Challenges to Digital Forensic Evidence

Fred Cohen
CEO - Fred Cohen & Associates
Research Professor - The University of New Haven
Chairman – SecurityPosture
and my day job...

fred.cohen at all.net
http://all.net/

Outline

- Background
  - Fred Cohen – Digital Forensics
- Basic notions – processes and challenges
- Evidence identification and collection
- Transportation and storage
- Analysis, interpretation, and reconstruction
- Presentation
- Summary and conclusions

Fred Cohen

- Fred Cohen & Associates
  - Almost 30 years of information protection leadership
  - Many complex research projects and subjects
  - Consulting for many leading organizations
    - Government and Academic research and advise
    - Corporate consulting at the highest levels
    - Digital forensics, private and LE investigations
    - Strategic intelligence and critical infrastructures
  - Research and Development
    - Evaluating technologies and lines of effort
    - Patented technologies and innovations

Forensics background

- Ph.D in electrical engineering - computer-centric
- ForensiX – Linux-based forensics tool
- Teach graduate forensics courses at UNH
- California SEARCH instructor
- California POST certified
- Research in forensics at Sandia National Labs
- Work on corporate civil cases here and there
- Testimony in Federal and State cases
- I tend to get the unusual cases

Outline

- Background
- Basic notions – processes and challenges
- Evidence identification and collection
- Transportation and storage
- Analysis, interpretation, and reconstruction
- Presentation
- Summary and conclusions

Processes and challenges

Process	Faults	Failures
Identification	Make/Miss	False +
Collection	Content Context Meaning Process Relationships Ordering Time Location Corroboration Consistency	False -
Transport		
Storage		
Analysis		
Interpretation		
Reconstruction		
Presentation		
Destruction	Accident/Intent	

Properties of digital evidence

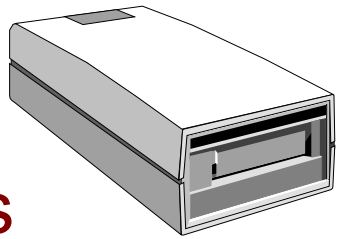
- Latent in nature
  - Can only be seen, understood, analyzed, and presented with and through tools.
- Often fragile and time sensitive
  - Sometimes exists for very short time periods
  - Easily destroyed or modified
  - Easily mishandled
- Meaning is only clear in context
  - Patterns of information combine to provide substance
- Like a puzzle you put together to get a picture
  - Easily misinterpreted
  - Often misleading
  - Often patently false

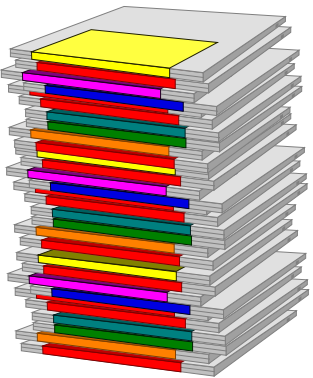
Outline

- Background
- Basic notions – processes and challenges
- Evidence identification and collection
- Transportation and storage
- Analysis, interpretation, and reconstruction
- Presentation
- Summary and conclusions

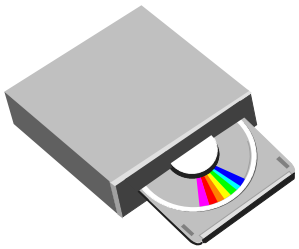
Data at rest

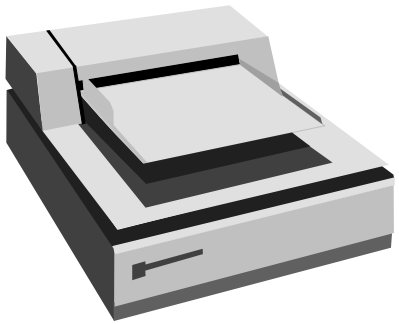
- File and non-file representations of content on physical media
  - Networked computers and storage arrays
  - Disks, memory, tapes, CDs, cameras
  - PC boards, PCMCIA cards, network boards
  - PDAs, cell phones, USB rings, pen recorders
  - Printers, answering machines, watches, etc.

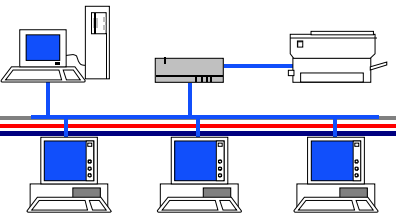




open, close, read, write, seek







Data in motion

- Mostly network services and communications
  - Server
    - Program opens a network port and waits for input
    - As input arrives, the program does its thing
      - Web server waits for ‘get’ and types out files
      - Mail server waits for email and delivers it
      - Telnet server waits for connection for login program
      - Ftp waits for file transfer requests and transfers files
      - IRC waits for chat sessions and relays content
  - Client
    - Program opens up port for output, makes request, awaits responses, shows to user, interacts
- Don't forget VoIP, tunnels, encryption, POTS, radios, cell systems, satellites, optical, etc.

Data in use

- Who is doing what and how?
  - Tagging users to content by behaviors / presence
  - Authentication
    - Something they have
    - Something they know
    - Something they are
  - What they do and what the computer does for them
- How do I get it?
  - Often present in audit trails across diverse systems
  - Often evidenced within content in storage
  - Sometimes has to be collected in real-time

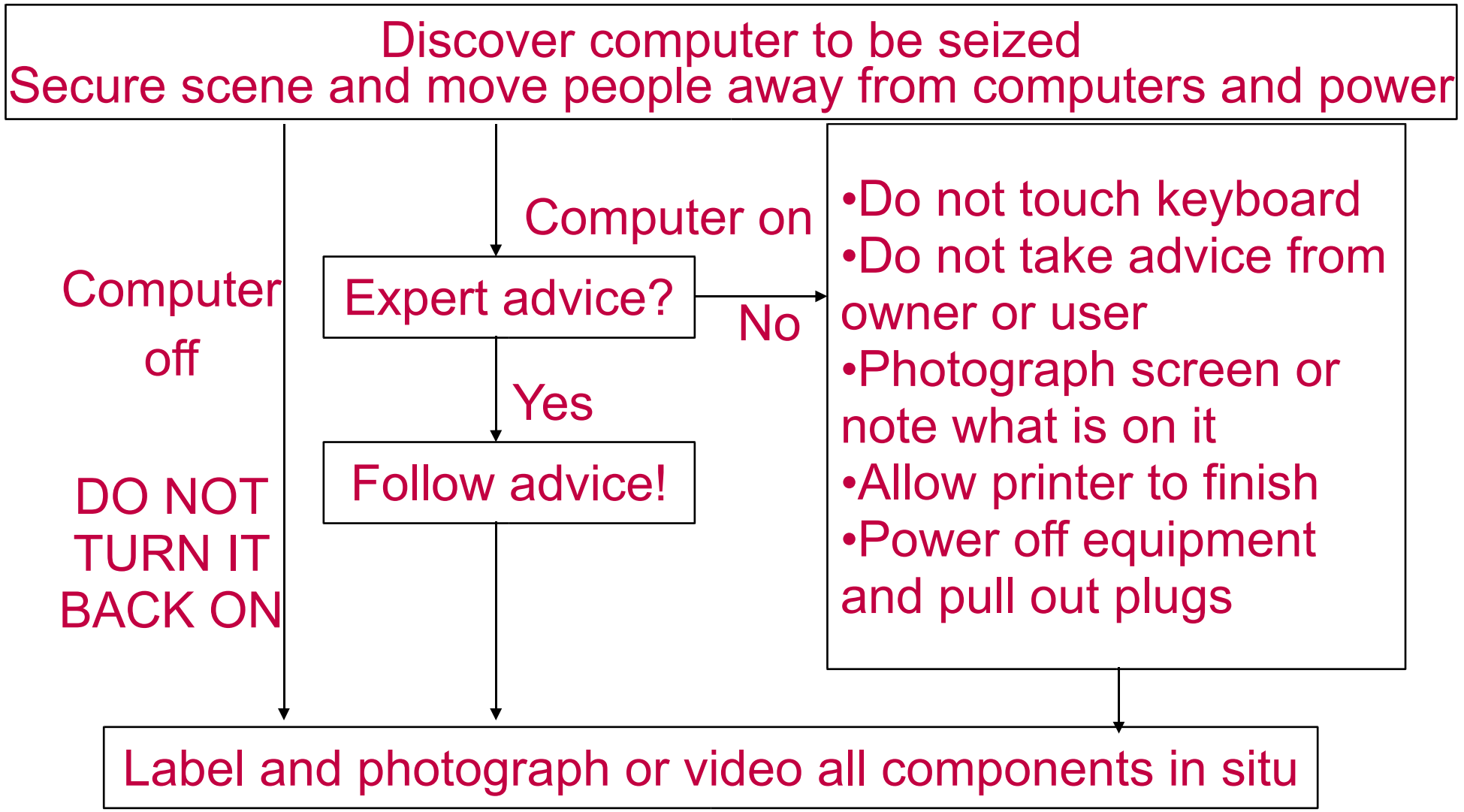
What is often missed/made?

- Missed
  - Failure to identify evidence as present
  - Failure to collect it while it is fresh
  - Failure to identify relevant materials for warrant
  - Failure to properly label and record
- Made
  - Identity things as evidence that are not
  - Collect things that are not allowed in the warrant
  - Mislabel things
  - Create forgeries (throw down computer)

Ceasing sources

- Take possession
  - Prevent destruction
  - Keep people safe
- Film the process?
- Photographs and labels
- Interview subjects
  - Get passwords, access codes, functional descriptions, etc.

"Good Practice"



"Good Practice" (cont)

- Remove and label all connection cables
- Remove all equipment, label, and record details
  - do this carefully - label and note serial numbers, etc.
- Ensure that all equipment is properly labeled
- Search the area for diaries, notebooks, papers
  - especially look for passwords or other similar notes
- Ask the user for passwords and record these
- Submit equipment for forensic examination

"Good Practice" (cont)

- What to seize
  - Main system box
  - Monitor, keyboard, mouse, leads and cables, power supplies, connectors, modems
  - Floppy disks, DATs, tapes, Jazz and Zip disks and drives, CDs, hard disks
  - Manuals and software, papers, circuit boards, keys
  - Printers, printouts, and printer paper
  - If in doubt, seize it!

Outline

- Background
- Basic notions – processes and challenges
- Evidence identification and collection
- Transportation and storage
- Analysis, interpretation, and reconstruction
- Presentation
- Summary and conclusions

"Good Practice"

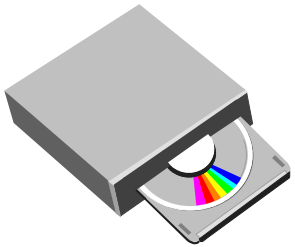
- Transport
  - Handle everything with care
  - Keep away from magnetic sources like loudspeakers, heated seats, radios, etc.
  - Place boards and disks in anti-static bags
  - Transport monitors face down buckled into seats
  - Place organizers and palmtops in envelopes
  - Place keyboards, leads, mouse and modems in aerated bags

Transporting evidence

- Many digital evidence forms are delicate
  - temperature sensitive - both hot and cold
  - shaking can damage them
  - bending connectors makes them fail
  - static electricity can effect them
  - dust, chemicals, water, and other factors harm them
  - time can cause them to decay
  - spores and fungi can damage them
  - magnets and light can damage them
- Accidental harm tends toward randomization
  - Can turn good evidence to bad evidence
  - Extremely unlikely to turn nothing into something

How long does it last?

- Tape, CDs, disks
  - 1-3 years if kept well
  - can fail in minutes
    - excessive heat
      - a car on a sunny day
      - a radiator or heater
      - a match
  - or in seconds
    - electromagnetic
      - a strong magnet
      - high impulse vibrations
      - overwritten



- Paper (non-acid)
  - hundreds of years
  - can fail in minutes
    - excessive heat
      - fire
      - heaters / radiators
    - shredding
      - a shredder
      - eating it
- Audit trails
  - some are never stored
  - others last minutes, hours, days, weeks, months, years

Retention requirements

- Companies don't have to keep it forever…
  - 7 years for some corporate / legal data
  - 4 Years for most civil data
  - Accounting records 7-10 years or longer
  - Email as a business record
  - Log files as business records
  - Legal process can force retention
- Check with your legal staff

Loss of data

- Bits go away with time unless they are maintained
  - Electromagnetism withers with time
  - Optical disks are susceptible to fungi
  - Magnetic tapes become brittle
  - Disks become warped, crack, etc.
  - Microfiche degrades and becomes a combustible
- Only active maintenance on live systems can really keep it right for long time periods
  - But live systems are more susceptible to other events

Storage of media and data

- Cool, dry, free of fungi
- No temperature cycling
- No humidity cycling
- No EMP effects
- Periodic review and rereading
- Crypto-checksums
- CRC codes and other data recovery methods
- Redundancy

Storage & transport challenges

- Miss
  - Content lost / stolen / decayed
  - Process failures and data losses / corruption
  - Chain of custody issues
- Make
  - Content altered / took too long to get there
- Most issues go against the prosecution
  - Incompetency shown and pointed out
  - Lost evidence “might” have been exculpatory
  - “You altered it during the missing 13 minutes”

Outline

- Background
- Basic notions – processes and challenges
- Evidence identification and collection
- Transportation and storage
- Analysis, interpretation, and reconstruction
- Presentation
- Summary and conclusions

The biggest source of problems

- When it gets to the lab the problems start
  - This is the really complicated part of the work
  - Nobody has all the expertise needed
  - Any little mistake can be highly problematic
- The goal of the analyst is to
  - Do no harm
  - Get at the facts
  - Make them sensible in context
  - Identify the presence/absence or event sequence

Principles (Best Practices)

- Principle 1:
  - No action taken by police or their agents should change data held on a computer or other media
- Principle 2:
  - In exceptional circumstances where examination of original evidence is required, the examiner must be competent to do it and explain relevance and implications
- Principle 3:
  - Audit records or other records of all processes should be created and preserved.
  - An independent third party should be able to reproduce the actions with similar results
- Principle 4:
  - The officer in charge is responsible for adhering to these principles

More best practice?

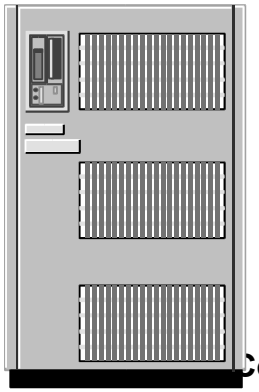
- There are decisions to be made at every step
- Make sensible judgments
- Base your judgment on science
  - refutable theory
  - experiments
    - confirm theory
    - can refute theory
- Philosophy of science
  - Popper

Outline

- Background
- Basic notions – processes and challenges
- Evidence identification and collection
- Transportation and storage
- Analysis, interpretation, and reconstruction
  - Imaging
- Presentation
- Summary and conclusions

Imaging the contents

- Image without alteration
- Evidence of integrity
  - Crypto-checksums
  - Process of collection
  - Process of analysis
  - Ability to reproduce
- Chain of custody issues
- Imaging techniques vary with the media imaged



- Mount copies for analysis
  - read-only
  - verify checksum before and after analysis
  - take good notes on the process along the way
  - provide for replay of the whole analysis if needed
  - maintain the chain of custody
- Better to analyze on a different system than it was produced on?

What is Best Practice?

- Nobody knows for sure
- Just because you are not perfect, doesn't mean it's not good enough
  - It's better to be better
- It's fundamental to have a philosophical rational
  - It's better to have thought it through
- Image without alteration
  - Heisenberg's Uncertainty
  - No experiment without alteration
- Purity established early
  - write protect
  - crypto-checksum
  - keep original pure
  - validate purity over time
- Record process
  - the ability to repeat it

Imaging challenges

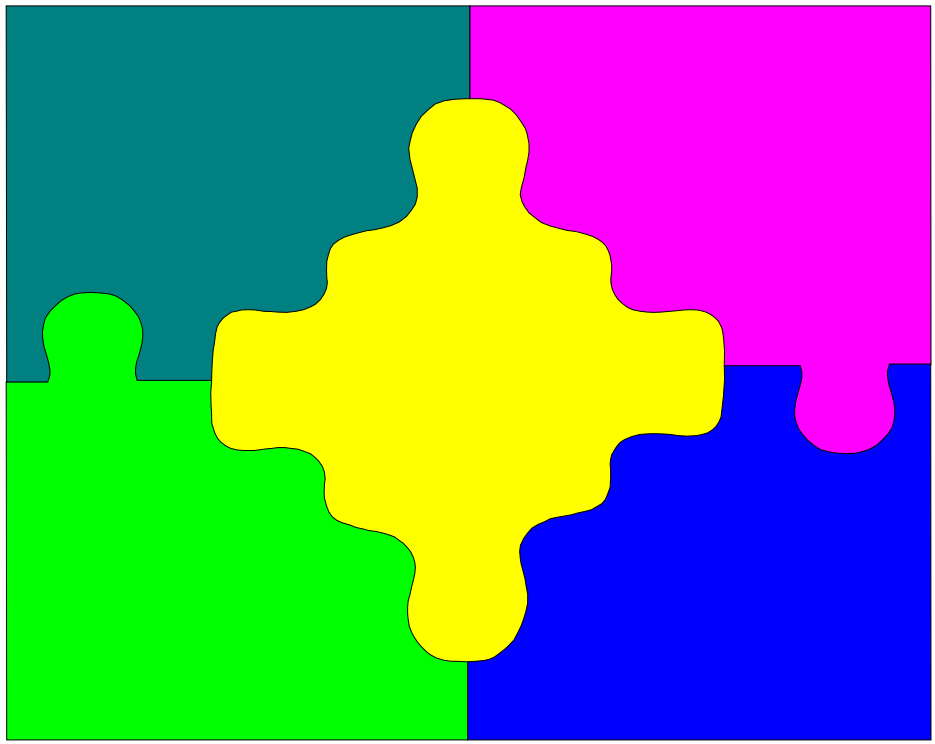
- Make / miss content
  - Was the ORIGINAL unaltered before imaging?
    - Chain of custody issues? Time available issues?
    - Could alterations be detected if present?
  - Was the copy exact and if not how not?
    - Was a cryptographic checksum used? Which one?
    - What tool was used? How good is it for this purpose?
    - Was the image onto a properly prepared disk?
    - Were there errors on the original? Copy?
    - Was hardware-level copy done? Does disk allow it?
  - Was the copy altered at any point and how?
    - Can we verify the cryptographic checksum again?
    - Was it verified at the end or periodically?

Outline

- Background
- Basic notions – processes and challenges
- Evidence identification and collection
- Transportation and storage
- Analysis, interpretation, and reconstruction
  - Analysis and interpretation
- Presentation
- Summary and conclusions

Analyze the data

- Putting together the puzzle
  - Generating and following leads
  - Identifying suspects
  - The process of elimination
  - Data formats
  - Derived evidence
  - Inculpatory and Exculpatory
  - Reconstructing the crime
  - Collating to collected data
  - Collecting more data
  - The smoking gun
  - Digital data is only a part of the overall picture



Generating & following leads

- Data gathered provides leads and indicates places to look
  - Example: files on a computer lead to web sites and postings which lead to user IDs which leads to other systems which provide more evidence
- How can I tell it's a lead?
  - The process of elimination takes out a lot of information
  - You are investigating something in particular?
  - Experience and the usual suspects

Identifying leads and evidence

- Whatever data you start with generates leads
  - IP addresses
  - Names in data files
  - System names
  - File types and content
  - Techniques used
  - Programs present
  - Associations and job
  - Unusual knowledge
  - And so forth...

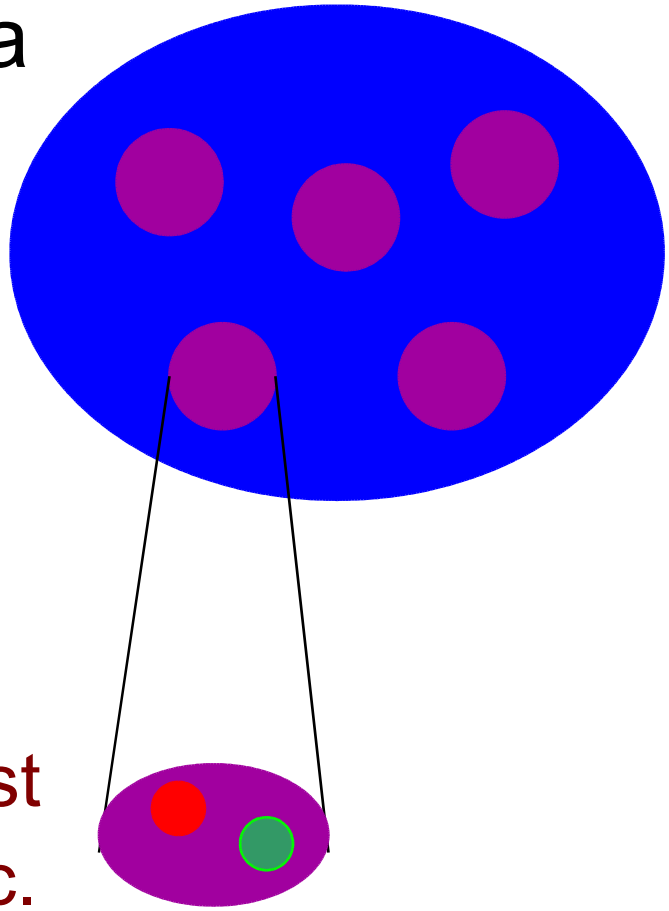
Means, motive, and opportunity should be present

Errors generated

- Each process can generate errors
  - Identify and challenge errors not eliminated
    - IP addresses – can be forged
    - Names in data files – how can they get there?
    - System names – how are they arrived at?
    - File types and content – defeat in detail
    - Techniques used – did they follow process?
    - Programs present – who put them? legitimate uses?
    - Associations and job – other candidates?
    - Unusual knowledge – who else has the knowledge?
    - And so forth…

The process of elimination

- Start with lots and lots of data
- Throw away known irrelevant data
  - e.g., standard OS stuff
  - data outside times of interest
  - data from and to elsewhere
- Focus examination on what's left
  - Unusual names, location, content
  - Encrypted or encoded data
  - Data with dates and times of interest
  - Erased data, logs, missing data, etc.
  - Pictures, scripts, programs, etc.



Evidence is evidence exculpatory counts too mostly dead ends

Elimination challenges

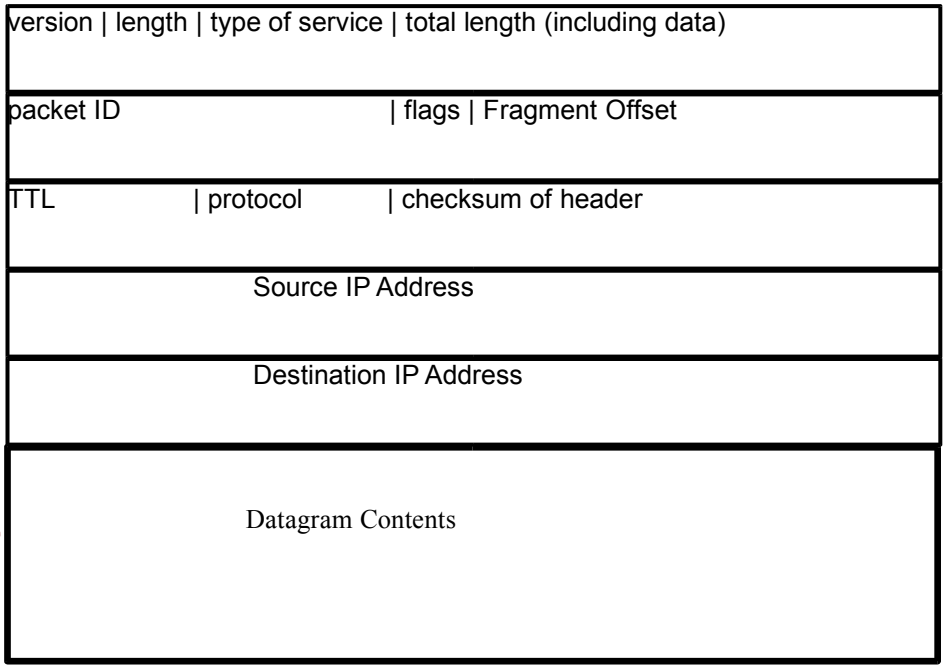
- Miss – usually through misinterpretation
  - Content seems more important than it is and is kept and considered instead of eliminated
  - Meaning exaggerated in context to create false line
  - Process analysis misses missing parts, leaps too far
  - Relationships not eliminated when false links followed
  - Ordering identified when none is actually present
  - Time setting and changing errors or offsets ignored
  - Location assumptions made and not fully checked
  - Consistency errors not identified to investigate more
  - Misinterpretation of meaning leads to false evidence

Elimination challenges 2

- Make – usually through over-interpretation
  - Content threw away something should have kept
  - Meaning missed resulting in key item eliminated
  - Process analysis misses leap to find new evidence
  - Relationships eliminated when they shouldn't be
  - Ordering missed when it is present
  - Time not properly correlated when it can be
  - Location not identified when it can be
  - Consistency errors falsely identified
  - Misinterpretation of meaning fails to find real evidence

Data formats

- Lots and lots of them
  - standard and non-standard
  - languages and character sets
  - compressed and packed
  - embedded and encoded
  - encrypted and transformed
  - context-dependent
  - combinations and recursive
- Some are often missed or misinterpreted leading to missed content



Derived evidence

- Evidence is often derived from other evidence:
  - Two records at the same time in different places
    - 1 record is in central time, the other in pacific time
    - System time and time zone and clock skew must be combined with times indicated to determine that the events are at the same time
  - Search of a system yields data on other systems
    - A web cache file indicates pornographic content
    - The remote we site contains records of the activities
    - One of the activities involves drugs
    - The system with the drug information links to bank transfers
    - And so on

More derived evidence

- Crimes are sequential events in digital systems
  - Picking out the different sequences can be very hard
    - IP traffic can be collected at 100,000,000+ bits per second and can involve data from hundreds of sources and destinations
    - The data we look for involves sequences between select pairs of participants
  - Correlation with system data is complex
    - Correlate traffic with activity logs at end points
    - Correlate traffic with activity logs in intervening infrastructure
    - Correlate traffic with dial-in times, phone bills, etc.

Still more derived evidence

- Data is stored or transmitted encrypted
  - Decryption yields user IDs and passwords
- Audit trails indicate an attack is underway
  - Based on what the attack might do, we try to ascertain motive and skill levels
  - Using likely intent, we seek to search the source system
  - The search turns up tools capable of achieving the assumed intent
  - Information on the suspect system indicates success against other targets

Inculpatory and Exculpatory

- Evidence is evidence
  - If it says guilty, you say guilty
  - If it says innocent, you say innocent
- Forensics deals in facts and interpretation
  - interpretation is opinion based on experience and data
  - facts are collected and documented
- Most interpretation can be pretty objective
- Some interpreters can be pretty subjective
- The truth can be verified by experiments

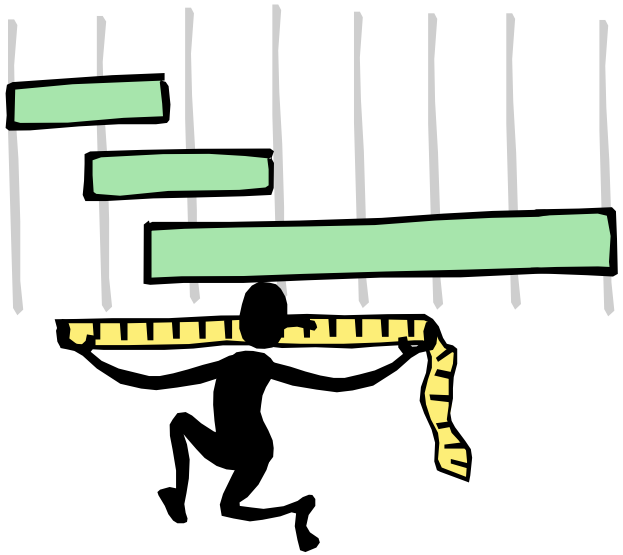
–**The scientific method**

Collecting more data

- As you start to see the mosaic, pieces are missing
  - How much data do you need?
  - How do you collect more?
  - What's the cost?
- But…
  - The next piece could prove my client innocent
    - Indeed it could…
    - Or it could continue to make the picture clearer
    - Or it could be irrelevant
  - People make judgements and people aren't perfect

What does it tell us

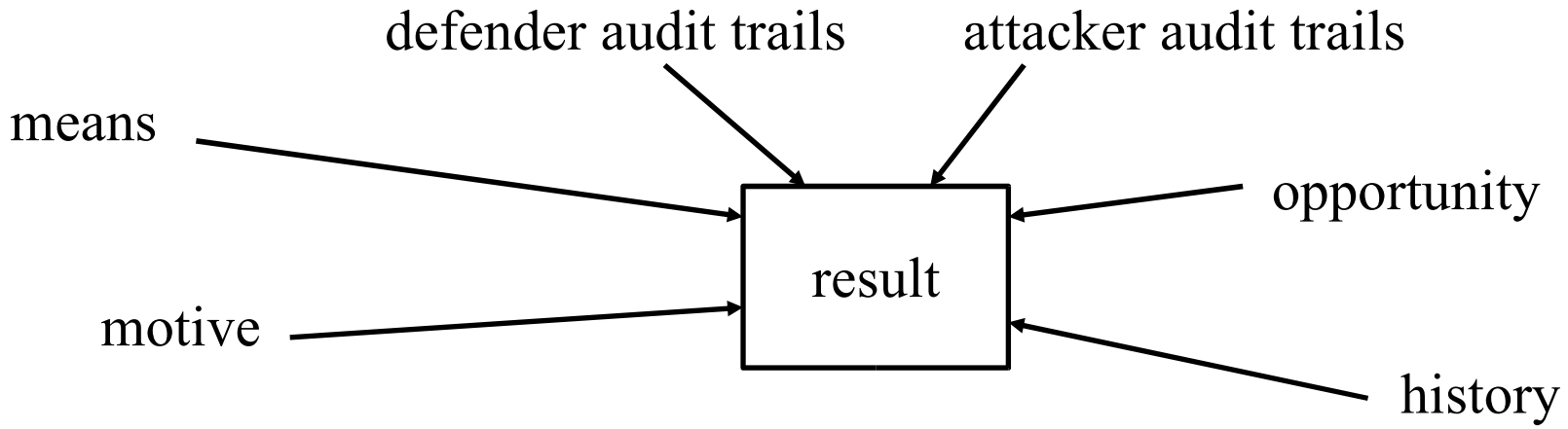
- Details of things that happened inside of computers
  - At this time
  - This thing was done
  - By this program
  - Acting for this user
  - With this result



- The timeline and pattern of these is interpreted to demonstrate criminal activity, cause, & intent

How to be more certain...

- More related facts limits alternative explanations
- Other forms of evidence help convince juries
- In some cases we can do mathematical analysis
  - to show that specific other explanations are unlikely
  - to show how complex it would be to generate in other ways

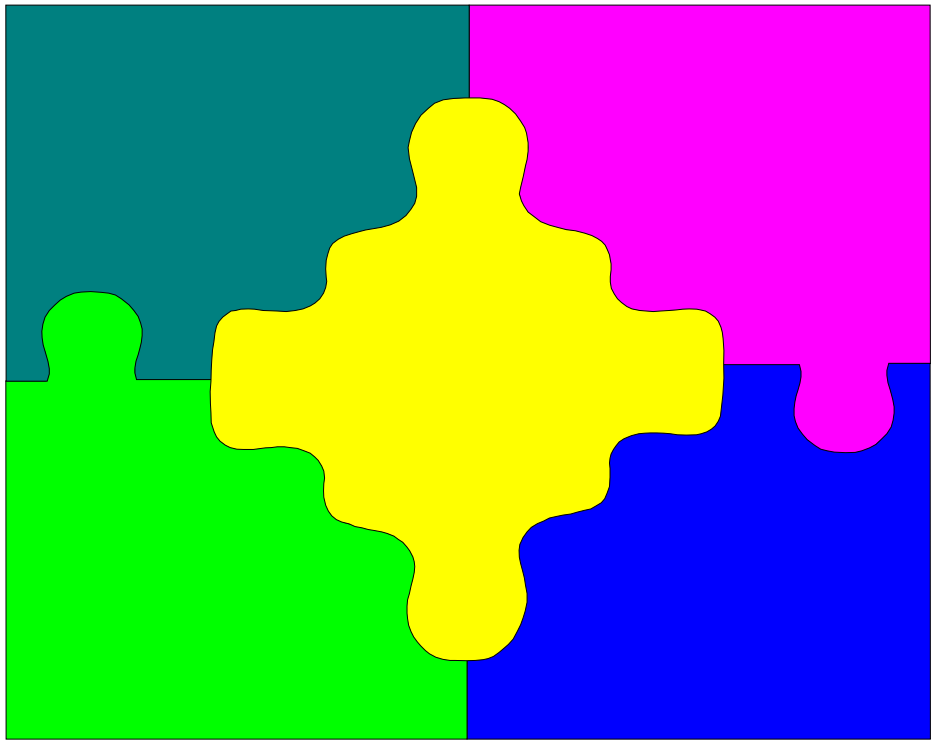


But I wasn't at the keyboard!

- I was on a break…
  - I was in the other room…
    - I leant the system to my brother in law…
      - Someone broke in and did it…
        - It was a mistake…
- There are a million explanations - but only one truth
  - Find the most consistent set of facts
  - Present them in an accurate light
  - Refute any erroneous facts given by any party
  - Let the jury decide

Digital data is only a part of the overall picture

- Physical evidence
- Means, motive, opportunity
- Follow the money
- Statements of witnesses
- Corroborating evidence
- Exculpatory evidence
- Statements of the accused
- Paper trails and pictures
- Eye witnesses



If it looks like a duck and quacks like a duck...

Validation - case study

- Civil trial - expert witness for one side turned into a special master for the court
- The ‘special master’ was clearly biased and had made many unsupportable statements
- In the end, two things won out:
  - refutation of several ‘expert’ opinions by physical evidence (analysis of a video tape, limits of time to do things the claimant was accused of by the SM)
  - evidence of the defendant violating the court’s order by perpetrating acts during the search and seizure

Outline

- Background
- Basic notions – processes and challenges
- Evidence identification and collection
- Transportation and storage
- Analysis, interpretation, and reconstruction
  - Reconstruction
- Presentation
- Summary and conclusions

Reconstructing the crime

- Crime scene reconstruction and fidelity
  - Build a copy of the crime scene
    - To the level of fidelity feasible based on resources and value
  - Try to reproduce the crime as postulated
    - Do the same things you claim the perpetrator did
  - Look for confirmations and refutations
    - Look for changes in system state and event sequences
      - Every time it runs it will run differently
      - Find the differences and commonalities
      - Compare to the evidence available
  - Look at the mosaic as a whole picture

Reconstruction challenges

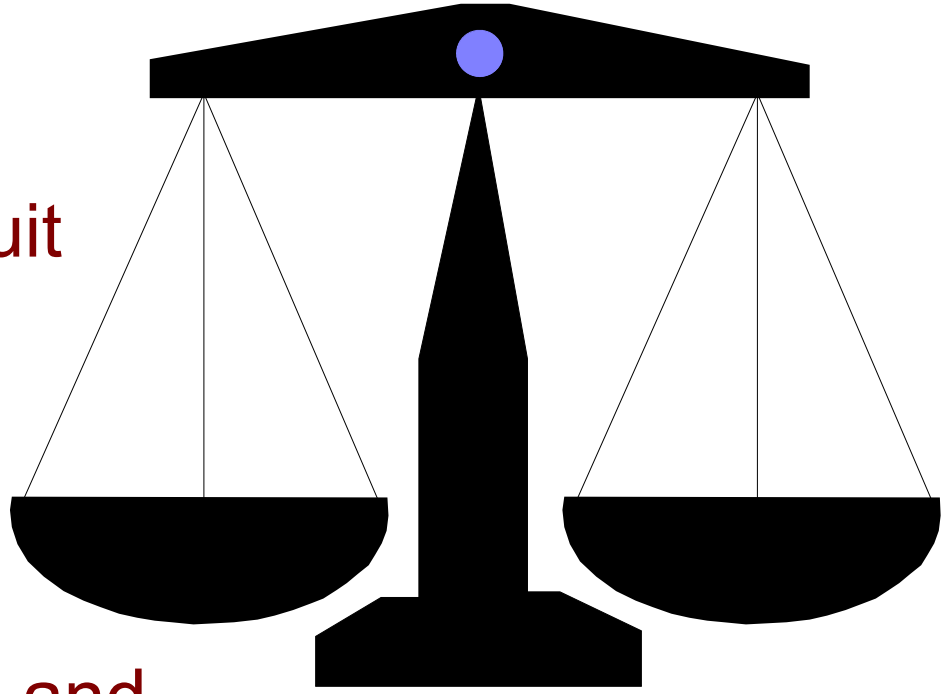
- We are trying to create a before and after picture
  - How did we get the before?
  - How well did we do it?
  - How can we tell how well we did it?
  - How do we compare before to after?
  - How accurate is the reconstruction?
  - What is relevant and irrelevant?
  - Ocam's Razor?
    - The simplest it can be - but no simpler
  - What is a model?

Outline

- Background
- Basic notions – processes and challenges
- Evidence identification and collection
- Transportation and storage
- Analysis, interpretation, and reconstruction
- Presentation
- Summary and conclusions

Process requirements

- Properly obtained evidence
  - Permission of the owner
  - Permission of the suspect
  - Search warrant / hot pursuit
- Chain of custody
  - Properly found, collected, annotated, taken apart, transported, searched, analyzed, stored, tracked, and copies provided to the other side
- Exculpatory & inculpatory evidence presented



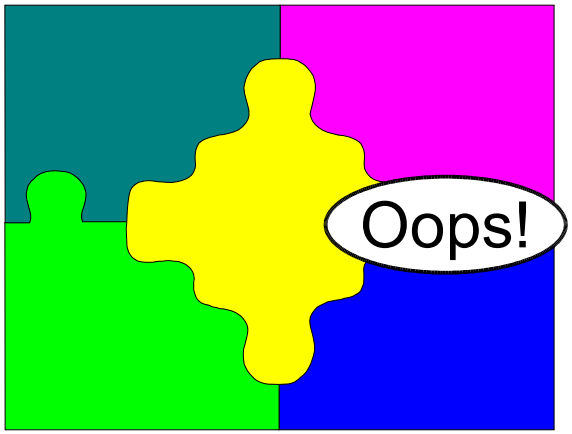
Chain of custody issues?

- Just like any other evidence
  - must get to it in time
  - must collect it properly
  - must transport it properly
  - must hold it securely
  - must analyze it carefully
  - must leave evidence in tact
  - must provide repeatability
  - must be available for defense
  - must be presentable in court
  - must be explainable in court

Just because it's in a computer doesn't make it right

The mosaic

- Mosaics are rarely perfect in digital evidence
  - There are missing pieces
  - There are slightly off pieces
  - There are extra pieces
- There are reasons for this that can be demonstrated
- In the end, the real issue is
  - Does it look like the suspect?
  - How revealing it is?
  - Is it consistent with the other evidence?
- Is it more probative than prejudicial?
  - More often than not – they let the jury decide



Presentation strategies

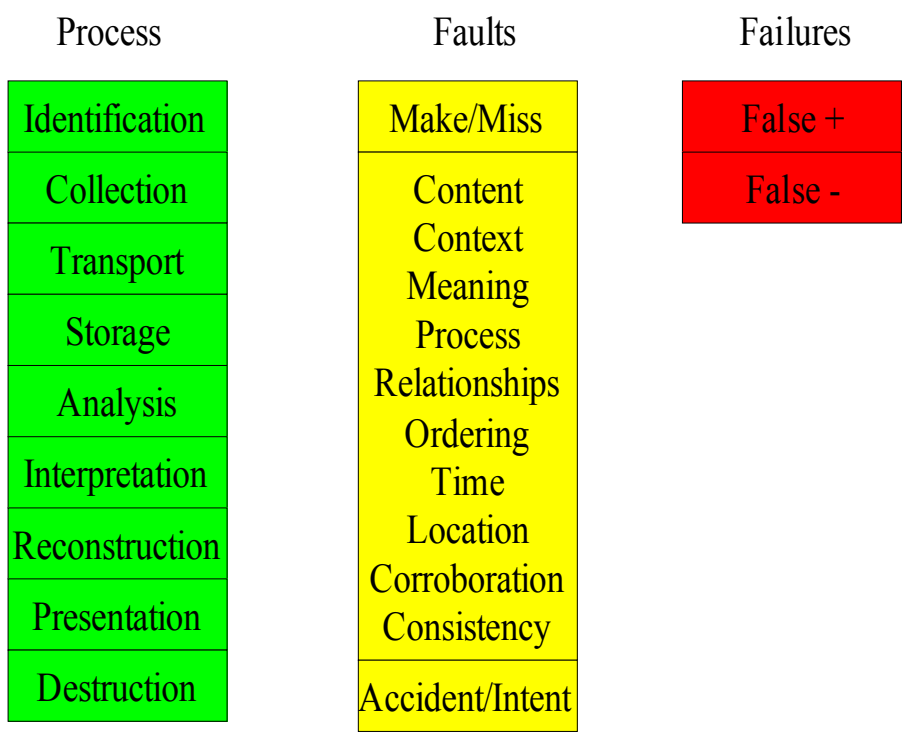
- Present the mosaic of the case in all its glory
  - Show the process that you assert took place
  - Step through the evidence supporting your process
- Address strengths and weaknesses along the way
  - Most favored position first and last
  - Alternatives in the middle
- Summarize the evidence in context of the process
  - Address anomalies before your opponent does
- Draw conclusions - if any
  - Provide the basis for drawing those conclusions
  - Address other possible interpretations and their basis

Outline

- Background
- Basic notions – processes and challenges
- Evidence identification and collection
- Transportation and storage
- Analysis, interpretation, and reconstruction
- Presentation
- Summary and conclusions

Meeting the challenge

- There are two sides to every story
- The other side is obliged to do their best
- The goal is to have the best evidence you can get
- No case or evidence is perfect
- No person is perfect
- Expect to be challenged
- Don't become defensive
- Don't take sides



Accuracy?

- Computer records are NOT ALWAYS accurate
  - Times and dates are often off
    - compare them to a standard
    - http://all.net/
      - press “what time is it?”
    - note your computer’s time
    - note the Navy’s time
  - File time/date stamps altered
    - echo “test” > aa;cat aa
    - ls -l aa;sleep 60;touch aa
    - ls -l aa;cat aa;rm aa
- Inaccuracies can be intentional
  - cd /u/local/attacks
  - ls genocide/log-wipers
    - ‘cloak’ and ‘cloak2’
      - wipe you from /var/adm/*
      - wipe you from utmp
      - fully automatic
      - need to be root
    - ps hiders
      - change your process names to hard-to-detect ones
  - ls sabotage/rootkit;cd
    - create backdoors

Timeliness?

- Real-time capture
  - network traffic, telephone calls, IRC sessions
- Rapid capture (hours-days)
  - ISP dial-ins, system logs, backups, cache files
    - telnet defender
    - less . netscape/cache/index.db
    - find the evil files
    - 15 minutes
- Timely analysis
  - PDAs
    - run out of power
    - lose memory
  - disks/tapes/CDs
    - 1-2 years expected lifetime
    - some last much longer
    - heat/magnets can destroy
  - data leading to other data

Completeness?

- Most computer records are fairly minimal
  - date, time, major event
  - sometimes only start, not stop
  - sometimes only stop, not start
  - content often missing
  - user information limited
- Better logs are easy to get
  - keystroke logging
  - tapping specific IPs / ports
  - log files take space
- Computer records can be missing things
  - entire records can be missed
  - attackers try to destroy logs
  - attackers try to avoid logs
  - attackers try to forge logs
  - almost always some evidence of alteration
  - use redundancy

Admissibility?

- Most computer records come in under the business records exemption from heresy
- They come in through expert testimony
  - Systems administrator declares that they were taken in the normal course of business
  - Indicates specific actions taken to collect records
  - Shows them in light of other records taken and kept
  - Expert witness explains and interprets the records
  - Opposing experts make their claims

How valid is it?

- Computer data is easily altered
  - by attackers it is done in the normal course of events
  - by defenders to make it look like an attack?
  - by accident all of the time
    - but rarely by hardware faults
    - sometimes by software faults
- It's usually fairly easy to tell if it was altered
  - it's gone completely over a portion of time
  - inconsistencies show up across audit trails
- It's hard to alter undetectably

Overcoming challenges

- You'd better have some more evidence!
  - Digital evidence - so far - has NEVER been the whole case in any successful litigation
  - The money paid for the secrets when sold?
  - The time stamps indicate that nobody else could have used tty31 at that time?
  - What web sites did you visit? What did you do there?
    - Check against those web site audit records
      - I don't really remember any of the details, I just like to cruise the web
        - Proxy? Review the logs…
- A pretty good defense!

How valid is it?

- Unaltered audit information is not always correct
  - Forged email, sessions, etc. look just as real
  - If I break in to user ‘Joe’ it looks like ‘Joe’ did it
  - Audit records fail during high load conditions
  - Audit records fail under unanticipated conditions
  - Some programs don’t produce records consistently
- Audits can often be avoided
  - Program audits are easily bypassed by not using the program to alter the data

What can go wrong?

- Pull the plug or not?
  - astute points on both sides
- Marking it properly
  - bag and tag techniques
- Transport sensitivity
  - shaking
  - temperature
  - dust, fumes, magnets, etc.
- Storage sensitivity
  - time in storage
  - also like transportation
- Analysis errors
  - not working the copy
  - modifying/deleting evidence
  - missing evidence
  - misreading evidence
  - not getting redundant data
  - looking excessively
- Presentation errors
  - talking technical
  - not using pictures
  - denying weaknesses of digital evidence

Legal Challenges

- Jurisdiction
  - Global nature of IT
  - Most communications are interstate or international
  - Cooperative agreements
  - Evidence problems
- Case Law
  - There isn't enough
  - We are always guessing
- Qualifications
  - No standard qualifications for expert witnesses
- Privacy
  - It's hard to get some things w/in the rules
- Search Warrants & Permission
  - Validity of broad searches
  - What can you look for?
  - What do you ask for?
  - Expand as you learn more
  - Permission==agreement?
- Privileges
  - Doctors, Lawyers, Clergy

Questions?

Thank You

http://all.net/ **fc@all.net**