Monitoring DBMS Activity for Detecting Data Exfiltration by Insiders

Customer Need: Detect and Respond to Insider Threats

- Corporate Awareness
- Employee Behavior
- Types of Insiders
- Malicious
- Unwitting
- Rule Bender
- Types of Breaches
- Unauthorized Access
- Intellectual Property Theft
- Exposure of Sensitive Data
- Other Data Theft
- Average Time to Detect: 32 Months
- Damage to Enterprise
- Cost
- Reputation
- Operations
- Lives

Background

Hypothesis
Exfiltration causes an anomalous state that can be distinguished from the legitimate actions executed in a DBMS system.

Challenge
Identify the events that represent signs of cyber-insider actions:
- "How do we define and identify user queries that are anomalous?"
- "Which data sources does an insider target?"
- "What information should be collected to detect such actions?"
- Provide core algorithms
- Develop Proof of Concepts
- Test in lab environment
- Transform PoCs to “shrink-wrapped” prototypes
- Integrate COTS products
- Maintain during evaluation

Approach (Technical)
- Build accurate DBMS access profiles (patterns of normalcy) using Role Based Access Control (RBAC) model
- Detect and respond to anomalous user behavior and events
- Observe deviations from profiles in real-time
- Alert security operators
- Respond according to set policies and forensics

Approach (Programmatic)
- Role profiling
- Enhanced machine learning algorithms
- Analysis of query optimizers for use in profiling the selectivity of role queries (e.g., for data-based anomaly detection)
- Application program profiling
- Profile and monitor application programs with respect to their database accesses
- Use concolic testing to capture the application behavior.
- Response mechanisms

Benefits
- Dynamic and automated generation of behavioral profiles
- Near-real time alerts of anomalous database activity
- Policy-defined (automated) response
- History and explanation for forensics

Current Status (Prototype 1)

Evaluation Results

<table>
<thead>
<tr>
<th>Detector Type</th>
<th>Evaluation Method</th>
<th>True Positives</th>
<th>False Positives</th>
<th>Average Values</th>
<th>Average Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>Human Evaluation</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Bayesian Detector</td>
<td>AD Score - Alerts only</td>
<td>41.73%</td>
<td>14.54%</td>
<td>56.17%</td>
<td>30.00%</td>
</tr>
<tr>
<td></td>
<td>Human Evaluation</td>
<td>39.31%</td>
<td>8.50%</td>
<td>47.81%</td>
<td>17.00%</td>
</tr>
<tr>
<td>Binary Detector</td>
<td>AD Score</td>
<td>66.37%</td>
<td>55.72%</td>
<td>61.04%</td>
<td>50.00%</td>
</tr>
<tr>
<td></td>
<td>Human Evaluation</td>
<td>48.79%</td>
<td>12.75%</td>
<td>61.04%</td>
<td>34.84%</td>
</tr>
<tr>
<td>For Reference</td>
<td></td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Summary Using Only Parsed Data

<table>
<thead>
<tr>
<th>Detector Type</th>
<th>Evaluation Method</th>
<th>True Positives</th>
<th>False Positives</th>
<th>Average Values</th>
<th>Average Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>Human Evaluation</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Bayesian Detector</td>
<td>AD Score - Alerts only</td>
<td>61.04%</td>
<td>19.78%</td>
<td>58.14%</td>
<td>11.81%</td>
</tr>
<tr>
<td></td>
<td>Human Evaluation</td>
<td>58.14%</td>
<td>11.81%</td>
<td>58.14%</td>
<td>11.81%</td>
</tr>
<tr>
<td>Binary Detector</td>
<td>AD Score</td>
<td>89.20%</td>
<td>75.68%</td>
<td>80.92%</td>
<td>60.00%</td>
</tr>
<tr>
<td></td>
<td>Human Evaluation</td>
<td>65.21%</td>
<td>18.08%</td>
<td>65.21%</td>
<td>18.08%</td>
</tr>
<tr>
<td>For Reference</td>
<td></td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Next steps (Prototype 2)

Three Phases over three years
- Prototype 1: Initial key features in controlled lab environment
- Prototype 2: Expanded features in controlled lab environment
- Pilot: Operational environment at select government agency

Customer Need:
Detect and Respond to Insider Threats

- Exposure of Sensitive Data
- Unauthorized Access
- Other Data Theft
- Intellectual Property Theft
- Corporate Awareness
- Malicious
- Unwitting
- Rule Bender

Exposure of Sensitive Data
Unauthorized Access
Other Data Theft
Intellectual Property Theft

Purdue:
Elisa Bertino
Lorenzo Bossi
Syed Rafiul Hussain
Asmaa Sallam

Northrop Grumman:
David Landers
R. Michael Lefler
Donald Steiner

Approach (Programmatic)
- Provide core algorithms
- Develop Proof of Concepts
- Test in lab environment
- Transform PoCs to “shrink-wrapped” prototypes
- Integrate COTS products
- Maintain during evaluation

Approach (Technical)
- Build accurate DBMS access profiles (patterns of normalcy) using Role Based Access Control (RBAC) model
- Detect and respond to anomalous user behavior and events
- Observe deviations from profiles in real-time
- Alert security operators
- Respond according to set policies and forensics

Approach (Programmatic)
- Role profiling
- Enhanced machine learning algorithms
- Analysis of query optimizers for use in profiling the selectivity of role queries (e.g., for data-based anomaly detection)
- Application program profiling
- Profile and monitor application programs with respect to their database accesses
- Use concolic testing to capture the application behavior.
- Response mechanisms