Building a new scientific theory and practice of digital forensics

2013-05-22

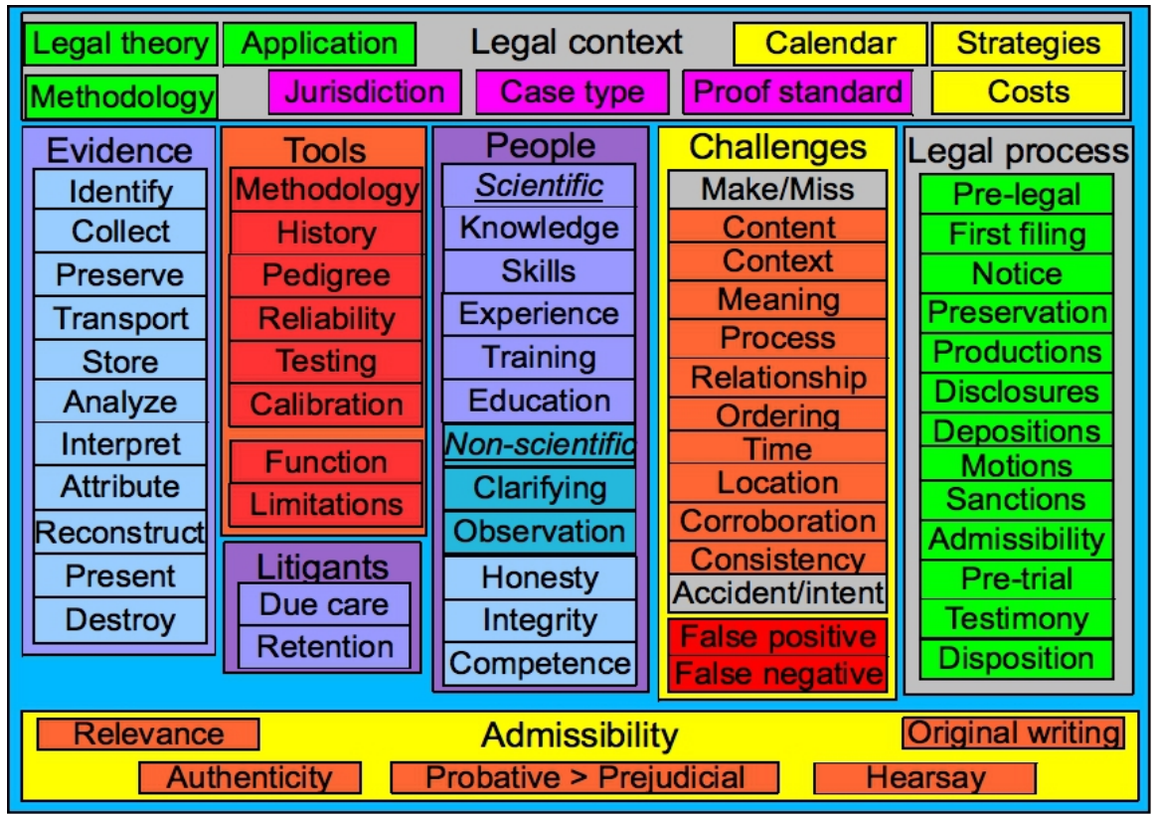
1st International Summit of Digital Forensics

Keynote Address

Dr. Fred Cohen
CEO – Management Analytics
CEO – Fred Cohen & Associates

Outline

- Then, now, and then again
  - Political systems and approaches
  - Technological changes
  - Scientific basis for digital forensics
- Building our scientific future together
- Discussion



Past – Present – Future

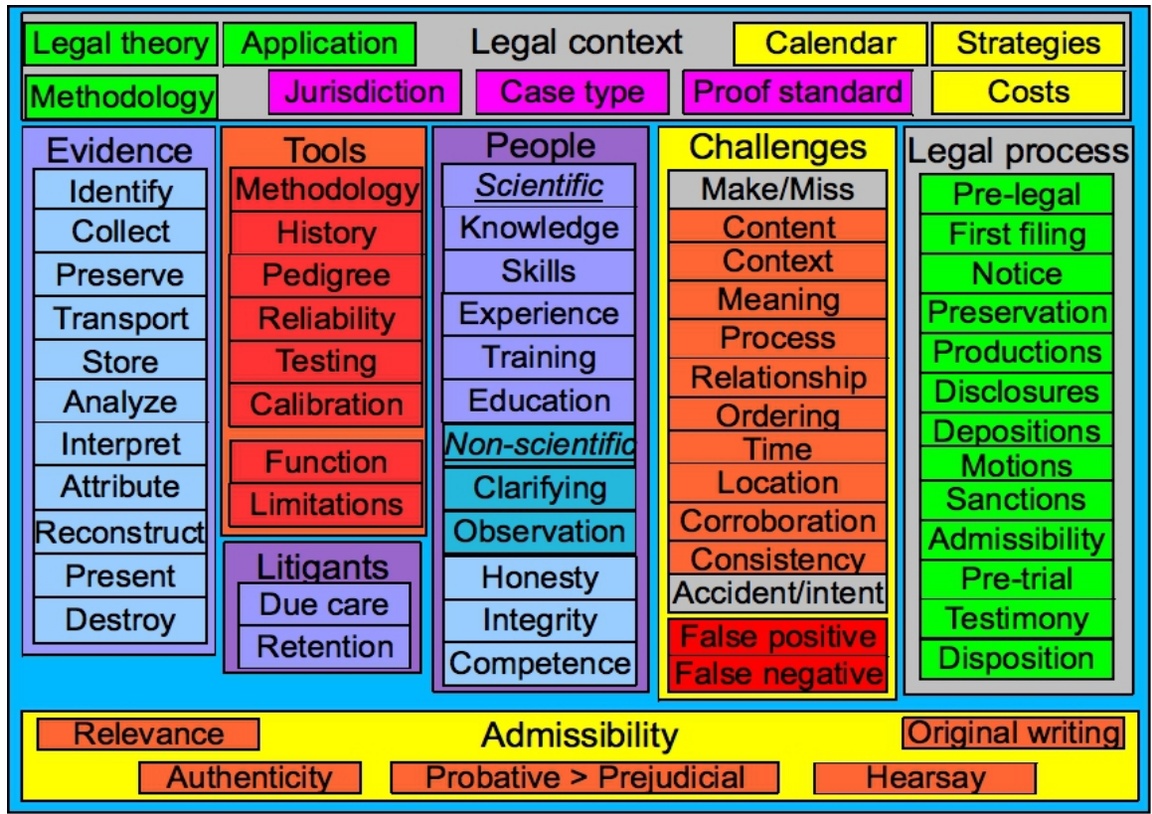
- This talk is arranged as follows
  - Each topic area is covered
    - Political systems and approaches
    - Technological changes
    - Scientific basis for digital forensics
  - Within each, we go past – present – future
  - Each builds on the others with examples
  - Then we talk about how to move forward together

Timeline exemplars

- Past
  - Pre-Internet – mainframes and midrange – connected via serial lines or low-speed links – local access via identified terminal lines – in enclosed limited access facilities – small number of local users – globally 1-10M users?
  - Internet circa 2000 – servers, workstations, limited mobile, serial lines except servers on T1s, largely fixed IP addresses, increasing deployment to wider audiences – 100M users worldwide – 100,000 programs or more
- Present
  - Social media – Billions of users – global reach – cellular and WiFi mobility widespread – many millions of programs
- Future - Massive sensor grid tomorrow – everyone and 100B things connected – unlimited dependencies

Outline

- Then, now, and then again
  - Political systems and approaches
  - Technological changes
  - Scientific basis for digital forensics
- Building our scientific future together
- Discussion



Political systems and approaches

- Different jurisdictions have different political systems
  - China / US / Iran / Zimbabwe / European Union / etc.
  - Example over time:
    - In the 1960s, Iran had the Shah, Iraq had Hussein
    - Now Iran has the Ayatollah, Iraq has mixed governments
- Different political systems have different rules
  - The US and others have an oppositional system of justice
    - Defendants have experts who challenge the prosecution
  - China and others have a cooperative system of justice
    - Prosecution has experts that check on each other
- And the rules change with time
  - Legal changes and international law is evolving

Within governments as well

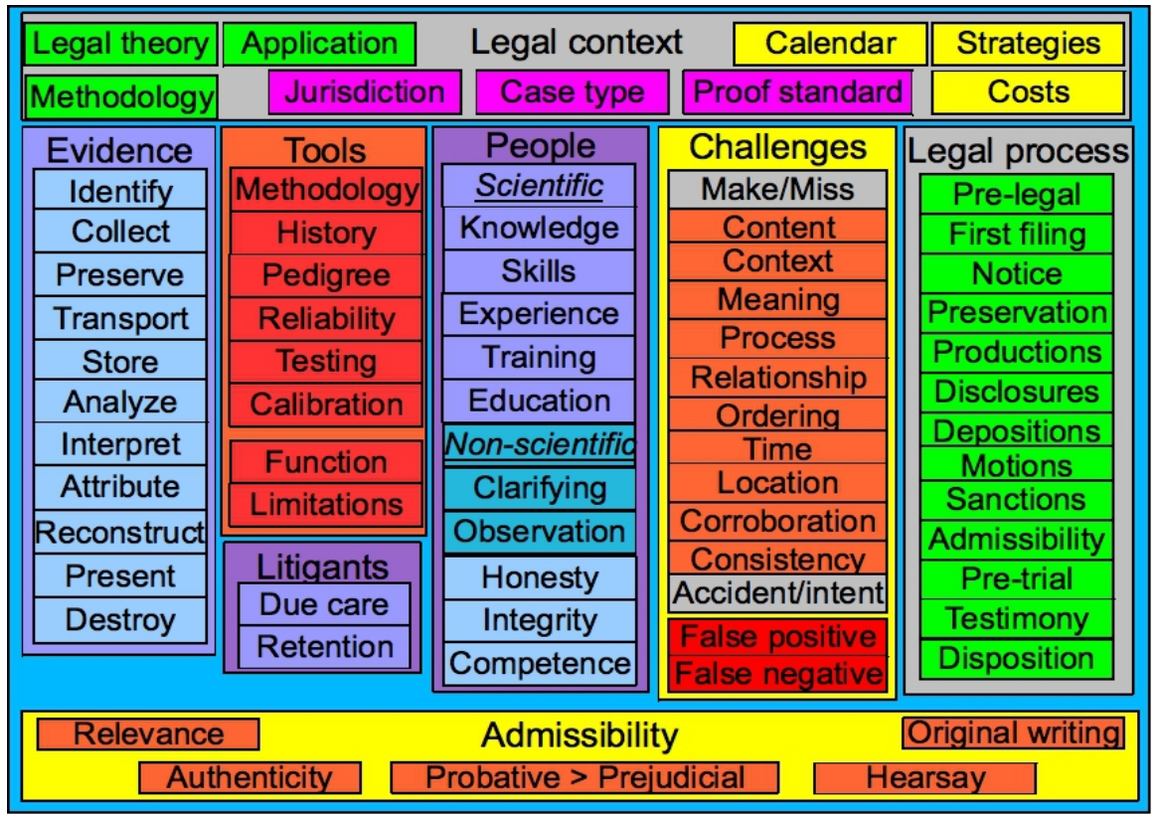
- The US has highly differentiated national v. international
  - National is all Federal, State, Local laws and police
  - International is largely intelligence agencies and military
  - As a result, Law Enforcement (internal) is handicapped compared to international
    - Example: Intelligence (external) has massive R&D
    - Result: They build detection, attribution, and deception
    - Example: Law Enforcement has no R&D function
    - Result: They largely purchase commercial and apply
    - Net effect: Law enforcement (internal) is handicapped
  - HOWEVER: That is a policy decision of the political system
- Many other countries have unified internal and external
  - Thus they can apply higher quality methods internally

Political systems also affected changes

- The US political system also produced startling advances
  - Example: I did security protocols for DISN in the 1970s
    - ARPA developed the Internet in the 1970s as well
    - The Internet overtook [most things] as a resilient system
  - Example: There was social networking in the 1970s
    - Social networking as a business has changed the world
    - Twitter and Facebook have helped revolutions
- But we don't necessarily know how to cope with them
  - Digital forensics pre-Internet
  - Digital forensics Internet
  - Digital forensics social media
  - Digital forensics massive sensor grid

Outline

- Then, now, and then again
  - Political systems and approaches
  - Technological changes
  - Scientific basis for digital forensics
- Building our scientific future together
- Discussion

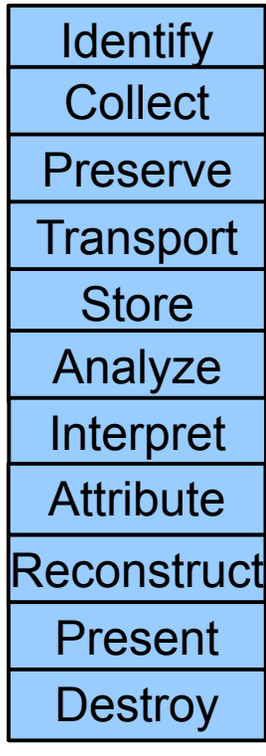


Digital forensics over time

- Political systems
  - Pre-Internet:
    - Almost all local to a well known jurisdictions and laws
    - Rare cross-jurisdictional identified by phone records
  - Internet:
    - Jurisdiction unclear, almost always multiple jurisdictions
    - Much of it international in nature, records unclear, subpoenas slow and difficult but still doable
  - Social media:
    - Jurisdictions often unknown till after much investigation
    - Massive collections with little attribution other than self-indicated and via multiple 3rd, 4th, etc. parties
  - Massive sensor grid
    - Equities issues start to abound, global treaties needed

Digital forensics over time

- Identify / collect / preserve
  - Pre-Internet:
    - Find computer(s) / image them / secure images
  - Internet:
    - Find some computers or content
    - Collect online and try for more subpoenas
    - Secure some content without system context
  - Social media:
    - Find some links and do link analysis for more
    - Collect online and through other for a, loop for more
    - Secure some content and link analysis with little context
  - Massive sensor grid
    - Ask people to send in what they have…



Examples

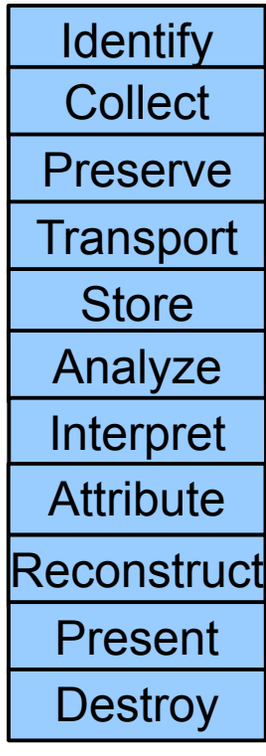
- Toll fraud circa 1970s
  - Go to telco switching center for records
  - Find phone lines in use and trace to physical location
  - Secure records and suspect
- Internet investigations circa 2000
  - Get IP address of source of bad thing
  - Get location from ISP and/or DNS records
  - Secure disk images or records from various computers
- The Boston Marathon Bombers circa now
- Anonymous and beyond circa who knows when

Examples

- Toll fraud circa 1970s
- Internet investigations circa 2000
- The Boston Marathon Bombers circa now
  - Ask the public to send in cell phone and CCTV pictures
  - Collect via email and Web sites or physically
  - Get information on sources for later recall and testimony
- Anonymous and beyond circa who knows when
  - Identify relevant intelligence (massive surveillance)?
  - Do vast scale global search for relevant content?
  - Secure likely relevant parts before retention time runs out?
- Where do the resources for this come from?

Digital forensics over time

- Transport / Store
  - Pre-Internet:
    - Carry devices on a vehicle to a laboratory
    - Store in a lockup
  - Internet:
    - Transport information via Internet or physically
    - Store in a file server and physically
  - Social media:
    - Transport all information via Internet only
    - Store only in local file stores
  - Massive sensor grid
    - Transport by whatever Internet they have?
    - Store in mass scale shared storage (cloud)?

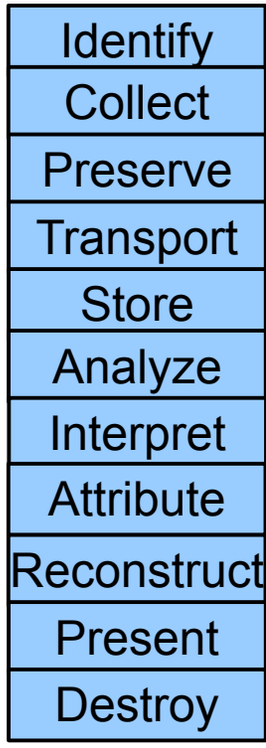


Examples

- Toll fraud circa 1970s
  - Transport records in a car to police lockup
- Internet investigations circa 2000
  - Carry the disks or records in a car / FedEx to a lockup
- The Boston Marathon Bombers circa now
  - Internet transport except for CCTV – carry in a car
  - Internet transport to file server / CCTV copies to file server
- Anonymous and beyond circa who knows when
  - Internet transport to cloud services
  - Stored in cloud service provider storage

Digital forensics over time

- Analyze / Interpret / Reconstruct
  - Pre-Internet:
    - A person reviews it all, interprets, documents
    - Rare reconstruction – limited non-forensic tools
  - Internet:
    - A person reviews it all, interprets, documents
    - Limited reconstruction – some forensic tools
  - Social media:
    - People write programs to analyze / try to interpret results
    - Link analysis and timeline tools limited – crime recon.
  - Massive sensor grid
    - Only automated analysis feasible, interpretation unclear
    - Only gestalt reconstruction – no tools yet

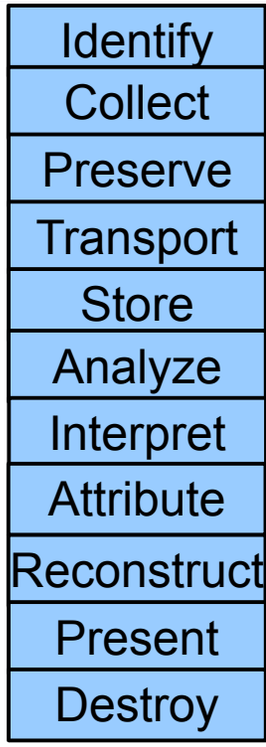


Examples

- Toll fraud circa 1970s
  - Expert reads the details, analyzes, interprets
  - Phone company reconstructs all details to demonstrate
- Internet investigations circa 2000
  - Limited automated analysis (e.g., search), expert interprets
  - Rare reconstruction using models built to purpose
- The Boston Marathon Bombers circa now
  - Largely human reviews, limited color and dress recognition
  - Repurposed tools with much manual effort to reconstruct
- Anonymous and beyond circa who knows when
  - Sources and targets too broad / diverse for task forces
  - Global intelligence agencies are the only real hope

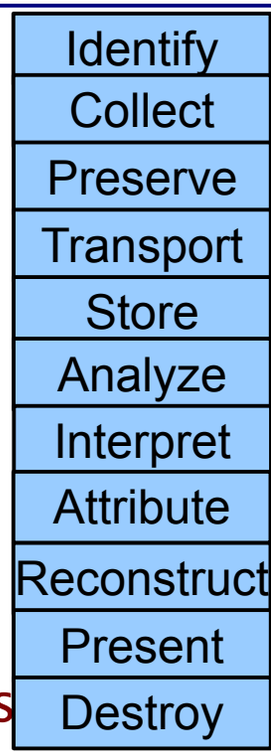
Digital forensics over time

- Attribute / Present / Destroy
  - Pre-Internet:
    - Limited suspects/physical access @place @time
    - Present on a few poster boards with paper output
    - Shred / burn / clear (overwrite) to destroy
  - Internet:
    - Many possibles / attribution via IP address but…
      - IP addresses are problematic for this for many reasons
      - Jurisdictional issues are often present and problematic
      - Anonymity and pseudonymity make it harder
    - Present in many forms with complex explanations
    - No real way to destroy
  - Social media:
  - Massive sensor grid



Digital forensics over time

- Attribute / Present / Destroy
  - Pre-Internet:
  - Internet:
  - Social media:
    - Enormous numbers of possible suspects
      - Location highly problematic and indirect
      - Attribution depends on unique users at many points
      - Widespread lack of adequate and understood records
    - Presentation complex especially for large networks
    - Destruction practically impossible
  - Massive sensor grid
    - Attribution depends on sensor network – many covert
    - Presentation of a very small subset / challengeable
    - Destruction completely out of control and non-owned

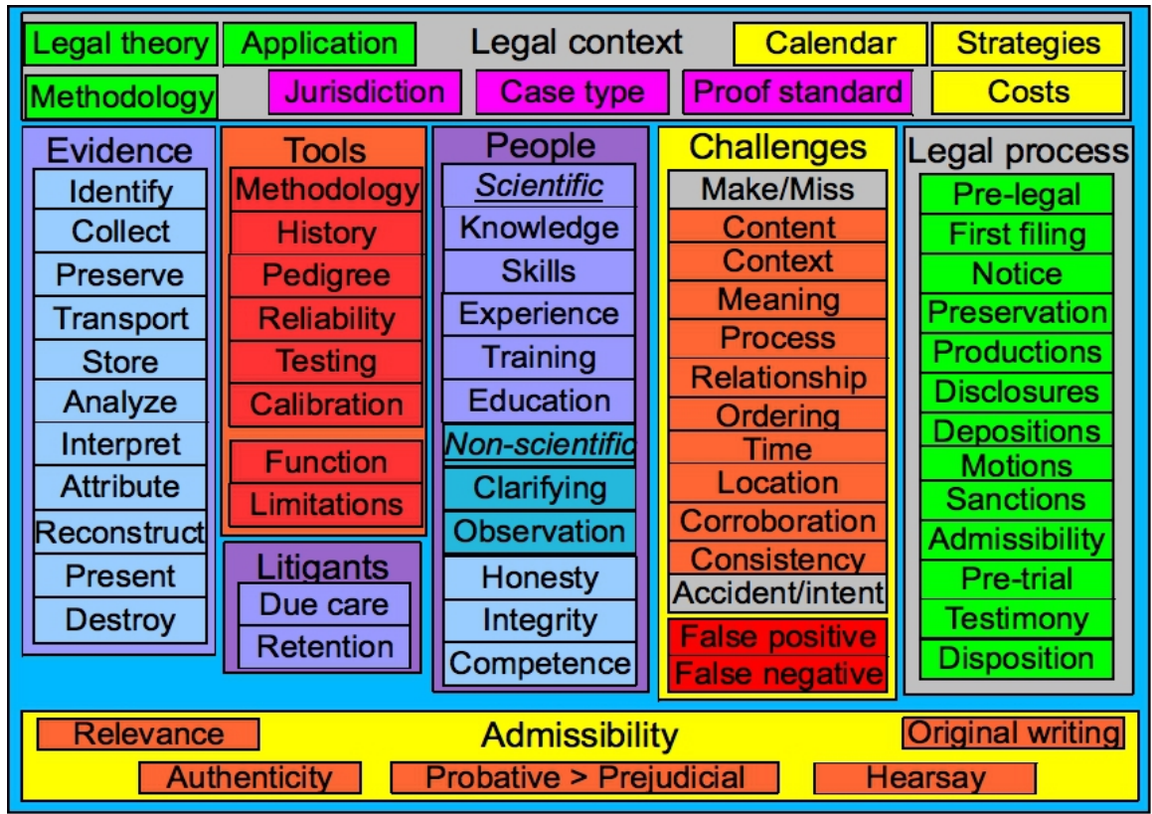


Examples

- Toll fraud circa 1970s
  - Physical limits augmented with detailed records attribute
  - Simple to present at low volume / destruction by burning
- Internet investigations circa 2000
  - IP address and related records with testimony to attribute and present. Destruction by burning / overwriting
- The Boston Marathon Bombers circa now
  - Attribution by large numbers of correlated pictures/data
  - Presentation complex fused near-movie / no destruction
- Anonymous and beyond circa who knows when
  - Attribution largely infeasible for a long time today
  - Presentation of secret stuff problematic / no destruction

Outline

- Then, now, and then again
  - Political systems and approaches
  - Technological changes
  - Scientific basis for digital forensics
- Building our scientific future together
- Discussion



Science over time

- The scientific basis / or lack thereof ($C \rightarrow^{M} E$)
  - Pre-Internet: Physics and electrical engineering
    - Few possible causes (C) and limited effects (E)
    - Limited available mechanisms(M) for $C \rightarrow^{M} E$
    - Reconstruction easy and relatively direct
  - Internet:
    - $C \rightarrow^{M} E$ too large and lack of adequate tools/methods
    - Reconstruction still feasible / tools still improving
    - Consistency analysis emerging as a valid scientific approach
    - But inadequate experimental support for many current claims
  - Social media:
    - $C \rightarrow^{M} E$ not even discussed today – some psychological research is promising, but highly speculative today
  - Massive sensor grid: Not invented yet

Identify
Collect
Preserve
Transport
Store
Analyze
Interpret
Attribute
Reconstruct
Present
Destroy

Examples

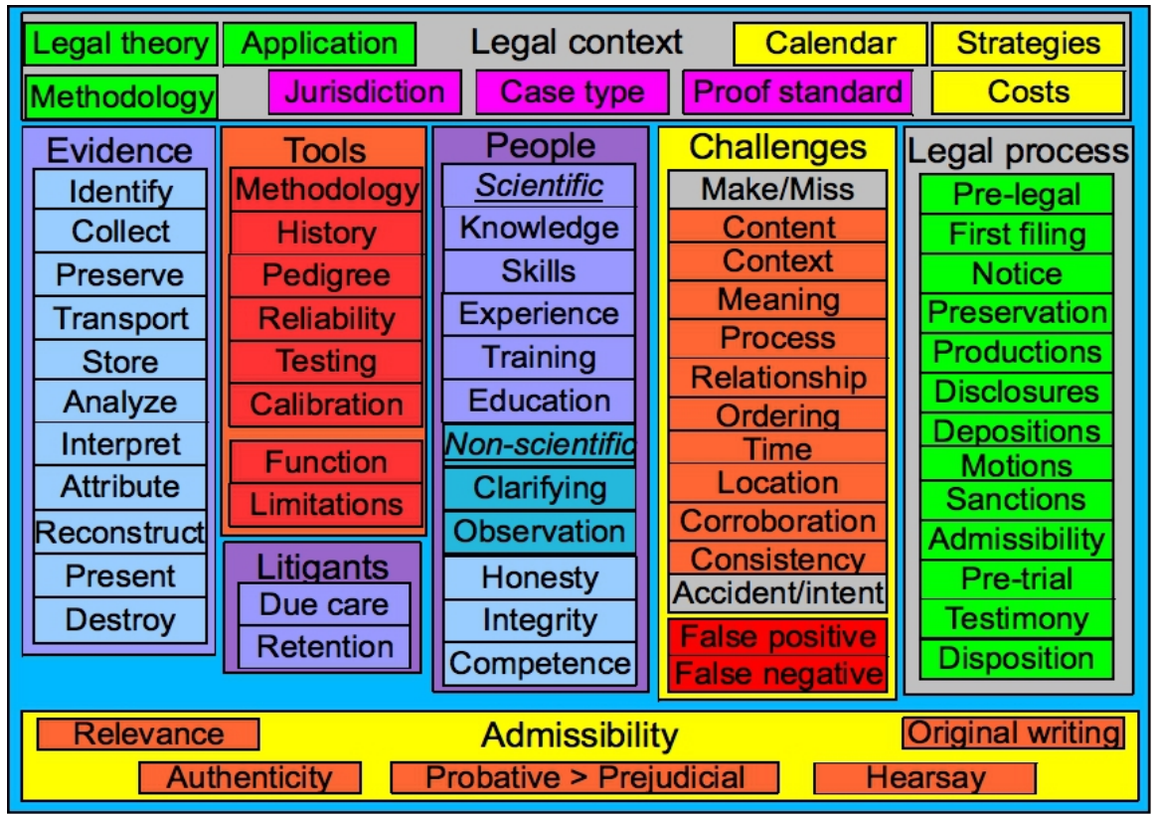
- Toll fraud circa 1970s
  - Only Joe and Mary were ever present at the known location
  - Joe had an alibi for 2 of the times→Mary is it. **(Convincing)**
- Internet investigations circa 2000
  - IP address often relate to locations (as in this case through subpoenas and evidence found on site). Related records are consistent with theory of the case and no alternative that is also consistent has been offered. **(less certain)**
- The Boston Marathon Bombers circa now (the CSI effect)
  - Video evidence validated by known methods / obvious forgeries detected by same methods / all evidence we show is consistent. Look at the movie **(Over-Convinced?)**
- Anonymous and beyond circa who knows when?

The science behind it

- As we move forward, there is less of a scientific basis
  - We have technology – lots of it – more and more of it
  - But is it becoming more prejudicial than probative?
    - **Boston bomber jump to false conclusions (NY Post)**
  - Much of the technology is not scientifically validated
    - Almost no commercial tools other than imaging
    - Many search tools have limitations not clear to examiners
  - We don't really have a science / consensus yet
    - **Studies have shown a lack of consensus over basics**
  - Group think and confirmation bias common
    - Generally, **refutation is <u>not</u> sought** – rather confirmation
    - Confirmation bias long-studied and readily demonstrable
    - **The Madrid bombing case and the Oregon attorney**

Outline

- Then, now, and then again
  - Political systems and approaches
  - Technological changes
  - Scientific basis for digital forensics
- Building our scientific future together
- Discussion



Science starts with agreed methodology

- We don't yet agree on scientific methodology
  - We need a consistent bent toward $C \rightarrow^{M} E$
    - Rather, almost always use $E \rightarrow C$ without the M
  - We need to agree on a testable theory of measurement
  - We need an experimental basis for claims
  - We need to use reconstruction to demonstrate consistency
  - We need more refutation tests and fewer confirmation tests
  - We need expert witnesses who are very well qualified
    - Medical: MD/equivalent + internship + specialty + board certified + licensed + experience are starting conditions for testimony
    - Digital: Few have MS, very few Ph.D., few internships, no well defined specialties, no real licenses or boards
  - We need specialized expertise to make specialized claims
    - What makes you an expert in email? Audit trails? Networks?

Scientific bases: C→^{M}E, measurement

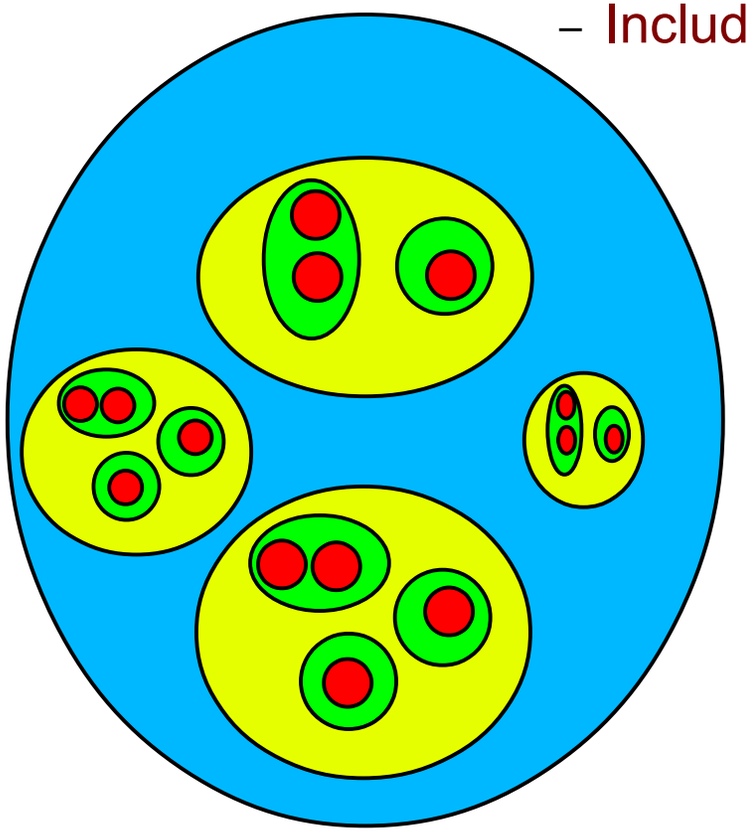
- Toll fraud circa 1970s
  - PBX access → creation of false access codes → calls charged to PBX owner
    - Records of dial-in for PBX access through audit mechanisms
    - Records of access code creation and subsequent use
    - Charge records for telco usage through audit mechanisms
- Internet investigations circa 2000
  - Datagrams coming from IP → access and exploitation → bad consequence
    - Records of datagrams at hops and destination via audit records
    - Records from DNS and ISPs lead to legal process
    - Corresponding records at source implicate causality
- The Boston Marathon Bombers circa now
- Anonymous and beyond circa who knows when

Scientific bases: C→^{M}E, measurement

- Toll fraud circa 1970s
- Internet investigations circa 2000
- The Boston Marathon Bombers circa now
  - Humans go to locations → leave explosives → Explosions kill people
    - Many videos of acts on different devices from different sources
      - Include accurate time, GPS location, source, witness for each
    - Explosion recorded from many angles on same videos
    - Hospital, LE, fire, and many official records
- Anonymous and beyond circa who knows when
  - Actors undertake acts → event sequences transpire → Bad things happen
    - Sensors record undertaken acts – or are they just actors?
    - Event sequences recorded in many places – or are they?
    - Records show bad things happened – or did they?

Drill-down into causality

- Causality at each level shown by the next lower level
  - Humans go to locations → leave explosives → Explosions kill people
    - Many videos of acts on different devices from different sources
      - Include accurate time, GPS location, source, witness for each
        - GPS signals include time signals to authenticate time
          - Time in GPS system is kept by atomic clocks
            - Atomic clocks are maintained by an official body
              - Body bases its methodology on physics
                - Physics uses measured c →[m] E or electron orbits
                - Orbits demonstrated through statistical methods
        - Witness for each states they took this video where and when
          - Video mechanism associated with computer activities
            - Video processing HW/SW produces known characteristics
              - Characteristics present in actual video consistent with mechanism
                - …
            - Storage uses OS software and storage HW
              - OS is designed to accurately record signals from input
                - Records consistent with normal OS operation
                - Surrounding records are consistent with the records at issue
                - Mechanisms tested after-the-fact to verify proper operation
              - Storage is based on specific storage media and formats
                - Storage formats consistent with media and mechanism
                - Content checksums correctly authenticate images
                - Formats include timestamps consistent with asserted times
              - …
            - …



Many recursive causal mechanisms may have to be defined, tested, and understood to get to overall conclusions of digital forensics.

Some challenges

- Capacity to create false records will be vast
  - Simulated event sequences to the fidelity of the sensors
    - Fake videos that are very realistic
    - Projected images of faces, movement, etc. for sensors
    - Injection post-sensor pre-recording
    - Alteration post-recording
- Chain of custody non-existent or out of control
  - Any or all of the "original" data may be fake
  - Data too massive for police to store and/or process
  - Selected subsets may fail to prove or miss exculpatory
    - More and more sources may emerge over time
    - Explanations for massive missing data may be complex
- Analytical capacity overwhelmed by quantity

Future crime

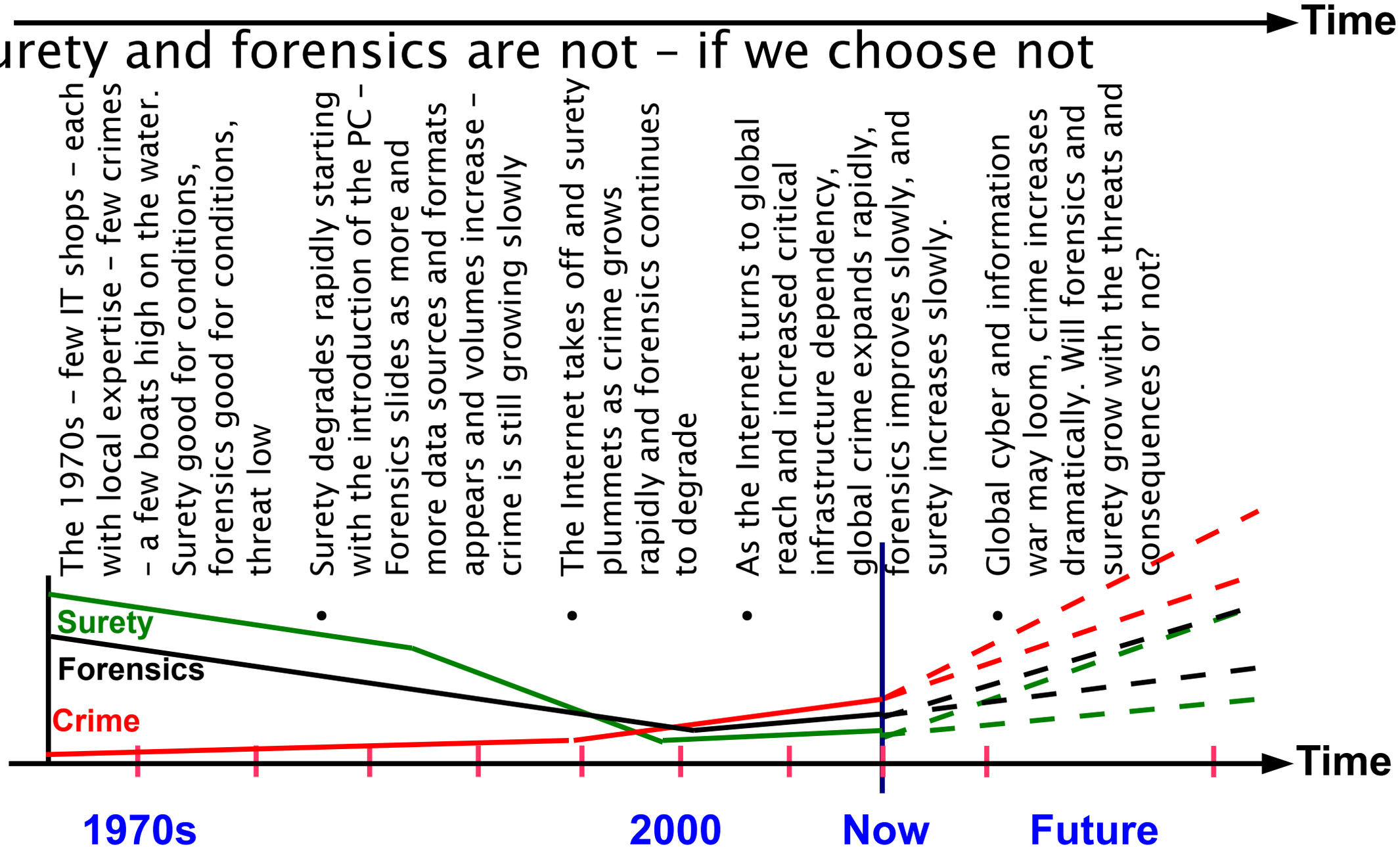
- Some examples of future crimes we may see
  - Attack on critical infrastructure used to disable sensors / transport / storage and/or misdirect forces away from real crime scheme where criminals undertake bad acts
  - Break-in to computer systems used to alter DNA for gene treatment and kill people who were supposed to be cured
  - Flash mobs used to wreck havoc on a competitor at key times, distracting them from a critical function
  - Trans-border infiltration into companies to get advanced information on pending deals and leak it for advantage
  - Mass hallucination by creating simulated attack on White House to take advantage of rapid market swings
  - Intellectual property theft on unprecedented scales
- Oops… these are present-day crimes… It will only get worse

Possible solutions and resourcing

- (1) Enough storage, computing, and sensors to do the job
  - The costs can be enormous – and it is not always needed
  - Or you can outsource as needed to cloud providers
    - But how do you assure reliable evidence/processing?
- (2) Unification of Law Enforcement and National Security
  - A very dangerous precedent when rich and powerful can control all of the levers of power over the lives of others
- (3) Raising all boats – bring the cloud up to legal standards
  - Turn cloud providers into high surety operations
  - Turn sensors into high surety devices
  - Create high integrity and confidentiality transport
  - Allow LE to use commercial providers with constraints

Raising all boats?

- Threats and consequences are out of our control
- Surety and forensics are not – if we choose not



Raising all boats

- The laws and judiciary need to be well informed
  - Substantial educational challenges face society
  - The information has to be balanced and scientific
    - Theory and methodologies have to be made clear
    - Metrics and their application have to be valid
    - Limits on results have to be clearly stated
    - Statements have to be in the diplomatics language
    - Formalisms of process and procedure must be defined
  - If you go a bridge too far, you do more harm than good
- Law enforcement and experts have to know how and why
- Providers have to meet standards of practice

Raising all boats

- The laws and judiciary need to be well informed
- Law enforcement and experts have to know how and why
  - Substantial educational challenges face law enforcement
  - Methods have to be systematized and specialized
    - Properly trained specialists produce consistent results
    - Consistency demonstrated to within defined limits
    - Common measurement methods applied uniformly
    - Statements and usage in common diplomatics language
    - Standardized formalized documentation methods needed
  - Careless use of terminology or slang destroys credibility
  - Standardization within and across departments help
- Providers have to meet standards of practice

Raising all boats

- The laws and judiciary need to be well informed
- Law enforcement and experts have to know how and why
- Providers have to meet standards of practice
  - Standards of practice should be defined as a requirement
    - Common standards of practice exist in the large
      - http://all.net/SecDec/index.html as an example
    - Augmented for LE purposes in the Cloud
      - A relatively small effort to start
    - Gain global consensus for minimum acceptable practices
    - LE and governments require these for their uses
  - The net effect will be differentiated services
    - Support for higher surety with a built-in market
    - The price will justify the value for other users

Outline

- Then, now, and then again
  - Political systems and approaches
  - Technological changes
  - Scientific basis for digital forensics
- Building our scientific future together
- Discussion
  - How can we work together on a global basis to achieve it?
    - Digital Diplomatics conference http://www.cei.lmu.de/digdipl13/call-for-papers
    - Cloud security alliance - https://cloudsecurityalliance.org/
    - This conference and others like it
    - IFIP TC-11and others
  - Working together is a fundamental requirement – how?

Thank you

http://all.net/ - fc at all.net

Thank You