Digital Forensics: Research Challenges and Open Problems

Dr. Yong Guan
Litton Assistant Professor
Department of Electrical and Computer Engineering
& Information Assurance Center
Iowa State University

December 4, 2007
Why & What for Today’s Talk

- Digital Forensics
  - What is it about?
  - Where have you learned about it first?
    - CSI
    - CourtTV
    - Front-page news, Patricia Cornwell novels, Sir Arthur Conan Doyle, …
    - FBI catches botmasters: Bot Roast II (Nov. 29, 2007)
  - Why are the problems unique and challenging?

Goal of this Talk:
- Understand Digital Forensics: Status and On-going Efforts
- Discuss research challenges & open problems
- Motivate young students and researchers to work in this exciting field
- Seek your insights and vision
Outline for Today’s Talk

- Motivation, Goals, and History
- Digital Evidence: Principles and Procedures
- Current Status and Efforts
- Research Challenges & Open Problems
Cyber Crime  A painful side-effect of the innovations of Computer and Internet technologies

- The possibility of becoming a victim of cyber crime is the number one fear of billions of people.

- 1995-2007 CSI/FBI Computer Crime and Security Surveys confirm that cyber crime is real and continues to be significant problems, and cause ruinous financial damage.
  - Low percentage of cyber crime cases reported to law enforcement. (Only 16% in 1996; and 25% in 2006)

- Total losses for 2006 were $52,494,290 for the 313 respondents that were willing and able to estimate losses.

- In 2007, average annual loss more than doubled (from $168,000 in 2006 to $350,424 this year)
A Few Concerns

- Massive base of installed infrastructure with little support for security

- Paradigm Shift of Incident Response
  - Priority: identifying the cause of problem vs. Fix the problem

- Legal Influence
  - Evidence presented, examined, and challenged by the jury and the judges in the courtroom

- Digital forensics is a young maturing field
  - Need many more forensically sound solutions
A Few Concerns

What are we seeing?

- Increased use of encryption
  - Child pornography case
- Increased numbers and complexity
  - ISP, device diversity
- Increased size & urgency
  - 1.8TB in 4 hours
Concerns => Problems

- Concern of negative publicity
  - Attracting other cyber-attackers
  - Undermining the confidence of their customers, suppliers, and investors
  - Inviting the ridicule of their competitors

- Growing sophistication and stealth of cyber criminal activities/attacks
  - Much harder to detect than crimes in the physical world
  - Often insiders and international involvement
  - Indirectly through various hiding techniques
    - Botnets
    - Information hiding: steganography, covert channel, etc
    - Anonymity proxies
    - Stepping Stones
    - …

- Vast majority of cyber crimes never get caught or prosecuted !!!
Motivation

- It’s clear that we need better (more secure) IT technology and training.
  - But that, by itself, is not enough!
    - Unexpected interactions and failures
    - Users with privileged access but with improper training or untoward motives

- We need reliable tools and methods for investigation when an untoward event occurs.
  - Fix collateral damage
  - Identify the causes
  - Prosecute the responsible person

- We need to find who is the source of the attack and where they are
Context of Digital Forensics

- Homeland Security
- Information Security
- Corporate Espionage
- White Collar Crime
- Child Pornography
- Traditional Crime
- Incident Response
- Employee Monitoring
- Privacy Issues
- ...
Digital Forensics

- Also known as
  - Cyberspace Forensics (or Cyber Forensics)
  - Computer and Network Forensics

- Can be defined as the art of discovering, retrieval of information about a crime in such a way to make it admissible to the court.

- Science?
Digital Forensic Science

Digital Forensic Science (DFS):

“The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations.”

Source: (2001). Digital Forensic Research Workshop (DFRWS)
Digital Forensics

Legal Requirements

Policies/Processes

Education/Training

Research/Scholarship

Technologies/Tools
Legal Requirements

- Status as scientific evidence?
  - Frye Test (1923)
    - General acceptance in the field
  - Federal Rules of Evidence
    - Rule 702
      - Scientific knowledge, assisting trier of fact
  - Daubert criteria (1993) (*Daubert v. Merrell*)
    - Whether the theory or technique has been reliably tested;
    - Whether the theory or technique has been subject to peer review and publication;
    - What is the known or potential rate of error of the method used; and
    - Whether the theory or method has been generally accepted by the scientific community.
  - Kumho Tire extended the criteria to technical knowledge
Current Situation

- Communities with overlapping but somewhat different objectives

<table>
<thead>
<tr>
<th>Area</th>
<th>Primary Objective</th>
<th>Secondary Objective</th>
<th>Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Law Enforcement</td>
<td>Prosecution</td>
<td></td>
<td>After the fact</td>
</tr>
<tr>
<td>Military IV Operations</td>
<td>Continuity of Operations</td>
<td>Prosecution</td>
<td>Real Time</td>
</tr>
<tr>
<td>Business &amp; Industry</td>
<td>Availability of Service</td>
<td>Prosecution</td>
<td>Real Time</td>
</tr>
</tbody>
</table>

- Non-unified, ad hoc, fragmented approach
- Vendor driven technologies
- Backdrop of a constantly changing technological environment
- Increased judicial scrutiny
**Historical Context**

- **Predominantly Law Enforcement Domain**
  - 1984 FBI Computer Analysis and Response Team (CART)
  - 1991 International Law Enforcement meeting to discuss computer forensics & the need for standardized approach
  - 1997 Scientific Working Group on Digital Evidence (SWGDE) established to develop standards
  - 2001 Digital Forensic Research Workshop (DFRWS) development of research roadmap
  - 2003 American Society of Crime Laboratory Directors – Laboratory Accreditation Board (ASCLD-LAB) recognized digital evidence as a full fledged forensic discipline

- **Reactive in nature**
  - Criminal Investigations
    - Fraud
    - White Collar Crime
  - More and more Academia, industry R&D, Private Sectors involvement
Major Initiatives

- National White Collar Crime Center (NW3C)
- National White Collar Crime Research Consortium (WCCRC)
- NIST
  - Computer Forensics Tool Testing project - CFTT
  - National Software Reference Library - NSRL
- NIJ Electronic Crime Program
- Office of Law Enforcement and Technology Commercialization (OLETC)
- Lab Accreditation
- DoD Cyber Crime Center
- Scientific Working Group on Digital Evidence (SWGDE)
- National Center for Forensic Sciences (NCFS)
- Digital Forensics Research Workshop (DFRW)
- Computer Forensic Educator’s Working Group (CFEWG)
- Cyber Tools Online Search for Evidence (CTOSE) - European
- International Organization on Computer Evidence (IOCE)
- SADFE
- IFIP 119 WG Digital Forensics Conference
- Anti-Phishing Working Group (APWG)
Major Forums

- IFIP 11.9 WG International Conference on Digital Forensics
- Digital Forensics Research Workshop (DFRWS)
- Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE)
- DoD Cyber Crime Conference
  - Sponsored by DoD Cyber Crime Center & JTF-GNO
  - DC3 Challenge (Annual Competition)
Forensic Principles

“Rules of Evidence: A Quick Review”
Edmond Locard (1877-1966): Exchange Theory of Transference

- Pioneer French criminalist – 1928-30
- Recognized and documented that:
  - When two objects come into contact, there is an exchange of material
  - **Principle of Exchange**
    - “...when a person commits a crime something is always left at the scene of the crime that was not present when the person arrived.”
- Leads to the notion that a suspect cannot enter a space without leaving trace evidence.
- The purpose of an investigation is to locate identify and preserve evidence-data on which a judgment or conclusion can be based.
5 Rules of Evidence

- **Admissible**
  - Must be able to be used in court or elsewhere

- **Authentic**
  - Evidence relates to incident in relevant way

- **Complete (no tunnel vision)**
  - Exculpatory evidence for alternative suspects

- **Reliable**
  - No question about authenticity & veracity

- **Believable**
  - Clear, easy to understand, and believable by a jury
A New View on Forensic Principles

- **Object-oriented Technology**
  - Language: C++, Java
  - Object and its associated methods
    - Object has state (i.e., value)
    - The state of the object can be changed due to some event.
    - The state of the object can only be accessed through its associated methods

- **Digital Evidence (object) and its associated investigation process (methods)**
  - Integrity
  - Authentication
  - Reproducibility
  - Non-interference
  - Minimization

<table>
<thead>
<tr>
<th>Digital Evidence</th>
<th>Digital Investigation Process (Collection &amp; identification, Storage, preservation, and transportation, Presentation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desired Properties: Integrity, authenticity, complete, reliable, non-repudiation, etc</td>
<td>Desired Properties: Reproducibility, Real-time, Confidence interval, reliable, believable, etc</td>
</tr>
</tbody>
</table>
Digital Forensics Process

Search & Seizures “Electronic Data Recovery/Bag & Tag”
Incident Response Methodology (PDCAERF)

Preparation → Detection → Containment → Analysis → Eradication → Recovery → Follow-up → Feed Back

Source: (2004). Marc Rogers, Purdue University
Disk Imaging and Analysis

Disk Imaging

- In many aspects, the general digital evidence collection issue has been **satisfactorily** addressed.
  - Encase
  - ILook
  - Safeback
  - Many others.

Remaining Challenges on File System Analysis

- Increasing usage of encryption
- New storage media with larger capacity
- Lack of automated tools
- Lack of tool testing & validation
**Live System Memory Imaging**

- In some cases, evidence may only be found in main memory.
  - E.g., child pornography cases

- Read from the source and write to a non-volatile destination, called **image Imaging**

- **Solutions**
  - Software-based approaches
  - Hardware-based approaches
Research Challenges & Open Problems
Why Are Digital Forensics Problems Challenging?

- Technologies/Tools
- Policies/Processes
- Legal Requirements
- Research/Scholarship
- Education/Training
Challenges

- Device Diversity
- Volume of Evidence
- Distributed Evidence
- Trust of Audit Trails
- Testing and Validation
- Anti-forensics

- ... You can name a lot
Challenge #1: Device Diversity

- Traditional storage devices: Simple data and image files
- We are seeing
  - Video, audio, GIS materials, VoIP systems, sensor net data, SCADA systems, etc.
  - Increasing usage of USB thumb drive, iPod, cell phone/PDA, digital camera, remote storage devices, removable media
  - Long-term storage in appliances and home media blur the notion of “local storage”
  - Peer-to-peer file sharing
  - Data outsourcing: Google Docs, Yahoo Photo Album, and many others
Challenge #2: Image Large, Active Disk Farms

- How to image large, active disk farms dynamically?

  - Imagine asking amazon.com or ebay to discontinue service while the drives are being copied
Challenge #3: Anti-Forensics

- Encryption
  - Encrypted files & Whole drive encryption (EFS)
- Steganography and other information hiding
- Evidence elimination tools

### Table 4 - Categories of Data Hiding

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Least Significant Bit</td>
<td>Compressed Images</td>
<td>Compound Doc formats</td>
<td>Marked Bad Clusters</td>
<td>Distributed Systems</td>
<td>Reserved Packet Offsets</td>
<td>BIOS</td>
<td>Heap Space</td>
<td>Virus-like expression</td>
<td>Perception</td>
</tr>
<tr>
<td>Audio</td>
<td>Stego</td>
<td>metadata - reserved structures</td>
<td>Maintenance Tracks</td>
<td>RAM Stack</td>
<td>Email Spies</td>
<td>CMOS</td>
<td></td>
<td>Rootkits altering system calls</td>
<td>F憾names</td>
</tr>
<tr>
<td>Video</td>
<td>Stego</td>
<td>File Slack</td>
<td>Extra tracks</td>
<td>Modified Dir Entries</td>
<td>Protocols</td>
<td>RAM</td>
<td>System Libraries</td>
<td>Plain sight</td>
<td></td>
</tr>
<tr>
<td>Imagery</td>
<td>Stego</td>
<td>Hidden partitions</td>
<td>Unallocated Space</td>
<td>Boot Sector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Challenge #4: Trust of Audit Trails

How can we trust audit trails?

- Always possible that an intruder may edit or delete the audit trail on a computer, especially weakly-protected PC.

- Increasingly sophisticated rootkits that dynamically modify the kernels of running systems to hide what is happening, or even to produce false results.
Open Problems

There are various open hard problems. Here is just a list of samples:

- Forensic tool testing and validation
  - Open vs. Close Source
- Solutions against anti-forensics techniques
- Network attack attribution
  - Botmasters
  - Criminals using stepping stones or Tor
  - Anonymous VoIP threatening callers
- Fighting against online fraudsters
  - Click fraud
  - Auction frauds
  - Spammers
  - Phishing
- Insiders
- Digital right management related issues…….
- More and more cyber criminal activities are for profit!
OP #1: Fighting Against Anti-Forensics

- Various information hiding and evasion techniques
  - Steganography
  - Watermarking
  - Colluded attacks

- Various digital evidence elimination tools

- Anonymity and privacy protection techniques
OP #1: Fighting Against Anti-Forensics

- Tracing attackers using anonymity proxies/systems
  - Tor – Onion Routing
  - Anonymizer.com
  - .......

Fig. 2. Mix Network
OP #2: Tracking Botnets

- Tracing Botnet Controllers / Botmasters
  - General Issues
    - Command and Control: Indirect and hidden
    - Evasive Bots: “Low and Slow” (no spurious traffic)
  - Tracing botnet controller
  - Attributing botmasters
OP #4: Tracing Anonymous VoIP Calls

- VoIP Services: Vonage, AT&T, Skype, etc.

- General Issues in VoIP E911 Services
  - Caller-ID
  - Location Determination

- Design reliable and secure VoIP Caller-ID Services
  - Support Mobility
  - Location Determination

- Tracing Anonymous Calls
  - SIP-based, H323-based, Skype Systems
# OP #5: Fighting Against Online Frauds

## Sentinel Top Complaint Categories

*January 1 – December 31, 2006*

<table>
<thead>
<tr>
<th>Rank</th>
<th>Top Categories</th>
<th>Complaints</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Identity Theft</td>
<td>240,035</td>
<td>36%</td>
</tr>
<tr>
<td>2</td>
<td>Shop-at-Home/Catalog Sales</td>
<td>46,995</td>
<td>7%</td>
</tr>
<tr>
<td>3</td>
<td>Prizes/Sweepstakes and Lotteries</td>
<td>45,587</td>
<td>7%</td>
</tr>
<tr>
<td>4</td>
<td>Internet Services and Computer Complaints</td>
<td>41,243</td>
<td>6%</td>
</tr>
<tr>
<td>5</td>
<td>Internet Auctions</td>
<td>32,832</td>
<td>5%</td>
</tr>
<tr>
<td>6</td>
<td>Foreign Money Offers</td>
<td>20,411</td>
<td>3%</td>
</tr>
<tr>
<td>7</td>
<td>Advance-Fee Loans and Credit Protection/Repair</td>
<td>10,857</td>
<td>2%</td>
</tr>
<tr>
<td>8</td>
<td>Magazines and Buyers Clubs</td>
<td>8,924</td>
<td>1%</td>
</tr>
<tr>
<td>9</td>
<td>Telephone Services</td>
<td>8,165</td>
<td>1%</td>
</tr>
<tr>
<td>10</td>
<td>Health Care</td>
<td>7,467</td>
<td>1%</td>
</tr>
<tr>
<td>11</td>
<td>Business Opps and Work-at-Home Plans</td>
<td>7,460</td>
<td>1%</td>
</tr>
<tr>
<td>12</td>
<td>Travel, Vacations and Timeshare</td>
<td>6,712</td>
<td>1%</td>
</tr>
<tr>
<td>13</td>
<td>Office Supplies and Services</td>
<td>5,723</td>
<td>1%</td>
</tr>
<tr>
<td>14</td>
<td>Grants: Scholarships/Educational &amp; Non-Educational</td>
<td>5,310</td>
<td>1%</td>
</tr>
<tr>
<td>15</td>
<td>Employ Agencies/Job Counsel/Overseas Work</td>
<td>4,485</td>
<td>1%</td>
</tr>
<tr>
<td>16</td>
<td>Investments</td>
<td>3,630</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>Other Coded Complaints</td>
<td>12,399</td>
<td>2%</td>
</tr>
</tbody>
</table>

---

FTC: Consumer Fraud and Identity Theft Complaint Data (Feb. 2007)
## OP #5: Fighting Against Online Frauds

### Total Number of Internet-Related Fraud Complaints & Amount Paid
Calendar Years 2004 through 2006

<table>
<thead>
<tr>
<th>CY</th>
<th>Total No. of Complaints</th>
<th>Complaints Reporting Amount Paid</th>
<th>Percentage of Complaints Reporting Amount Paid</th>
<th>Amount Paid Reported</th>
<th>Average Amount Paid(^1)</th>
<th>Median Amount Paid(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>210,850</td>
<td>188,757</td>
<td>90%</td>
<td>$271,345,207</td>
<td>$1,438</td>
<td>$215</td>
</tr>
<tr>
<td>2005</td>
<td>197,084</td>
<td>160,574</td>
<td>81%</td>
<td>$336,559,907</td>
<td>$2,096</td>
<td>$342</td>
</tr>
<tr>
<td>2006</td>
<td>204,881</td>
<td>176,847</td>
<td>86%</td>
<td>$590,310,461</td>
<td>$3,338</td>
<td>$500</td>
</tr>
</tbody>
</table>

\(^1\)Average is based on the total number of consumers who reported amount paid for each calendar year: CY-2004 = 188,757; CY-2005 = 160,574; and CY-2006 = 176,847. Eighty-five consumers reported an amount paid of $1 million or more during CY-2006; 15 and 24 consumers for CY-2004 and CY-2005, respectively.

\(^2\)Median is the middle number in a set of numbers so that half the numbers have values that are greater than the median and half have values that are less. Calculation of the median excludes complaints with amount paid reported as $0.

FTC: Consumer Fraud and Identity Theft Complaint Data (Feb. 2007)
OP #5: Fighting Against Online Frauds

- Click Fraud Problem
  - US Online Advertising Market Grows fast.
  - In the online ad market, 14.6% are paid to Click Fraud.

- Auction Fraud: Billion-dollar worldwide market
  - eBay attracts more and more users, and ... Fraudsters, too.
    - Current feedback systems cannot guarantee a user with high scores are trustable.
    - Fraudsters can build excellent feedback promoted by a clique of accomplices

- Phishing
Summary

- Digital forensics is a young and maturing field
  - Many interesting and challenging problems

- Hope you have enjoyed the talk and it could help to:
  - understand more about digital Forensics: on-going efforts, research challenges & open problems
  - attract more young students and researchers/practitioners to work in this exciting field

- Share your insights and vision with us and help to shape the research agenda

- You are very welcome to participate in 2008 IEEE Symposium on Security and Privacy, the top conference in security.

  - A digital forensics workshop will be there!
Thanks and Questions

Yong Guan
Department of Electrical and Computer Engineering
& Information Assurance Center
Iowa State University
(515) 294-8378
http://www.eng.iastate.edu/~guan