Building Public Trust Through Forensic Science and Crime Prevention

Dr. Joshua I. James
Digital Forensic Investigation Research Laboratory
Graduate School of Forensic Science
SoonChunHyang University, Asan, South Korea

Overview

- A Brief History of Digital Crime
- Digital [Forensic] Investigations
- Digital Forensic Science
- Digital Forensic Investigation Research
- Case studies: Science meets practice
- Digital Crime Prevention
  - Measurement
  - Finding Patterns and Relations
  - Predicting Digital Crime
- Digital Crime Education

whoami

- Joshua I. James
  - B.Sc. Network Security, PhD. Digital Forensic Investigation
  - Lecturer: Live Data Forensics, Digital Forensic Practice
  - DF Trainer for the Centre for Cybercrime Investigation, INTERPOL, UNODC, Soonchunhyang University
  - Researcher:
    - DigitalFIRE Laboratory
    - Irish Police
    - KNPU International Cybercrime Research Center
    - KU Digital Forensic Research Center
  - Developer, DF Automation and Intelligence tools

Digital Crime

- From 2000 to 2012 there has been an estimated 566.4% worldwide growth in Internet users [1]
  - 2.4 billion users
- Approximately a 3.4% increase in U.S. based complaints per year (IC3) [2]
- Digital crime *requests* seem to be dropping in Korea
  - Not sure that is a good thing! (scope)

Digital Crime Investigation

- Relatively new field
  - Basic eDiscovery conducted in the 60's and 70's
  - Law Enforcement started investigating computers more in early 80's
    - Not really "forensically sound" as we know it now
- 1st Digital Forensic Research Workshop in 2000
  - Attempted to define "Digital Forensic Science"

Not Just Digital Crime

- Nearly every NYC crime involves a cyber component [3]
- How many of you have a cell phone?
  - Texts?
  - Camera?
  - GPS?
- How many of you have a car navigation system?
- How many of you have a Facebook account?

Digital Investigations

- Digital Investigations can be fruitful in traditional crimes
  - Murder
  - Burglary
  - Drugs
- Digital Investigations are required in digital-only crimes
  - Hacking
  - Malware

Digital [Forensic] Investigations

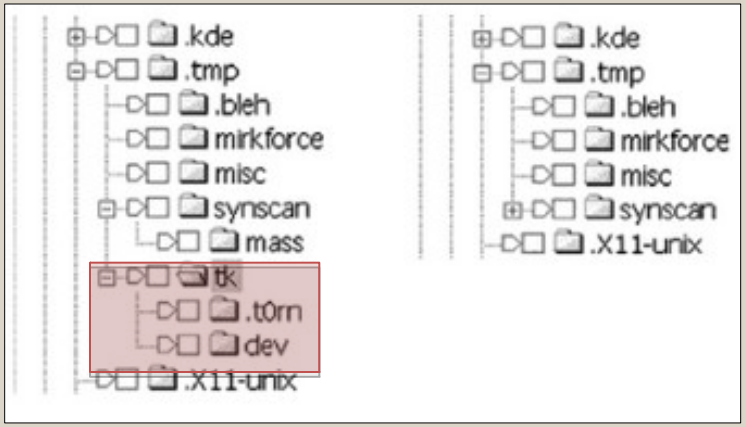
- **Digital Investigation**
  - process to answer questions about digital states and events [3]
- **Digital *Forensic* Investigation**
  - Special case of a digital investigation
  - Used procedures and techniques allow results to be entered in a **court of law**

Digital [Forensic] Investigations

- **Digital Forensic Investigation**
  - The collection, preservation, analysis, and presentation of computer-related evidence
  - All procedures and techniques must be “forensically sound” to be considered for admissibility in court

Digital Evidence

- **Digital Evidence** is data that supports or refutes a hypothesis that was formulated during an investigation [3]
- Digital evidence must be translated into a human-readable form [4]
  - Each layer of abstraction can introduce error or information loss
  - Result validation required!



Error in parsing a file system with two versions of the same program

What is “Forensically Sound”?

- “The application of a transparent digital forensic process that preserves the original meaning of the data for production in a court of law [5].”
- Derived evidence should be:
  - Reliable
  - Complete
  - Accurate
  - Able to be tested and verified

Evidence Dynamics

- Evidence dynamics is any influence that changes evidence, regardless of intent.
- Applies to digital evidence too!
- Some causes of evidence dynamics in digital investigations:
  - System administrators
  - Offender covering behavior
  - Victim actions
  - Secondary transfer
  - Witnesses
  - Nature/weather

Reliability of Evidence: Chain of Custody

- Chain of custody ensures an unbroken audit trail of seized exhibits to determine what was done, when and by whom
  - Who, when, where and how the exhibit was collected
  - Who, when, where and how the exhibit was transported
  - Who took possession? When?
  - How was the exhibit stored and protected in storage?
  - Who took it out of storage? When? Why? What did they do with it?

Reliability of Evidence: Authentication of Digital Exhibits

- Must be able to show that any changes on the original have no effect on the evidence (data)
  - Tools in a live environment modify system state, but not user data
  - How do we know that user data is not modified? Experiment!
  - Text or images don’t appear at random

Reliability of Evidence: Authentication of Digital Exhibits

- **Forensic data acquisition** is making an *exact* copy of the suspect data
- After forensic acquisition, the data should not change
  - If so, you must be able to demonstrate why and how the data was changed
  - Can verify the data has not changed by using a cryptographic hash

Reliability of Evidence: Authentication

- Cryptographic hash
  - SHA, MD5, etc.
    - Relatively small, unique string of characters generated based on a given input
    - MD5 (file.text) = **053ef45186fff3b4461485b14a554c37**
    - Only exactly the same input can produce the same output
    - If the resulting hash of two files is the same, they contain exactly the same data
    - If even one bit is changed in the file, the hash will change!

Goal of an Investigation

- An investigation attempts to support or deny a a question posed to the investigator
  - Question: Was the computer used to download illegal images?
- An investigation should attempt to answer the question **and** look for evidence of all (reasonable) explanations!
  - Reasonable explanation: A virus downloaded the illegal images.

Digital Forensics in Criminal Investigation

- Usual specialty areas:
  - Computer Forensics
  - Cell Phone Forensics
  - Database Forensics
  - Network Forensics
- Combination specialties:
  - Cybercrime investigation
  - Malware analysis
  - Financial crime analysis

Digital Forensics in Civil Investigation

- Not normally as thorough as criminal investigations
- Usual specialty areas:
  - Computer Forensics
  - Cell Phone Forensics
  - Database Forensics
  - Network Forensics
- Combination specialties:
  - eDiscovery
  - Financial crime analysis
  - Auditing

Normal Cases in Criminal Investigations

- Ireland (similar in Europe/U.S.):
  - Child Exploitation Material ~ 80% of time spent
  - Internet Investigation/fraud ~ 15% of time spent
  - Murder/hacking/kidnapping/drugs ~ 5% of time spent
- Korea
  - Appears to focus largely on hacking, DDoS and reputation defamation cases

UCD DUBLIN

Normal Cases in Civil Investigations

- Corporate Investigations:
  - eDiscovery
    - Keyword search
    - eMail/database search
    - Maybe financial inquiry (usually Audit department)
  - Return responsive data – very little analysis
  - Data Recovery
- Private Investigator
  - Investigation of cheating spouse
  - Second opinion in criminal case

Digital Forensics in Military

- Military and Intelligence rely heavily on digital forensic investigators
  - Operations involving technology
    - War zones
    - Data recovery/cracking
    - Spying
  - Internal investigations

Digital Forensics in Academia

- Areas of research:
  - Practical
    - Techniques/Forensic Programs
    - Law/Policy
  - Theoretical
    - Models
    - Philosophy

Digital Forensics in Academia

- There is a lot of practical work coming out of academia
  - Software programs/prototypes
  - Techniques and very technical applied work
- Problems:
  - Academics don’t always understand what practitioners need (not in the field)
  - Lack of consistency and long-term support

Digital Forensics in Academia

- There is a lot of theoretical work coming out of academia
  - Creating generic models to better understand digital crime
  - Considering what digital crime is
    - What is “cybercrime”?
    - What is “cyber war”?
    - How do you measure digital crime?

Digital Forensics in Academia

- Current problems with theory:
  - Cannot always be applied
  - If theory can be applied, Law Enforcement is usually about 5 to 10 years behind Academia
- Solution?
  - Digital Investigators should strive to be more scientific
  - Scientists should strive to be more applied

Digital Forensic Science

- Forensic science is based in the natural sciences: chemistry, physics, biology, etc.
- Digital Forensics should also be based on sciences: computer science, physics, etc.
  - Digital Forensics should involve the *systematic* study of the structures and behaviors of digital crime and how it affects physical reality
  - Should lead to more **objective** investigation (evidence based)

The Scientific Method

- Examiners are (should be) **neutral** finders of fact
  - Bias from personal beliefs
    - Very emotional case (child exploitation)
    - Influence from the media?
  - Bias from cultural beliefs
    - Westerners cannot eat very spicy food

The Scientific Method

- Scientific method
  - Standard procedure for developing a theory
    - helps increase objectivity
    - helps reduce bias

Scientific Method (simplified) [6]

1. Ask a question
2. Do background research
3. Construct a hypothesis
4. Test the hypothesis
5. Analyze data
6. Make conclusions
7. Present results

1. Ask a Question

- What is the investigating member trying to prove, exactly?
- What questions will the defense likely ask?

2. Do Background Research

- What type of case is it?
- What is the profile of the suspect?
- What information or data is available?
  - Forensic disk image?
  - Mobile device?
- What information are you likely to need to answer the questions posed by the investigating member?

3. Construct Hypothesis

- Hypothesis is driven by the research question
  - Question: “Was the computer used by a human to download illegal images?”
  - Hypothesis 1: “A web browser was used by a human to download illegal images.”
  - Hypothesis 2: “BitTorrent was used by a human to download illegal images.”
  - Hypothesis 3 (defensive): “A virus downloaded illegal images”
  - …

Recommended Reading: Carrier, Brian D. *A Hypothesis-based approach to digital forensic investigations.* [7]

4. Test Hypothesis

- For each hypothesis, experiment:
  - In similar system, simulate the same action
  - What traces are created in the system?
    - Hypothesis 1: Possible traces created in Temporary Internet Files
    - Hypothesis 2: BitTorrent client installed
    - Hypothesis 3: Traces of a virus on the system
- Read published articles / academic research papers

5. Analyze Data

- Analyze available data
  - Normally a forensic image of a suspect device
- Look for traces identified during the test phase
- Example:
  - Hypothesis 3: No virus found after scanning with several commercial virus scanners
  - Hypothesis 2: No active or deleted trace of BitTorrent client found on system
  - Hypothesis 1: Suspicious URLs found in IE history, suspicious URLs found in Windows Registry TypedURLs MRU list

6. Draw Conclusions

- What conclusions can we make?
- No evidence to support hypothesis 3 (virus)
  - Does that mean there was no virus?
    - NO! Just very unlikely!!
    - No evidence to support that the system was infected by a virus

6. Draw Conclusions (cont.)

- Some evidence to support hypothesis 1 (browser)
  - Does that mean a user used IE to download illegal images?
    - NO! Just very likely!!
    - Some evidence to support that Internet Explorer was used by a human to download suspected illegal images
- Second problem: <u>who</u> downloaded the images?
    - How to associate a human with the action

6. Draw Conclusions (cont.)

- No conclusions 100% definitely happened
- Found evidence increases or decreases the **probability** of a hypothesis
  - The goal is to derive enough evidence to prove a hypothesis *beyond a reasonable doubt*

7. Present Results

- Answer the initial question as clearly as possible
  - "Was the computer used by a human to download illegal images?"
    - We cannot say "the computer was definitely used by person X to download illegal images"
    - All we can say is, "The evidence suggests that a human used Internet Explorer to download suspected illegal images."

7. Present Results

- Never say a specific user was at the keyboard!
- Never make a claim that is beyond your scope of expertise
  - For example:
    - Indecent images of children
    - Never say “illegal image of a child”
      - Are you a Pediatrician (child doctor)?
      - Can you differentiate between a 16 year old and an 18 year old?
    - “Suspected image of a minor”

Validation of Digital Forensic Triage and Preliminary Analysis

Case Study: Science Meets Practice

Case: Theoretical Work

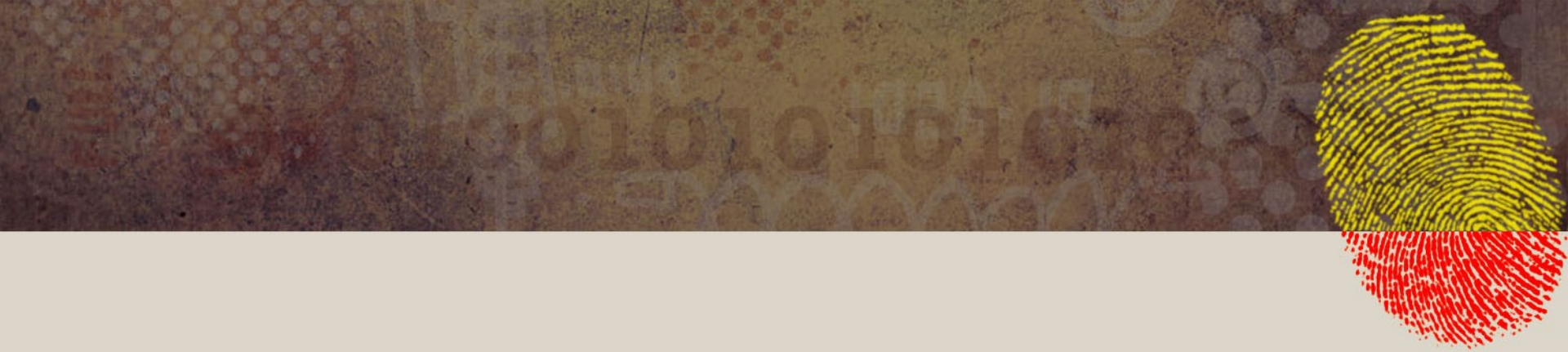
Signature-Based Detection of User Actions

- Locard's Exchange Principle: "with contact between two items, there will be an exchange"
- Locard's exchange principle also holds in the digital world
- With each event in a computer system, traces relating to the event are created

- Inferring user actions from trace observations:
  - If a user action causes a unique set of traces to be created, a signature can be created to detect the unique pattern of traces
  - A signature is equal to the knowledge of a system to be inferred (the user action)
  - A match of the signature is equal to observing the system

Signature-Based Detection of User Actions

- Individual traces have different update behaviors for the same user action
  - Some are always updated with every execution
  - Some are not always updated with every execution
- By examining trace update behaviors, signature categories can be created
  - Always updated traces allow for the last execution of the user action to be determined
  - Not always updated traces allow for multiple past executions of the user action to be determined
- From this, a model was created that generically applies to all digital devices



Some thoughts on

Crime Prevention

Understanding Crime

- To prevent crime we need a better understanding the crime
  - What motivates the crime?
  - What variables effect the crime?
  - How is this crime related to other crime?
- Relations between variables (and the strength of those relations) can be learned using statistical methods
  - Requires a lot of data

Understanding Crime

- Once we have a better understanding of the crime, we can begin to create strategies that focus on the strongest relationships
- Holistic view:
  - Using statistical methods we can look at variables associated with many different types of crime
  - Strategies can be broad or focused depending on our needs/resources
- Broad = Law/Policy Specific = actionable

Prevention Strategies

- Should not only be about policing
  - Many crimes occur because of social problems that are out of the scope of Law Enforcement
  - But Law Enforcement has all the raw data!
- Transparency and Public Relations
  - Helping to build/direct organizations focused on the variable
  - LE should make more data available for analysis and criticism

Measurement

- Measurement of the crime is necessary
  - Understand what the crime looks like
  - Understand how prevention strategies change the crime
- Remember that crime is dynamic
  - Pick metrics that represent a generalized model of the crime
  - Pick metrics that can be measured over time

Measurement

- Bad metrics:
  - Number of cases reported
    - Assumption: Less cases reported = less crime
  - Number of cases closed by LE
    - Assumption: More cases closed = police effecting the crime
- What does this mean?
  - Better metrics are needed, and might be a combination of different measurements

Prediction

- Once we understand the crime and related factors (in a measurable way) we can begin to predict
- Prediction is difficult
  - Requires a lot of data
  - Requires a thorough understanding of the data
  - Requires a clear question

Prediction

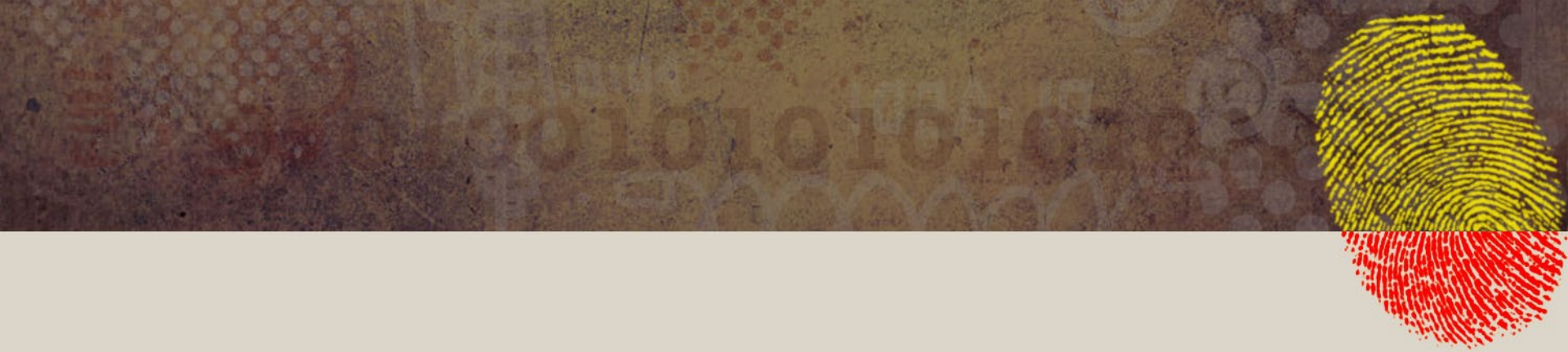
- Why is prediction useful?
  - At a high level, prediction can detect emerging patterns of crime before they become main-stream
  - At a low level, prediction can be used to determine when/where a particular crime is likely to take place

The Public

- What do the public know about digital crime prevention?
- Many of the crimes that happen today are made possible by the public
  - Phishing
  - Social Engineering
  - Malware
- Technology *can* be secure – people are a weakness

The Public

- The best crime prevention technique is EDUCATION
- Getting the public involved in securing their devices
  - Give the educational resources
    - Free online classes
    - Required tech security courses in school/university
    - Easy to understand!



Demonstration:

Understanding Crime Through Data Mining

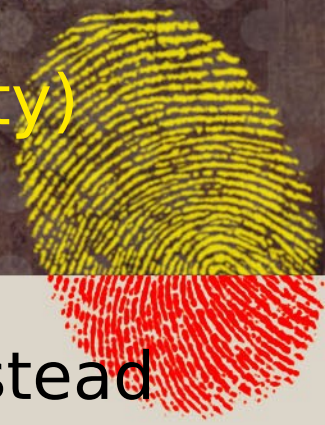
Data Mining Crime Data

- Data mining law enforcement case data can give insight into the crime and variables that affect the crime
- Law Enforcement has the best data to understand crime (but are not using it!)

Data Mining Crime Data

- Resources for learning data mining:
  - The R Project for Statistical Computing
    - http://www.r-project.org/
    - http://RStudio.org
  - Free Online Course
    - Coursera.org “Data Analysis”
  - Books
    - McCue, Colleen (2006) “Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis”. Elsevier.
    - Torgo, Luis (2011) “Data Mining with R”. Taylor & Francis Group.

Thoughts on Improving Prevention (Security) and Investigations



- Implementing policy based on evidence instead of gut feeling (research)
- Using the outputs of digital forensic investigations to create security policies
- Focusing more on past security research and how it affects us now
- Holding people accountable for their actions
- Thinking globally; cybercrime is not a country-specific problem
- **Education of everyone**

References

1. (2012, 30 June). "Internet Usage Statistics: The Internet Big Picture." Retrieved 27 Feb, 2013, from http://internetworldstats.com/stats.htm.
2. IC3 (2011). 2011 Internet Crime Report, Internet Crime Complaint Center.
3. http://www.theepochtimes.com/n2/united-states/nearly-every-nyc-crime-involves-cyber-says-manhattan-da-355692.html
4. Carrier, Brian D. (2006) Basic Digital Forensic Concepts. http://www.digital-evidence.org/di_basics.html
5. Casey, Eoghan. (2010) *Handbook of Digital Forensics and Investigation*. Elsevier Inc.
6. McKemmish, Rodney. (2008) *When is Digital Evidence Forensically Sound?* Advanced in Digital Forensics IV. Springer. http://link.springer.com/chapter/10.1007%2F978-0-387-84927-0_1?LI=true
7. Steps of the scientific method: http://www.sciencebuddies.org/science-fair-projects/project_scientific_method.shtml
8. Carrier, Brian D. (2006) *A Hypothesis-Based Approach to Digital Forensic Investigations. Purdue University.* https://www.cerias.purdue.edu/tools_and_resources/bibtex_archive/archive/2006-06.pdf
9. Vacca, John R. (2002) *Computer Forensics Computer Crime Scene Investigation.* Charles River Media, INC.
10. Kruse, Warren G., Jay G. Heiser. (2001) *Computer Forensics Incident Response Essentials*. Lucent Technologies.
11. Carrier, Brian D. (2002) *Open Source Digital Forensics Tools: The Legal Argument*. www.digital-evidence.org/papers/opensrc_legal.pdf