A Multi-Disciplinary Approach for Countering Insider Threats

Robert DelZoppo, Eric Brown, Matt Downey: Syracuse Research Corporation
Michael D’Eredita, Elizabeth D. Liddy, Joon S. Park, Anand Natarajan, Svetlana Symonenko, Shuyuan M. Ho: Syracuse University

Secure Knowledge Management (SKM 2004)

September 23-24, 2004
Marriott Buffalo-Niagara
Amherst, NY USA
Insider Threat

- Mission-critical information = High-value target
- Threatens US Intelligence Community (IC), other Government organizations and large corporations
- Probability is low, but impact is severe

Types of Threat posed by malicious insiders
  - Denial of service
  - Compromise of confidentiality
  - Compromise of integrity

High complexity of problem
  - Increase in sharing of information, knowledge
  - Increased availability of corporate knowledge online
  - “Low and Slow” nature of malicious insiders
**Malicious Insider, examples**

**Robert Hanson:** (1985-2001)
- Compromise: Exfiltrated over 6000 pages of classified material
- Impact:
  - Divulged Intel capabilities of FBI and other agencies
  - Identified three Soviet double agents (1 imprisoned, 2 killed)
- Cyber Activities:
  - Frequent need-to-know “violations”
  - Frequent queries looking for signs of an investigation targeting him

**Brian Patrick Regan:** (1999-2001)
- Compromise: Removed and hid over 800 pages of classified material, email contact to leaders in Iraq, Libya, and China
- Impact:
  - Suspected acquisition of classified imagery and reports to Iraq
- Cyber Activities:
  - Frequent need-to-know “violations”
  - High volume printing; Encrypted emails
Characteristics of Malicious Insider Behavior (current, projected)

- Technically competent to highly-skilled
- Attempts to cover up, destroy evidence
- Sophisticated search / query techniques
- Abuses security clearance to gain access to information (violates “need to know”)
- Downloads data to new devices (e.g., USB thumb drive)
- Encrypts data
- Changes system logs to hide activity
- Uses “stealthy” techniques to communicate with handlers (e.g., encrypted email)
Approach

- **Staged**: Detect anomalies in user behavior from cyber observables and, based on these anomalies, assess the risk of malicious insider behavior.

- **Multi-Perspective**: Detect anomalies in user behavior considering *user-to-user*, *user-to-content*, *user-to-resource* relationships.

- **Multi-Disciplinary**:
  - **Social Network Analysis (SNA)** - Apply concepts from SNA to detect anomalies in social behavior [*user-to-user*].
  - **Semantic Analysis (SA)** - Leverage Natural Language Processing (NLP) and machine learning techniques to analyze the textual data associated with insiders at a semantic (conceptual) level [*user-to-content*].
  - **Composite, Role-based Monitoring (CRBM)** – Analyze insider activity based on the organizational, application and operating system roles. [*user-to-resource*]
Research Objectives

- Advance the state-of-art in Insider Threat Countermeasures by developing techniques to:
  - Model behavior of insiders operating in an IC-based context
  - Distinguish between expected and anomalous user behavior
  - Detect indicators of malicious insider behavior (MIB)
  - Assess indicators of MIB for potential threat to the confidentiality and integrity of information.

- To reduce the overall effort in countering threat from malicious insiders:
  - Reduce the size of the problem space to a manageable number of indicators a system security / assurance administrator would need to look at
  - Provide early awareness of risk elevating situations
To provide a robust solution which:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Has Breadth</td>
<td>Incorporates a wide range of observable types and can assess multiple types of risk</td>
</tr>
<tr>
<td>Has depth</td>
<td>Can analyze observables at fine-grained levels (e.g., semantics)</td>
</tr>
<tr>
<td>Is scalable</td>
<td>Can model behavior at multiple levels (e.g., insider, role) and is minimally impacted as # of insiders increases</td>
</tr>
<tr>
<td>Is extensible</td>
<td>Can be extended to incorporate new threat scenarios and other sources of indicators (e.g., anomaly detectors)</td>
</tr>
<tr>
<td>Is reusable</td>
<td>Modules could be reused in another system or context</td>
</tr>
</tbody>
</table>
Assumptions

- Insiders with similar roles, goals and tasks will have similar behavior.

- Malicious insider behavior will differ, to a measurable degree, from behavior of typical insiders.

- Insiders’ actual behavior will be discernable through cyber-observations from sensors which currently exist or could be constructed.

- Anomaly-based or signature-based methods, by themselves, are insufficient for identification of Insider Threats.
Hierarchically organized by role/goal/task (RGT)

Allows for computation of non-deterministic behavior (e.g., multitasking)

Provides scoping mechanism

Can be used for both pattern matching and data generation
Risk Assessment

Risk is identified as indicators are asserted; indicators are asserted from the anomalies detected.

- Observables
- Anomalies
  - atypical access to system
  - high-degree of off-topic consumption
  - low-degree of expected interaction
- Indicators
  - “collector” behavior pattern
- Risk
  - Confidentiality compromise (High)
System Overview

Expected Behavior Model

Observable Activity
black boxed sensor input such as:
- web logs
- print logs
- email monitors
- phone logs
- system access logs
- Host sensor logs
- card key readers
- etc.

Risk Assessor

Anomaly Detectors

Social Network Analysis

Semantic Analysis

Composite Role-Based Analysis

Risks & Alerts
**Current Work:** Relational Matrix Analysis Tool  
(user-to-user, user-to-resource)

**Observables (from Scenario)**
<Observable>
  <Name>Terry</Name>
  <Role>analyst</Role>
  <Toi>Biological Weapons</Toi>
  <Aoi>Russia</Aoi>
  <Task>Report</Task>
  <Method>leave VM</Method>
  <ResourceLabel>Smith</ResourceLabel>
  <ResourceType>senior reporter</ResourceType>
  <Time>1071032734</Time>
</Observable>

**Given:**
- Observables
- Method Restrictions
- Insider Restrictions
- Resource Restrictions
- Generate Relational Matrices
  - Based on insider (constrained by RGT) versus a hierarchy of resources, goals, and interaction methods
  - Comparison level: specific (explicit resource) or generic (resource type)
- Perform Outlier Analysis

**Resource Restrictions:**
- TOI, AOI, task

**Insider Restrictions:**
- role, TOI, AOI, task

**Method Restrictions**

**Insider vs. Resource Matrix**

<table>
<thead>
<tr>
<th></th>
<th>datasource2</th>
<th>datasource4</th>
<th>datasource6</th>
<th>datasource7</th>
<th>datasource8</th>
<th>datasource9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Anderson</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Apok</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cipher</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Eric</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Fred</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Graham</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Herb</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hoss</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Joe</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>John</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Lars</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lucy</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Matt</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Outlier Indicators and Analysis**
Current Work: Semantic Analyses

Document clustering, based on geographic area-of-interest
Current Work: Semantic Analyses
(user-to-content)

Document clustering, based on *topic-of-interest*

Cluster 0

| Cluster | Size | Precision | Recall | F-Measure | Elbow | Elbow

| 0        | 16   | 0.906     | 0.057  | 0.294     | 0.196 |

Descriptive & Discriminating Features

| Cluster | Size | Elbow |

| 0        | 16   | 0.294 |

- Cluster 0:
  - Precision: 0.906
  - Recall: 0.057
  - F-Measure: 0.294
  - Elbow: 0.196

- Descriptive & Discriminating Features:
  - Cluster 0: Size 16, Elbow 0.294
  - Discriminative: missile 37.7%, Missile 37.7%, Weapon 23.8%, Weapon-of-Mass-Destruction 0.4%, Nukes 0.3%
  - Discriminating: missile 23.8%, Missile 23.8%, Weapon-of-Mass-Destruction 21.8%, Biological Weapon 11.0%, Anthrax 5.4%
System Architecture

- **Scenario Generator**
  - CPN Tools
  - IC Workflow Model

- **Observable Archive**

- **Risk Assessor**
  - **Controller / Rule Engine**
    - JESS

- **Expected Behavior Model**

- **Social Network Monitor**
  - JUNG

- **Semantic Analysis Monitor**
  - CNLP Technology

- **Risk Policy**

- **Risk Assessment Display**
  - i2 Analyst Notebook
  - MS Excel

- **Composite Role-based Monitor**
  - Role-based Research

- **Document Collection**

- **COTS**

- **R&D Leverage**

- **ARDA**

- **XML interface**

Copyright © 2004
Scalability of Solution

- **High Scalability / Extensibility**
  - Other anomaly detectors can be added to provide additional indicators
  - Risk Assessment Policy provides a means for writing new rules and sets of rules

- **Generalizability**
  - Methodology provides abstraction mechanisms for managing complexity
  - Approach can be generalized to other domains

- **Reusability / Interoperability**
  - Anomaly detectors can provide indicators to other types of systems
  - XML-based interfaces – provide “loose” couplings between modules
Limitations/Vulnerabilities

❑ **Non-cyber activities**
  • Mitigation: Security Administrator Application for entering / managing non-cyber indicators

❑ **Undetected cyber observables:**
  • Most non-textual media (Images, Audio, Video)
    » Example: Communications analyst inappropriately retrieving images unrelated to task
    » Mitigation: Analyze image meta-data to provide basic analysis of image content
  • Anonymous user behavior – Guest, and other potentially anonymous activities such as access through web-based applications
    » Mitigation: Can still monitor to identify risk
  • Account “masquerading”
    » Mitigation: Focus on individual insiders; detect shifts in behavior
Summary

- Currently under experimentation using controlled simulation with synthetic data sets (scenarios):
  - Baseline scenario – observables under normal conditions
  - “Threat” scenarios – baseline scenario with anomaly injection
  - Includes supporting UNCLASSIFIED document collections on a variety of topics (e.g., Terrorism/WMD)

- Preliminary results indicate
  - Role-Goal-Task-orientation of Expected Behavior Model provides a basis for modeling context-dependent behavior
  - Relational Matrix approach very well suited to anomaly detection in entity-to-entity interaction
  - Semantic Analysis approach works well to identify off-topic information access
Acknowledgements

- Advanced Research and Development Activity (ARDA) Advanced Countermeasures for Insider Threat (ACIT) Program (sponsor)

- Other ARDA Programs
  - Cyber Indications & Warning (CIW) Workshop (MITRE, Aug 03)
  - Advanced Question & Answering for Intelligence (AQUAINT)
  - Novel Intelligence from Massive Data (NIMD)

- Mitigating the Insider Threat to Information Systems - #2; Workshop Proceedings (RAND, Aug 00)