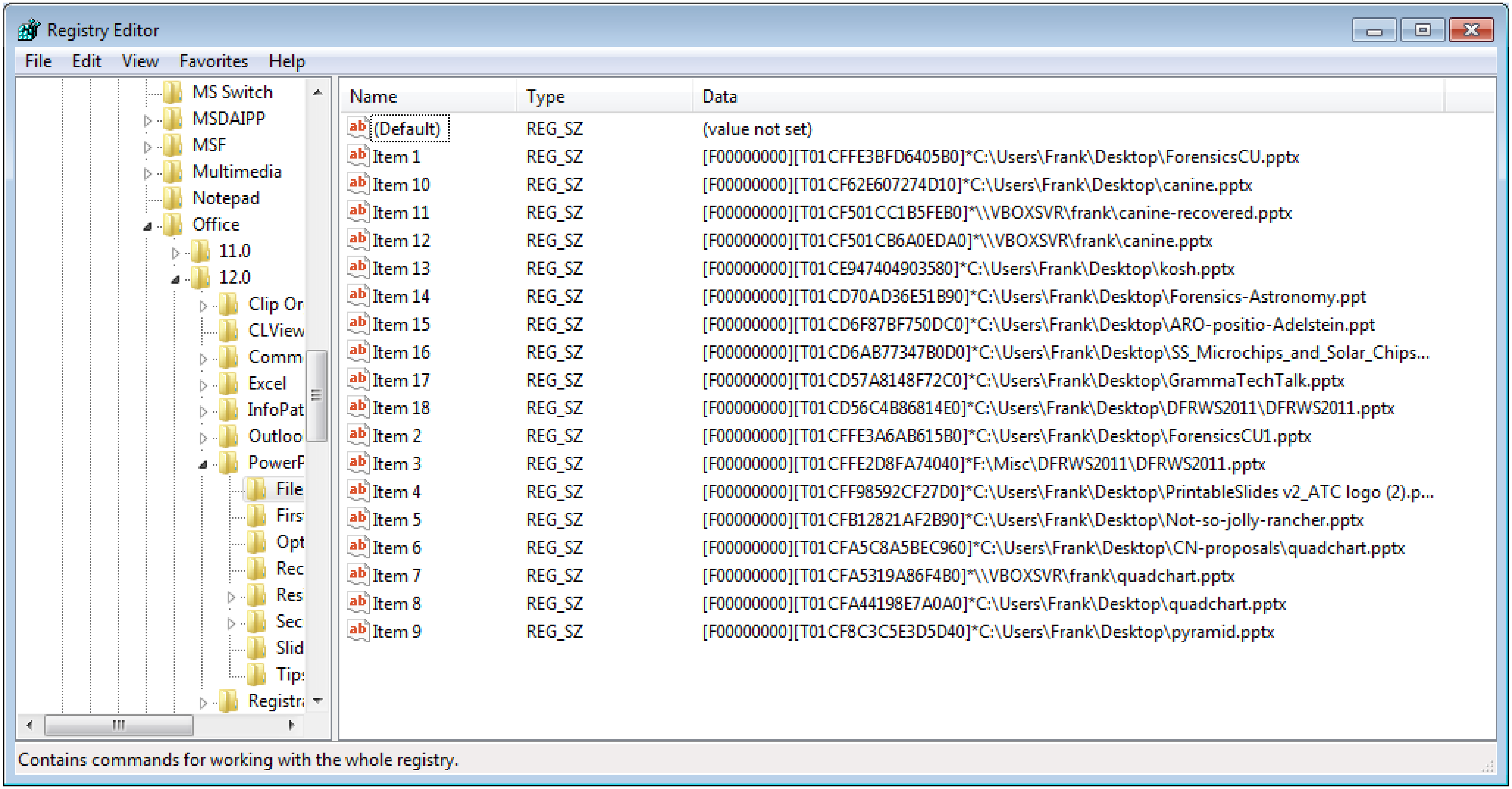
Registry

- Tree-like key/value store for Windows, represented as a linked list on disk
  - Hives (files)
  - Keys (path in a tree)
  - Value (path to leaf)
  - Data type (string, binary, 2 byte, 4 byte, multi-string (array), etc.)
  - Data (value of leaf)
- Programs store configuration and status data
- For forensic investigators, registry can either be…

Registry

- Recently used files
- Typed URLs
- Installed programs
- Previously installed programs
- Programs automatically run at log on
- Devices that had been plugged in
- Configuration settings, data locations

Recently Used Files

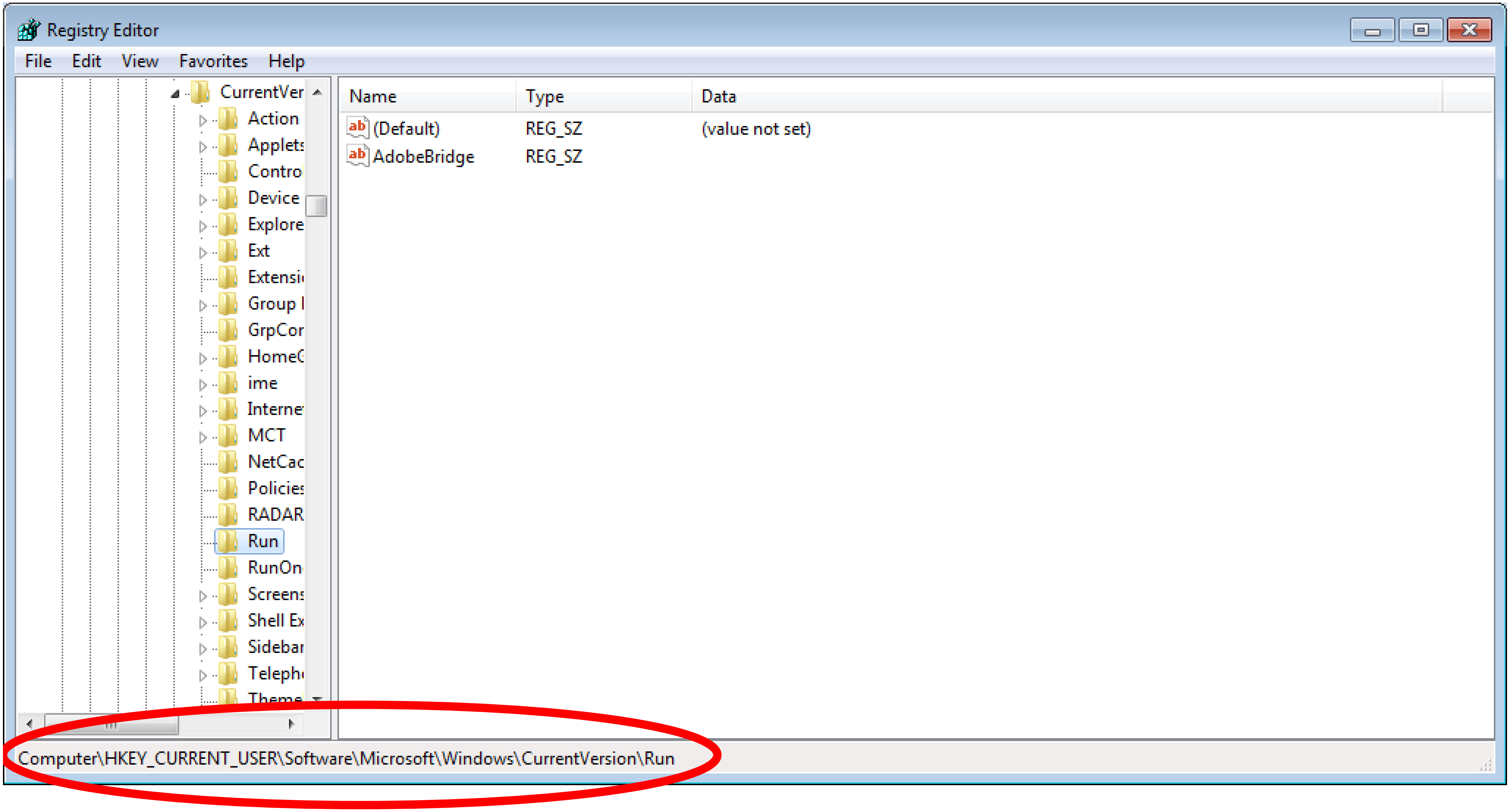


Name	Type	Data
(Default)	REG_SZ	(value not set)
Item 1	REG_SZ	[F00000000][T01CFFE3BFD6405B0]*C:\Users\Frank\Desktop\ForensicsCU.pptx
Item 10	REG_SZ	[F00000000][T01CF62E607274D10]*C:\Users\Frank\Desktop\canine.pptx
Item 11	REG_SZ	[F00000000][T01CF501CC1B5FEB0]*\\VBOXSVR\frank\canine-recovered.pptx
Item 12	REG_SZ	[F00000000][T01CF501CB6A0EDA0]*\\VBOXSVR\frank\canine.pptx
Item 13	REG_SZ	[F00000000][T01CE947404903580]*C:\Users\Frank\Desktop\kosh.pptx
Item 14	REG_SZ	[F00000000][T01CD70AD36E51B90]*C:\Users\Frank\Desktop\Forensics-Astronomy.ppt
Item 15	REG_SZ	[F00000000][T01CD6F87BF750DC0]*C:\Users\Frank\Desktop\ARO-positio-Adelstein.ppt
Item 16	REG_SZ	[F00000000][T01CD6AB77347B0D0]*C:\Users\Frank\Desktop\SS_Microchips_and_Solar_Chips...
Item 17	REG_SZ	[F00000000][T01CD57A8148F72C0]*C:\Users\Frank\Desktop\GrammaTechTalk.pptx
Item 18	REG_SZ	[F00000000][T01CD56C4B86814E0]*C:\Users\Frank\Desktop\DFRWS2011\DFRWS2011.pptx
Item 2	REG_SZ	[F00000000][T01CFFE3A6AB615B0]*C:\Users\Frank\Desktop\ForensicsCU1.pptx
Item 3	REG_SZ	[F00000000][T01CFFE2D8FA74040]*F:\Misc\DFRWS2011\DFRWS2011.pptx
Item 4	REG_SZ	[F00000000][T01CFF98592CF27D0]*C:\Users\Frank\Desktop\PrintableSlides v2_ATC logo (2).p...
Item 5	REG_SZ	[F00000000][T01CFB12821AF2B90]*C:\Users\Frank\Desktop\Not-so-jolly-rancher.pptx
Item 6	REG_SZ	[F00000000][T01CFA5C8A5BEC960]*C:\Users\Frank\Desktop\CN-proposals\quadchart.pptx
Item 7	REG_SZ	[F00000000][T01CFA5319A86F4B0]*\\VBOXSVR\frank\quadchart.pptx
Item 8	REG_SZ	[F00000000][T01CFA44198E7A0A0]*C:\Users\Frank\Desktop\quadchart.pptx
Item 9	REG_SZ	[F00000000][T01CF8C3C5E3D5D40]*C:\Users\Frank\Desktop\pyramid.pptx

Contains commands for working with the whole registry.

HKEY_CURRENT_USER\Software\Microsoft\Office\12.0\PowerPoint\File MRU

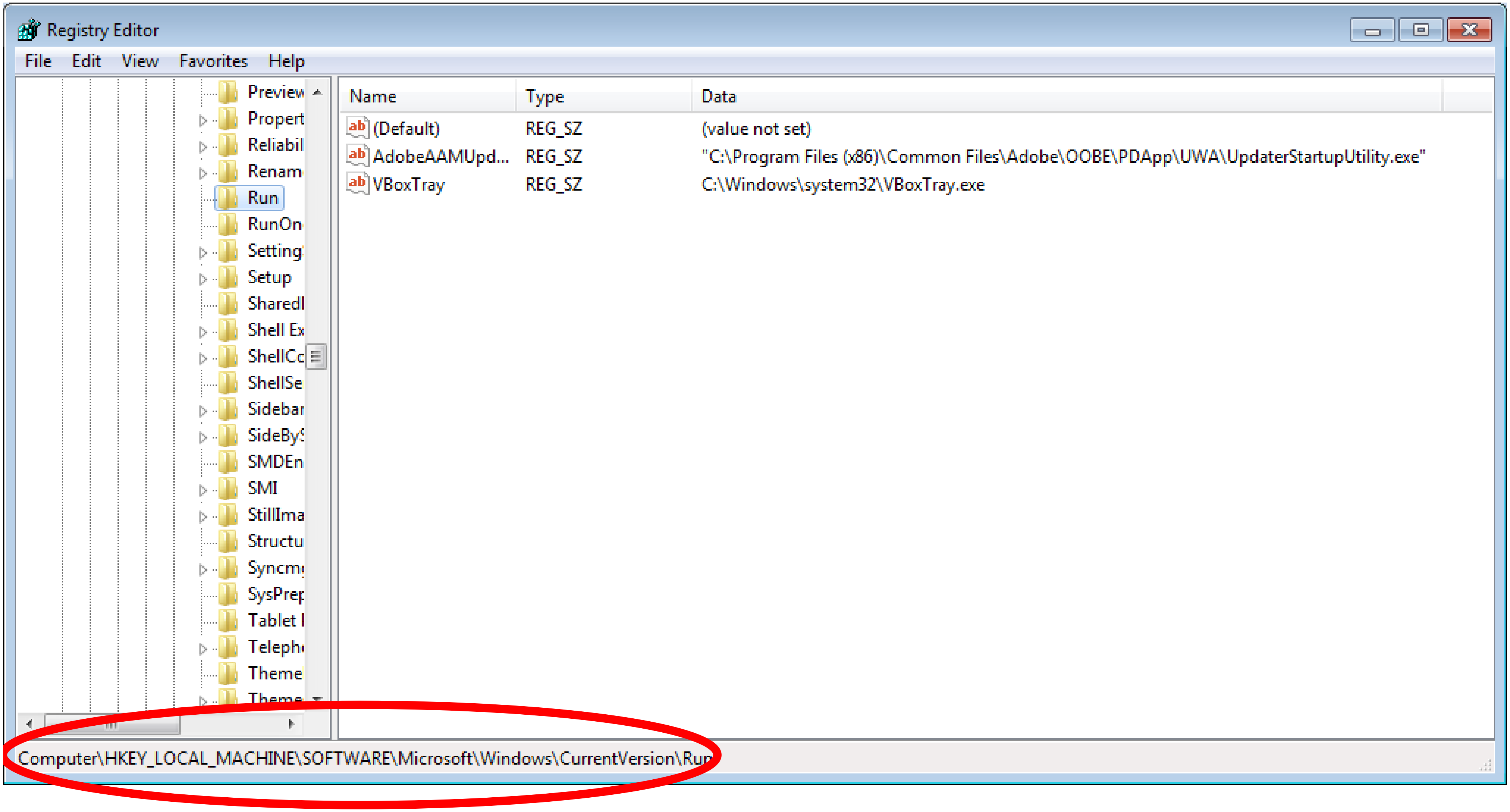
Runs When A Specific User Logs In



Name	Type	Data
(Default)	REG_SZ	(value not set)
AdobeBridge	REG_SZ	

Computer\HKEY_CURRENT_USER\Software\Microsoft\Windows\CurrentVersion\Run

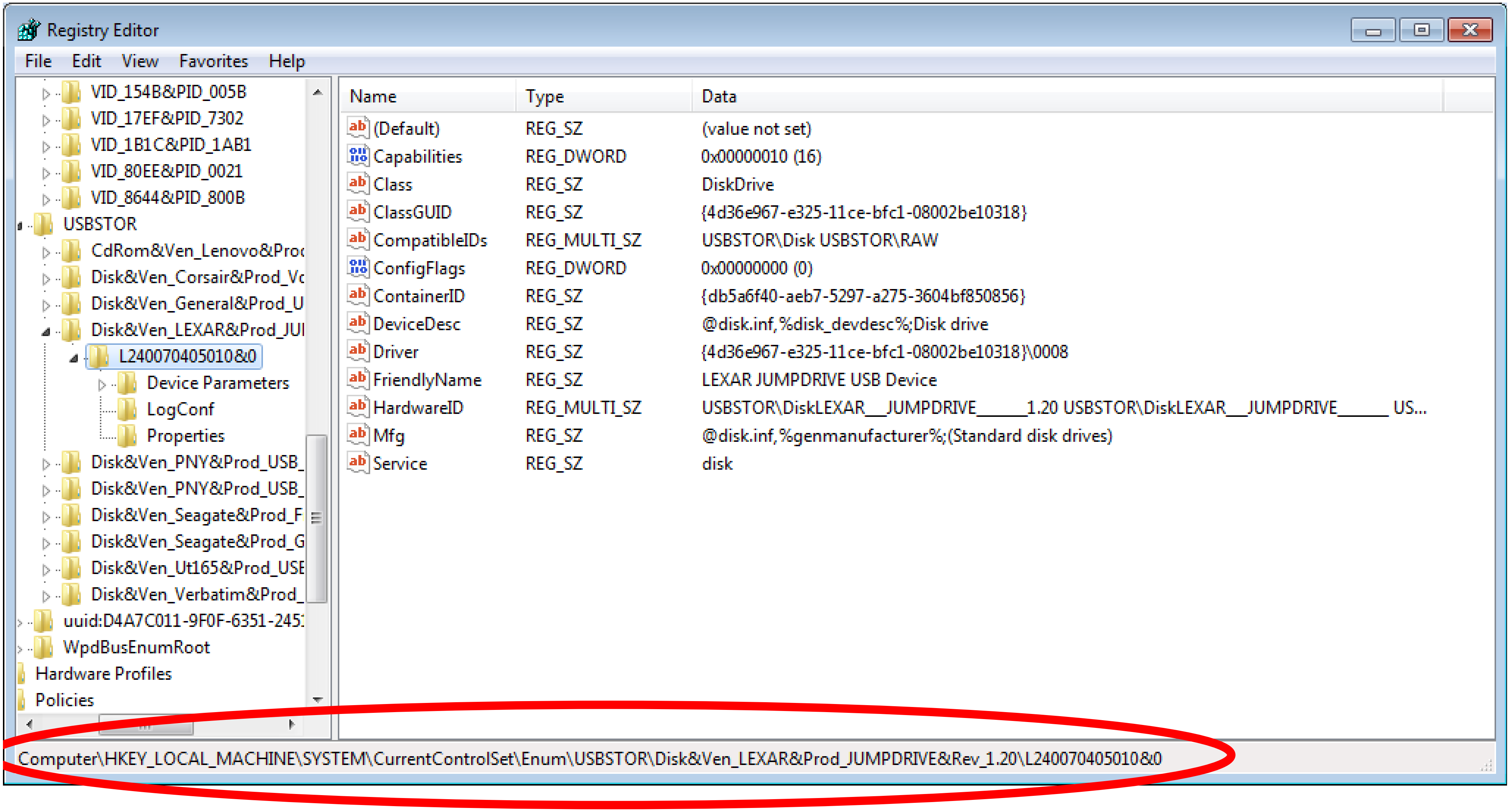
Runs When Any User Logs In



Name	Type	Data
(Default)	REG_SZ	(value not set)
AdobeAAMUpd...	REG_SZ	"C:\Program Files (x86)\Common Files\Adobe\OOBE\PDApp\UWA\UpdaterStartupUtility.exe"
VBoxTray	REG_SZ	C:\Windows\system32\VBoxTray.exe

Computer\HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows\CurrentVersion\Run

USB Devices the Computer Has Seen

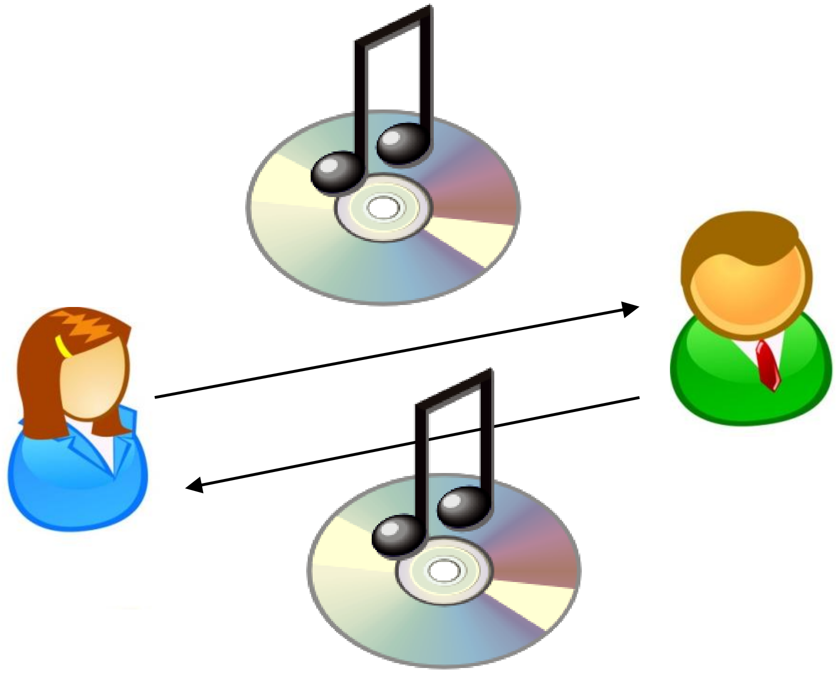


```
  VID_154B&PID_005B
  VID_17EF&PID_7302
  VID_1B1C&PID_1AB1
  VID_80EE&PID_0021
  VID_8644&PID_800B
USBSTOR
  CdRom&Ven_Lenovo&Proc
  Disk&Ven_Corsair&Prod_Vc
  Disk&Ven_General&Prod_U
  Disk&Ven_LEXAR&Prod_JUI
    L240070405010&0
      Device Parameters
      LogConf
      Properties
  Disk&Ven_PNY&Prod_USB_
  Disk&Ven_PNY&Prod_USB_
  Disk&Ven_Seagate&Prod_F
  Disk&Ven_Seagate&Prod_G
  Disk&Ven_Ut165&Prod_USE
  Disk&Ven_Verbatim&Prod_
uuid:D4A7C011-9F0F-6351-245:
WpdBusEnumRoot
Hardware Profiles
Policies
```

Name	Type	Data
(Default)	REG_SZ	(value not set)
Capabilities	REG_DWORD	0x00000010 (16)
Class	REG_SZ	DiskDrive
ClassGUID	REG_SZ	{4d36e967-e325-11ce-bfc1-08002be10318}
CompatibleIDs	REG_MULTI_SZ	USBSTOR\Disk USBSTOR\RAW
ConfigFlags	REG_DWORD	0x00000000 (0)
ContainerID	REG_SZ	{db5a6f40-aeb7-5297-a275-3604bf850856}
DeviceDesc	REG_SZ	@disk.inf,%disk_devdesc%;Disk drive
Driver	REG_SZ	{4d36e967-e325-11ce-bfc1-08002be10318}\0008
FriendlyName	REG_SZ	LEXAR JUMPDRIVE USB Device
HardwareID	REG_MULTI_SZ	USBSTOR\DiskLEXAR\__JUMPDRIVE______1.20 USBSTOR\DiskLEXAR\__JUMPDRIVE______ US...
Mfg	REG_SZ	@disk.inf,%genmanufacturer%;(Standard disk drives)
Service	REG_SZ	disk

Computer\HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Enum\USBSTOR\Disk&Ven_LEXAR&Prod_JUMPDRIVE&Rev_1.20\L240070405010&0

Peer-to-Peer



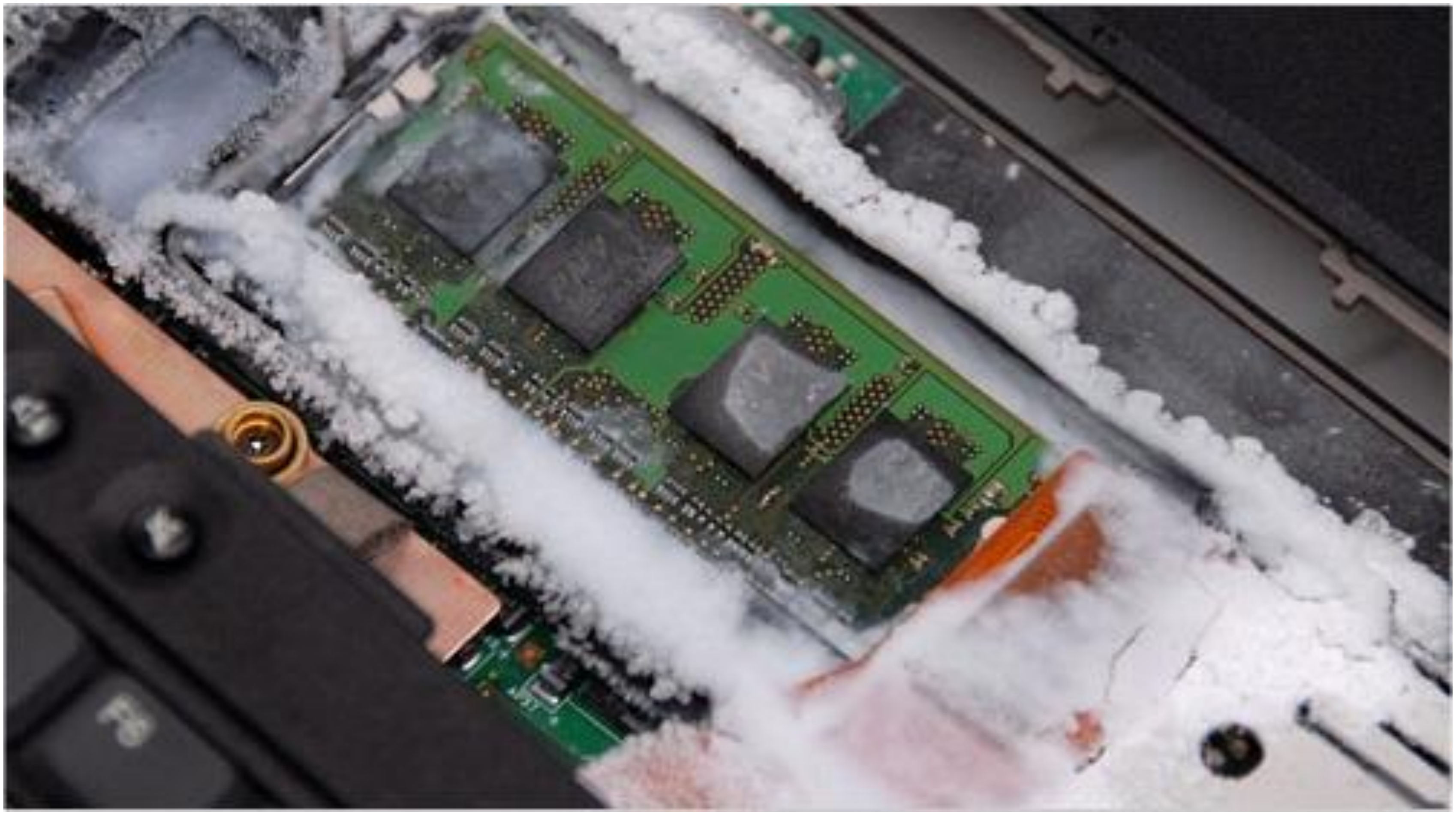
Current Trends: Peer-to-Peer

- Many clients (program) and servers (protocols)
- Each program stores files in different locations, has different defaults, and different configuration file
- Investigators need to determine
  - what programs are present
  - where the shared files are
  - where the metadata and logs are
  - what data has been shared/distributed
- Then they must get it all in a forensically sound way

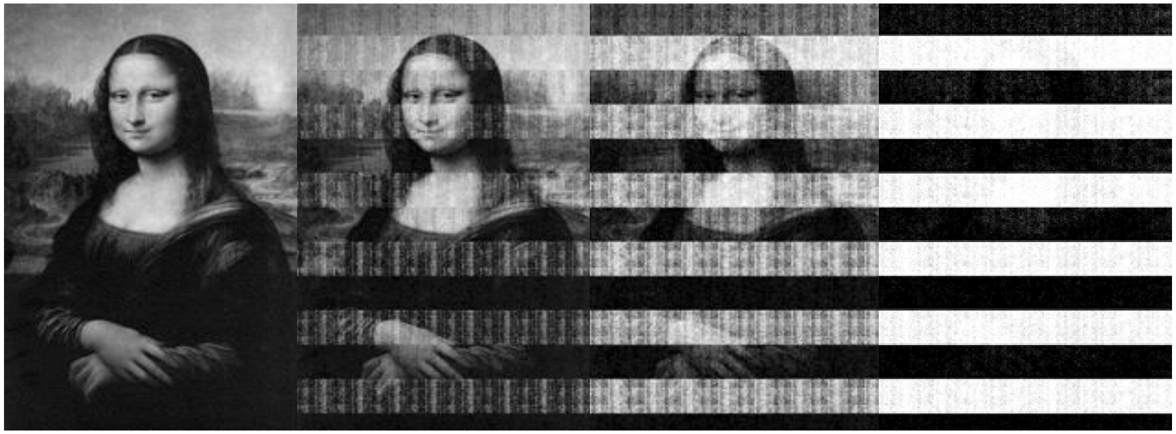
Trends: Live Forensics

- Get context of what’s happening on system NOW
  - Running processes
  - Active network connections
  - Memory dump (memory-only programs, encryption keys, etc.)
- Can pre-deploy agent (plan to get hacked)
- Push an agent on the machine (given credentials and contaminating/overwrite some evidence)
- Reboot into a forensic DVD-only OS (for some machines rebooting erases all memory)
- Or, break out a can of compressed air, and…

Data Remanence



Remanence at Room Temp



5 secs 30 secs 60 secs 5 mins (gone)

Canned air: MINUTES
Liquid Nitrogen: AN HOUR

Source: http://citp.princeton.edu/memory

-50°C Brrrrrrrrrrrrr!

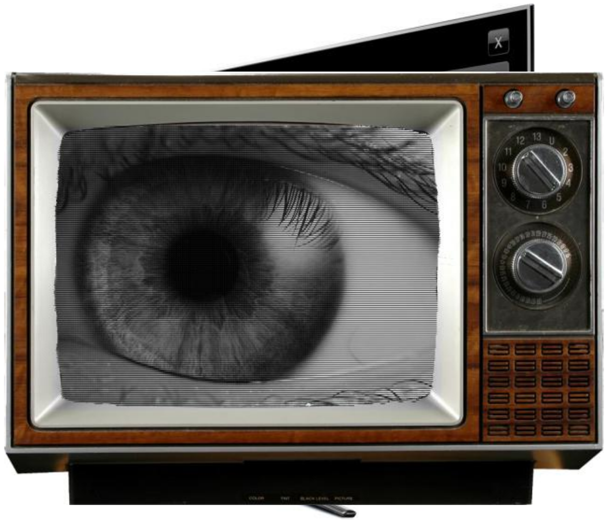
Bye, bye Bitlocker.

Trends

- Memory analysis
  - Best Practices 10 years ago was “pull the plug”
  - DFRWS 2005 Memory Challenge
    - Provided scenario, memory dumps, kernel, network capture, etc., and 7 questions to answer
    - Motivated a lot of work in the area of memory forensics
  - Identify “key schedules” in memory, reduce brute force search space and crack AES fast! (ref DFRWS09)
  - New tools when given a memory dump can reconstruct processes, file descriptors, application data structures, and more (Volatility, GRR)

Trends

- Link Analysis – graph who talks to whom
- New devices, new evidence sources
  - – Cell phones, tablets, watches
  - – Smart TVs store a **LOT** of information
    “In Soviet Russia, TV watches YOU”
- Reverse Engineering
  - – Understand what’s going on
  - – Defeat hiding mechanisms



Trends: Reverse Engineering for…

- Analysis of large installation break-ins
  - Break-ins at Target, Home Depot, JP Morgan, Chase, Walmart, (IRS?) just this year!
- “Advanced Persistent Threat”
  - Deeply compromised systems
  - Quietly compromised
- Bot-nets (more sophisticated C&C)
- Anti-forensics (detection)
- Line between IR and LE practitioners are fuzzy at best

Archival Forensics

- Donated collections, retiring professors, …
- OLD systems (‘90s era and beyond)
- Privacy is a huge issue—what can be released to the public?
  - Email
  - Contacts
  - Drafts of famous papers and books
- Is it “real” forensics? Yes!

Archival Forensics

- Goals:
  - Preserve information
  - Find out what happened, how it worked, catalog data and media, *intent*
- Methods
  - Use standard tools (Encase, FTK, The Sleuth Kit, …)
  - Create new tools as needed
- Constraints
  - Don’t change sources if possible (archival principles)
  - Limited information/cooperation from owner
- Main differences
  - Not presenting before a court of law
  - **Goal is to provide public access to the material**

Archival Principles

- Provenance/Authenticity
- Respect des fonds
  - Keep records emanating from the same source together
  - Group without mixing them with others
- Preserve order
- Don’t organize things in a way that distorts the original context
- Access
  - Least restrictive access, reasonable redaction
  - Donor agreement
- **Don’t speak for them**

Summary

- Forensics benefits from information leaks from applications, OS, and more
- Forensics and security are at odds (privacy too)
- Forensics benefits from software written without regard to privacy where deep state inspection was never considered
- Correlation of multiple data sources can provide a much more detailed picture of events
- Generally most sources were not intended to be used for forensics
- Often these sources change or are removed
- Privacy mechanisms can foil some analysis
- Analysis requires deep knowledge and meticulous detail

Challenges and Research Problems

- Handling big data (need to move beyond floppy disk mentality)
  - Using partial images/samples (legal and technical)
- Integrating more sources of evidence
- Automatic event reconstruction
- Privacy concerns
- Anti-forensics (tools and techniques to defeat forensic analyses)
- Reverse engineering
- Different groups have different goals, requirements, and restrictions
  - “If it’s good enough for law enforcement…” not true for everyone
- Automation while avoiding “black box” syndrome (“because the program said so” is not a valid response to “why did you claim my client is responsible?”)
- Little research money and limited commercial market

Thank you

Questions, comments now,
or later to: frank@notfrank.com

A Few References/Sources

Conference

- Digital Forensics Research Workshop, www.DFRWS.org (all papers and presentations)

Journal

- Digital Investigation
  - http://www.journals.elsevier.com/digital-investigation/

"Cool" Research

- *Lest We Remember: Cold Boot Attacks on Encryption* Keys, J. Alex Halderman, et al., 17th USENIX Security Symposium (Sec '08), San Jose, CA, July 2008. http://citp.princeton.edu/pub/coldboot.pdf

Fun book

- The Cuckoo's Egg by Cliff Stoll
  - Network forensics and good storytelling, 1989

Tools

- Scalpel https://github.com/sleuthkit/scalpel
- The Sleuth Kit (http://www.sleuthkit.org/)
- Volatility (memory forensics)
  - https://code.google.com/p/volatility/
- GRR, I.R. for remote live forensics
  - https://github.com/google/grr