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)

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

Outline

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

BACKGROUND

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

BACKGROUND

- 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

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

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.capegazette.com/pages/policrep.html

Case Concept Tools: USDOJ Manual

Case Type	Relevant Information Items
Email Threats/ Harassment / Stalking	Address books, diaries, e-mail/notes/letters, internet activity logs, legal documents, telephone records, financial/asset records, victim background research, images
Extortion	Date and time stamps, e-mail/notes/letters, history log, internet activity log, temporary internet files, user names

From USDOJ's Electronic Crime Scene Investigation a Guide for First Responder's

CASE DOMAIN MODELING

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

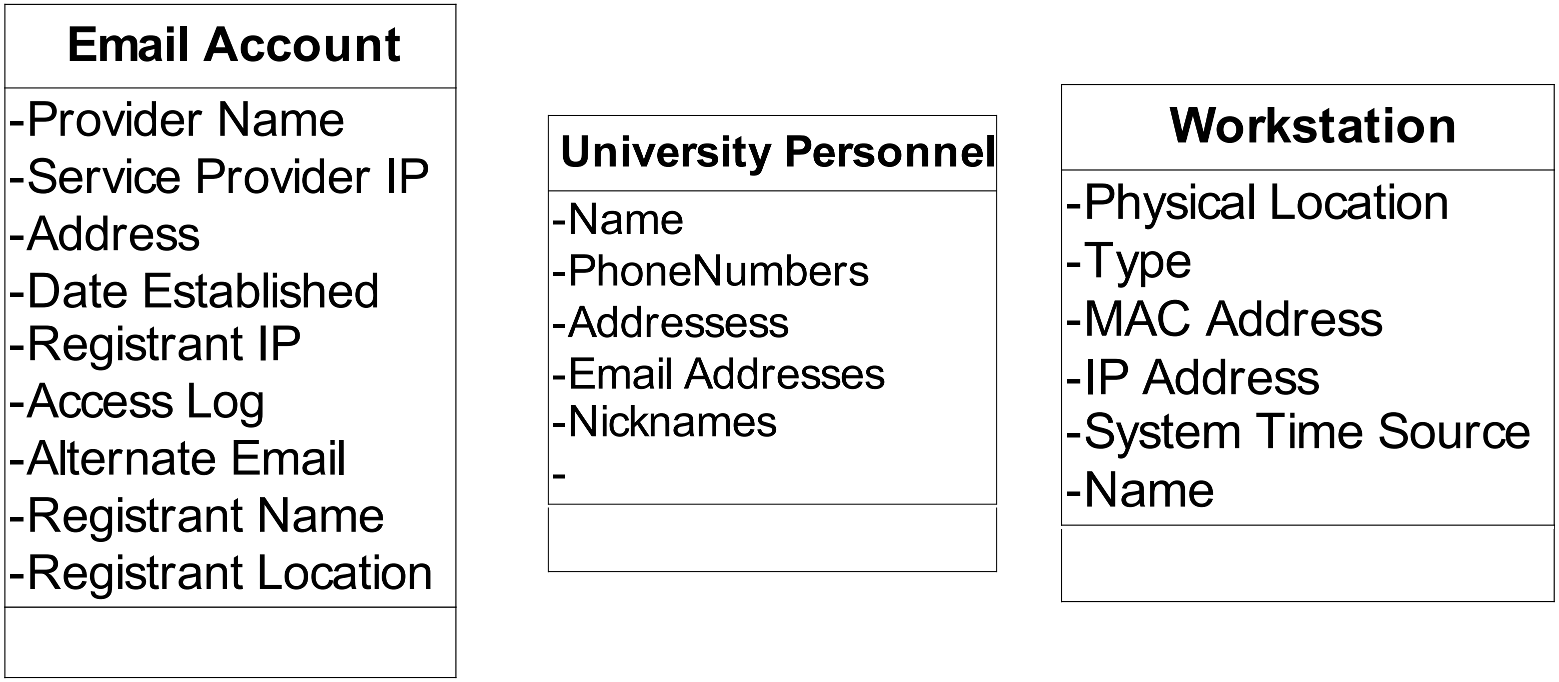
Concept Relationship Categories

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

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



Email Account

-Provider Name
-Service Provider IP
-Address
-Date Established
-Registrant IP
-Access Log
-Alternate Email
-Registrant Name
-Registrant Location

University Personnel

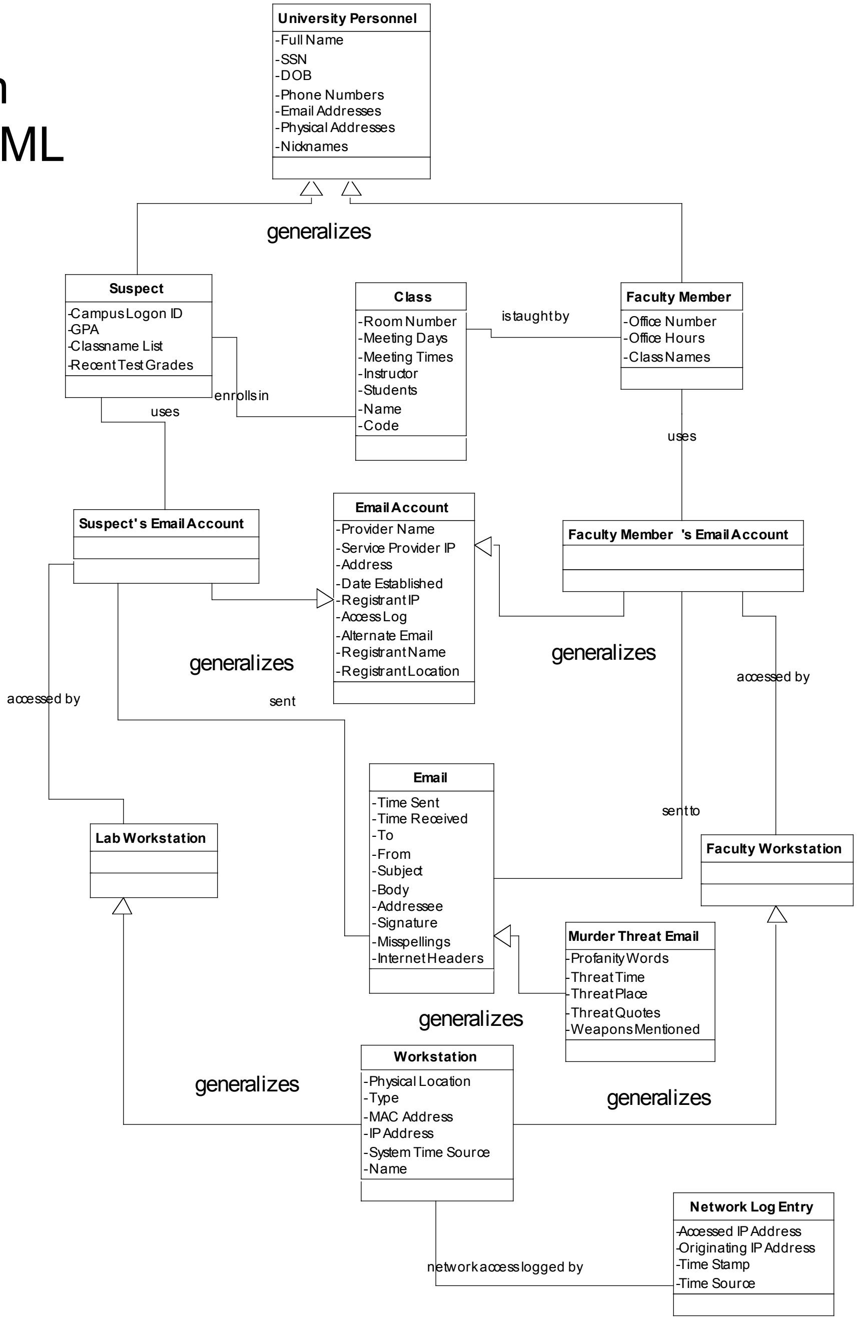
-Name
-PhoneNumbers
-Addressess
-Email Addresses
-Nicknames
-

Workstation

-Physical Location
-Type
-MAC Address
-IP Address
-System Time Source
-Name

Example Case Domain Model Represented By UML Conceptual Diagram

- •Student to Professor Death Threat
- •Public Use University Computer
- •Suspect Likely in Professor’s Class



University Personnel
-Full Name
-SSN
-DOB
-Phone Numbers
-Email Addresses
-Physical Addresses
-Nicknames

generalizes

Suspect
-Campus Logon ID
-GPA
-Classname List
-Recent Test Grades

Class
-Room Number
-Meeting Days
-Meeting Times
-Instructor
-Students
-Name
-Code

is taught by

Faculty Member
-Office Number
-Office Hours
-Class Names

enrolls in

uses

uses

Suspect's Email Account

Email Account
-Provider Name
-Service Provider IP
-Address
-Date Established
-Registrant IP
-Access Log
-Alternate Email
-Registrant Name
-Registrant Location

Faculty Member 's Email Account

generalizes

generalizes

accessed by

sent

accessed by

Email
-Time Sent
-Time Received
-To
-From
-Subject
-Body
-Addressee
-Signature
-Misspellings
-Internet Headers

sent to

Lab Workstation

Faculty Workstation

Murder Threat Email
-Profanity Words
-Threat Time
-Threat Place
-Threat Quotes
-Weapons Mentioned

generalizes

Workstation
-Physical Location
-Type
-MAC Address
-IP Address
-System Time Source
-Name

generalizes

generalizes

Network Log Entry
-Accessed IP Address
-Originating IP Address
-Time Stamp
-Time Source

network access logged by

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)

APPLICATIONS

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])

APPLICATIONS

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

Conclusions

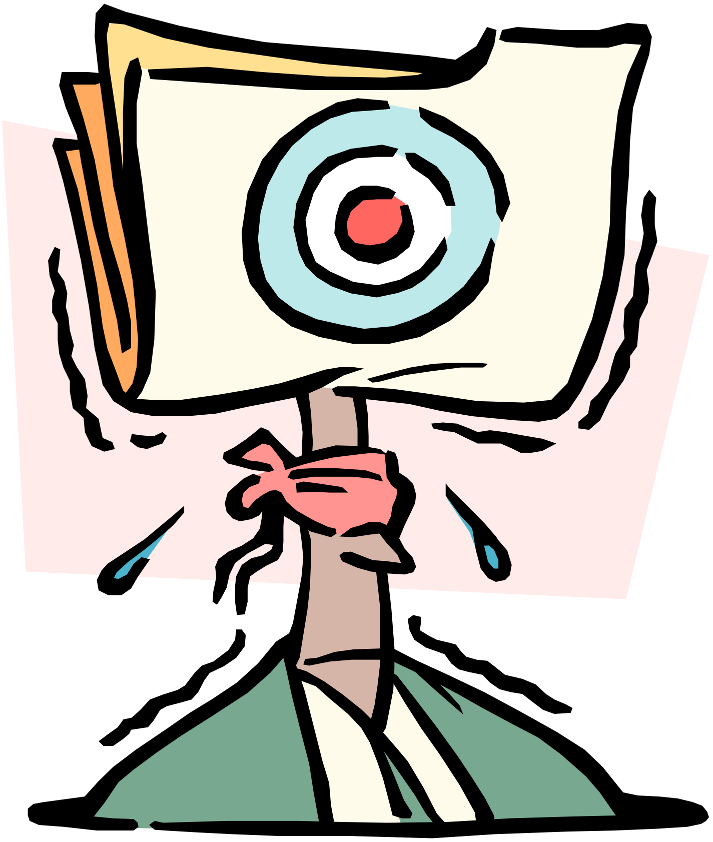
- Potential Benefits Large-Scale Investigations
  - A Structured Analytical Approach for Filtering and Organizing Information
  - Could Contribute to
    - Less Uncertainty
    - More Recovered Evidence
    - Improved Case Documentation
- May Be Too Burdensome for Smaller Cases
  - Require Less Planning
  - 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

CONCLUSIONS

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



CONCLUSIONS

Slide References

[1] V. Baryamureeba and F. Tushabe, "The Enhanced Digital Investigation Process Model," 2004; http://www.dfrws.org/ (current 2005 January 11).

[2] N. L. Bebe and J. G. Clark, "A Hierarchical, Objectives-Based Framework for the Digital Investigations Process," 2004; http://www.dfrws.org/ (current 2005 January 11).

[3] A. C. Bogen and D. A. Dampier, "Knowledge Discovery and Experience Modeling in Computer Forensics Media Analysis," presented at International Symposium on Information and Communication Technologies, Las Vegas, Nevada, 2004.

[4] A. C. Bogen and D. A. Dampier, "Modeling Evidence Recovery from Digital Media," *Naval Science and Engineering*, no., January, 2005,

[5] D. Bruschi and M. Monga, "How to Reuse Knowledge About Forensic Investigations," presented at Digital Forensics Research Workshop, Linthicum, Maryland, 2004.

[6] M. Carney and M. Rogers, "The Trojan Made Me Do It: A First Step in Statistical Based Computer Forensics Event Reconstruction," *International Journal of Digital Evidence*, vol. 2,no. 4, Spring, 2004, http://www.ijde.org/docs/04_spring_carneyrogers.pdf (current 9 July 2004).

[7] B. Carrier and E. Spafford, "Getting Physical with the Digital Investigation Process," *The International Journal of Digital Evidence*, vol. 2,no. 2, Fall, 2003, http://www.ijde.org/docs/03_fall_carrier_Spa.pdf (current 9 July 2004).

[8] P. Gladyshev, "Finite State Machine Analysis of a Blackmail Investigation," *International Journal of Digital Evidence*, vol. 4, no. 1, Spring, 2005,

[9] C. Larman, *Applying UML and Patterns An Introduction to Object-Oriented Analysis and Design*. Upper Saddle River, New Jersey: Prentice Hall, 1998.

[10] G. Palmer, "A Road Map for Digital Forensic Research," Utica, New York, technical report DTR-T001-0, 2001.

[12] G. Ruibin, T. Chan Kai Yun, and M. Gaertner, "Case-Relevance Information Investigation: Binding Computer Intelligence to the Current Computer Forensic Framework," *International Journal of Digital Evidence*, vol. 4, no. 1, Spring, 2005, pp. 1-13.

[12] P. Stephenson, "Using a Formalized Approach to Digital Investigation," in *Getting the Whole Picture, A Series of 12 Columns on End-to-End Digital Investigation (EEDI) Appearing in Elsevier Advanced Technology's, "Computer Fraud and Security" Publication in 2002 and 2003*, vol. 1: International Institute for Digital Forensic Studies & Elsiever Advanced Technology, 2003, pp. 7.

[13] P. Stephenson, "The Application of Formal Methods to Root Cause Analysis of Digital Incidents," *International Journal of Digital Evidence*, vol. 3, no. 1, Summer, 2004,

[14] United States Department of Justice Office of Justice Programs, "Electronic Crime Scene Investigation a Guide for First Responders," United States Department of Justice, Washington, DC July 2001.

CONCLUSIONS

Other References

[1] Association of Chief Police Officers (AOPO), "Good Practice Guide for Computer Based [3] 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]*, vol. 14, no. 1, 1999, pp. 20-26.

[2] N. Iscoe, G. B. Williams, and G. Arango, "Domain modeling for software engineering," presented at Software Engineering, 1991. Proceedings., 13th International Conference on, 1991.

[3] D. McGuiness and F. van Harmelen, "OWL Web Ontology Language Overview," 2004; http://www.w3.org/TR/owl-features/ (current

[4] N. Noy and C. Hafner, "The State of the Art in Ontology Design," *AI Magazine*, vol. 18, no. 3, 1997, pp. 53-74.

[5] R. Pressman, *Software Engineering A Practitioner's Approach*, 6th ed. New York, New York: McGraw Hill, 2005.

[6] R. Prieto-Diaz, "A faceted approach to building ontologies," presented at Information Reuse and Integration, 2003. IRI 2003. IEEE International Conference on, 2003.

[7] B. Schneier, "Attack Trees," *Dr. Dobb's Journal*, vol. 24, no. 12, December, 1999, pp. 21-29.

[8] P. Stephenson, "Applying DIPL to an Incident Post Mortem," *Computer Fraud and Security*, vol. 2003, no. 8, August, 2003, pp. 17-20.

[9] United States Department of Justice Office of Justice Programs Computer Crime and Intellectual Property Section, *Search and Seizure Manual: Searching and Seizing Computers and Obtaining Electronic Evidence in Criminal Investigations*, 1.0 ed. Washington, DC, 2002.

[10] M. Uschold and M. Gruninger, "Ontologies: Principles, Methods, and Applications," *The Knowledge Engineering Review*, vol. 11, no. 2, 1996, pp. 93-136.