Preparing for Large-Scale Investigations with Case Domain Modeling

By
Chris Bogen and David Dampier

Presented At
The Digital Forensic Research Conference
DFRWS 2005 USA  New Orleans, LA (Aug 17th - 19th)

DFRWS is dedicated to the sharing of knowledge and ideas about digital forensics research. Ever since it organized the first open workshop devoted to digital forensics in 2001, DFRWS continues to bring academics and practitioners together in an informal environment. As a non-profit, volunteer organization, DFRWS sponsors technical working groups, annual conferences and challenges to help drive the direction of research and development.

http://dfrws.org
Preparing for Large-Scale Investigations with Case Domain Modeling

A. Chris Bogen, M.S.
United States Army Corps of Engineers
Engineering Research & Development Center
Information Technology Lab
Vicksburg, MS

David A. Dampier, PhD
Assistant Professor
Department of Computer Science and Engineering
Mississippi State University
Outline

1) Background
2) Case Domain Modeling
3) Applications of Case Domain Modeling
4) Conclusions and Future Work
Digital Forensics Backgrounds

- Software Engineering Practice & Research Backgrounds
  - Dampier – Retired Army Officer, Software Engineer, now Asst. Prof. @ MSU CSE Department
  - Bogen – USACE Software Engineer, PhD Candidate w/ Forensics Focus, M.S. w/ Software Engineering Focus

- Computer Forensics Research & Instruction
  - MSU Center for Computer Security Research
    - http://security.cse.msstate.edu
  - MSU Forensics Training Center
    - http://security.cse.msstate.edu/ftc
  - NSA Center of Academic Excellence in Security since 2001

- Limited Computer Forensics Practice
  - Dampier – Consulting
  - Bogen – Brief Internship at MSAGO Cyber Crime Center

- Interested in CF Analytical & Modeling Methodologies
Software Engineering and Computer Forensics Similarities

- Common Underlying Philosophy
  - Quality Focus
  - Repeatable Processes
  - Application of Scientific Methods
  - Application & Development of Tool Support
- Existing Modeling Work in CF Suggests Similarities to SWE
  - Process Models
    - Baryamureeba & Tushabe [1], Bebe & Clark [2], Carrier & Spaford [7], Palmer [10]
  - Formal Methods
    - Carney & Rogers [6], Gladshav [8], Stephenson [12,13]
  - Patterns & Knowledge Reuse
    - Bruschi & Monga [5]
Analytical Challenges Encountered on Large Cases

- Several People, Places, Organizations
- Abundance of Digital Media
  - e.g. 30 Workstations & Servers
- Goals of Forensic Activities are Uncertain
  - What Are We Looking For?
  - How do We Characterize the Evidence?
- Unfamiliar Case Domain
  - Jargon
  - Technology
  - Business Process
Problem Focus

- Filtering Relevant Case Information
- Representing/Managing Forensic Case Data
- Knowledge Reuse
- Facilitating Investigator/Technician Communication
- Practical Analytical Methodologies/Framework
- We Propose an Adaptation of SWE Domain Analysis/Modeling to Address these Issues
Introduction to SWE Domain Analysis

- Originated from Artificial Intelligence, Knowledge Engineering, Ontology Development
- Performed in Early Requirements Phase of Object-Oriented Development
- Problem Domain is Populated by Specialized Knowledge
  - People, Places, Things, Policies, Processes, Science, etc.
- Goals:
  - Identify Sources of Domain Knowledge
  - Facilitate Knowledge Reuse & Communication
  - Filter the Relevant Domain Knowledge
  - Reach a Shared Understanding of Problem Domain
  - Contribute to a Quality set of Requirements & a Development Plan
Case Domain Modeling

• Golden Rule:
  – *If it is not relevant to the examination then don’t model it*
    • Use Process with Heuristics to Determine Relevance

• Analytical/Modeling Process (Adapted from Larman)
  1. Select Case Concepts
  2. Select Concept Relationships
  3. Identify Concept Attributes
  4. Instantiate the Model
     (Steps 1-3 May Occur Concurrently)

• UML Used as Example Representation
  – Currently We are Focusing on Analytical Framework, Not Representation
Identifying Case Concepts

- Brainstorm and Generate a Complete Concept List
  - Gradually Eliminate Irrelevant Concepts
- Select Reusable Concepts that Balance Between Generalization & Specialization
  - Concept Name: *Patrick Bateman* (worst)
    - Too Specialized, better to have a name attribute
  - Concept Name: *Person* (better)
    - Too general if there are lots of people with different roles
  - Concept Name: *Suspect* (best)
    - Reusable as a specialized role or type of person
# Case Concept Tools: Concept Category Table

<table>
<thead>
<tr>
<th>Concept Category</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical or tangible objects</td>
<td>Cell phone, Hard Drive, CDR disk</td>
</tr>
<tr>
<td>Descriptions of things</td>
<td>Marketing Report, Incident Report</td>
</tr>
<tr>
<td>Places</td>
<td>Home, Street</td>
</tr>
<tr>
<td>Transactions</td>
<td>Payment, Sale, Money Deposit, Email Transmission</td>
</tr>
<tr>
<td>Roles of people</td>
<td>Victim, Suspect, Witness</td>
</tr>
<tr>
<td>Containers of things</td>
<td>Databases, Hard Drives</td>
</tr>
<tr>
<td>Things in a container</td>
<td>Files, Transactions</td>
</tr>
<tr>
<td>Computer or Electro-mechanical systems</td>
<td>Internet Store, Credit Card Authorization System</td>
</tr>
<tr>
<td>Abstract noun concepts</td>
<td>Motive, Alibi, Insanity, Poverty</td>
</tr>
<tr>
<td>Organizations</td>
<td>Mafia, Corporate Department, Government Organization</td>
</tr>
<tr>
<td>Events</td>
<td>Robbery, Meeting, Phone Call, File Access</td>
</tr>
<tr>
<td>Rules and policies</td>
<td>Laws, Procedures</td>
</tr>
<tr>
<td>Records of finance, work, contracts, legal matters</td>
<td>Employment Contract, Lease, Receipt, Subpoena</td>
</tr>
<tr>
<td>Services</td>
<td>Internet Service Provider, Telephone Service, Cell Phone Service</td>
</tr>
</tbody>
</table>
Case Concept Tools: Noun Extraction

Woman charged for heroin possession

State police arrested Edna Krabapple, 38, of Springfield, after she was treated for overdosing on illegal drugs. Chief Wigham, state police spokesman, said troopers were dispatched to a Homer Street residence in Springfield shortly before 8 p.m., Aug. 3, to assist emergency medical workers with a patient who was disorderly. While en route, said Wigham, an ambulance driver called the dispatch center and said the patient had calmed down, so the trooper did not need to go to the residence.

The trooper went to the hospital to check on her condition, at which point he learned Krabapple had overdosed and her purse contained illegal substances. Police found a total of 27 packages of what later field tested positive as heroin. There were two groups of 12 and 13 packages respectively, that were banded together, and two packages that were loose. Additionally, said Wigham, there was an unlabeled bottle of pills and a glass pipe in the purse. There were 45 Soma pills and one methadone pill, he said. Soma is a drug prescribed for acute, painful muscle strains and spasms. Methadone is a medication used to treat narcotic withdrawal and dependence. Krabapple was charged with possession with intent to deliver heroin, possession of drug paraphernalia, maintaining a dwelling for keeping controlled substances and drugs not in their original container. Krabapple is also suspected of being involved in an Internet-based drug distribution network. She was released to the custody of relatives on $6,000 unsecured bond.

Adapted From http://www.capecazette.com/pages/policrep.html
## Case Concept Tools: USDOJ Manual

<table>
<thead>
<tr>
<th>Case Type</th>
<th>Relevant Information Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email Threats/ Harassment / Stalking</td>
<td>Address books, diaries, e-mail/notes/letters, internet activity logs, legal documents, telephone records, financial/asset records, victim background research, images</td>
</tr>
<tr>
<td>Extortion</td>
<td>Date and time stamps, e-mail/notes/letters, history log, internet activity log, temporary internet files, user names</td>
</tr>
</tbody>
</table>
Identifying Concept Relationships

• Not as Important as Concepts & Attributes
  – But Can Reinforce Understanding
  – Especially when We Are Interested in Relationships Between People & Organizations

• Don’t Try to Include Every Relationship
  – Too Many Relationships Obscure Domain Model
  – Scalability Becomes an Issue when Illustrating the Domain Model
  – Include Essential Relationships that Reinforce Understanding
# Concept Relationship Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>A is a physical part of B</td>
<td>DVD Drive – Workstation</td>
</tr>
<tr>
<td>A is a logical part of B</td>
<td>Network Mapping – Network Intrusion</td>
</tr>
<tr>
<td>A is physically contained in/on B</td>
<td>Used CDR Media – CD Case</td>
</tr>
<tr>
<td>A is a description for B</td>
<td>Readme file – Executable Program</td>
</tr>
<tr>
<td>A owns B</td>
<td>Suspect – Vehicle</td>
</tr>
<tr>
<td>A is a member of B</td>
<td>Suspect – Gang</td>
</tr>
<tr>
<td>A is an organizational subunit of B</td>
<td>Information Technology Division – Company</td>
</tr>
<tr>
<td>A uses or manages B</td>
<td>Systems Administrator – Company Network</td>
</tr>
<tr>
<td>A is a specialized version of the generalized B</td>
<td>Systems Administrator – Company Employee</td>
</tr>
<tr>
<td>A communicates with B</td>
<td>Suspect – Associates</td>
</tr>
<tr>
<td>A is known/logged/recording/reported in B</td>
<td>Email Registration – Network Logs</td>
</tr>
</tbody>
</table>
Identifying Attributes

- Select the Defining Characteristics of Each Concept
- The “Meat” of the Model
  - Attribute Values Seed the Examination
  - e.g. Email attribute source IP
- Some of the Eliminated Candidate Concepts May Serve as Attributes
## Attribute Examples

<table>
<thead>
<tr>
<th>Email Account</th>
<th>University Personnel</th>
<th>Workstation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provider Name</td>
<td>Name</td>
<td>Physical Location</td>
</tr>
<tr>
<td>Service Provider IP</td>
<td>PhoneNumbers</td>
<td>Type</td>
</tr>
<tr>
<td>Address</td>
<td>Addresses</td>
<td>MAC Address</td>
</tr>
<tr>
<td>Date Established</td>
<td>Email Addresses</td>
<td>IP Address</td>
</tr>
<tr>
<td>Registrant IP</td>
<td>Nicknames</td>
<td>System Time Source</td>
</tr>
<tr>
<td>Access Log</td>
<td></td>
<td>Name</td>
</tr>
<tr>
<td>Alternate Email</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Registrant Name</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Registrant Location</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example Case Domain Model Represented By UML Conceptual Diagram

- Student to Professor Death Threat
- Public Use University Computer
- Suspect Likely in Professor’s Class
Instantiate the Case Domain Model

- The Generalized Case Domain Model Must be Instantiated for a Specific Case
- Simply Fill in the Known Attribute Values
  - E.g. Suspect {name=Patrick Bateman}
- If Important Attribute Values are Unknown
  - Resume Investigative Efforts
  - Revisit Methodology
Training and Information Sharing

• Concepts are Abstract & May be Reused on Similar Case Types
• Useful for Providing an Investigative Training Templates
  – Using Existing, Expert Built Models
    • What Questions Should Be Asked in An Interview?
  – Following the Methodology Even on Smaller Cases
    • Allow Inexperienced Investigators to Develop Analytical Skills (maybe especially good for “new wave” of CF)
Deriving Keyword Search Terms with Case Domain Models

• Keyword Lists
  – Sometimes Required for Warrants
  – Useful in Forensics Software Tools
    • Password Crackers
    • File Searching

• Method For Deriving Candidate Seed Keywords
  – Select Appropriate Concepts From the Case Domain Model
  – Select Relevant Attributes
    • Ones You Can Find with a Keyword Search
  – Construct a Keyword List for Each Attribute
    • Elaborate on Different Synonyms and Representations
    • May be Automated (see Ruibin et al. [12])
Knowledge-Based Forensics Tools

• Requires More Formalized Knowledge Representation
  – Complex & Very Difficult for General Use

• Investigators Can Develop Informal Models Then Knowledge Engineers Can Formalize Them

• See Ruibin et al. [12]
  – Forensic Expert System
A “Unified” Forensics Modeling Methodology

In Software Engineering, Methods such as UML present Multi-View Models of a System
  - Requirements Views
  - Architectural Views
  - Implementation Views

• Forensics Modeling Views
  - Process View
  - Domain View
  - Hypothesis View
  - Examination Activity View

• Subject of Our Upcoming SADFE Paper
  - See You in Taiwan!
Conclusions

- Potential Benefits Large-Scale Investigations
  - A Structured Analytical Approach for Filtering and Organizing Information
  - Could Contribute to
    - Less Uncertainty
    - More Recovered Evidence
    - Improved Case Documentation
- May Be Too Burdensome for Smaller Cases
  - Require Less Planning
  - Are Very Familiar
  - Little or No Uncertainty with respect to Forensic Goals
- Methodology Needs Tuning for Practical Use
  - Needs Tailoring for Non SW Developers
- Adoption is Highly Dependent Upon Tools & Model Representation
  - Stanford Medical Informatics’ open-source Protégé tool is a Good Starting Places
Future Work

- Experiments on Case Domain Modeling Applied to Keyword Search Term Derivation
  - Evaluate Required Effort
  - Evaluate Amount of Evidence Recovered
  - Evaluate Practicality with Practitioners

- Prototype Case Domain Modeling Tool
  - Initial Prototype for Experiments
Q & A
Slide References

Other References

B. Chandrasekaran, J. R. Josephson, and V. R. Benjamins, "What are ontologies, and why do we
need them?," Intelligent Systems and Their Applications, IEEE [see also IEEE Intelligent Systems],


http://www.w3.org/TR/owl-features/ (current

53-74.

Hill, 2005.


no. 8, August, 2003, pp. 17-20.

Property Section, Search and Seizure Manual: Searching and Seizing Computers and Obtaining