EXTRACTING ATTACK SCENARIOS USING INTRUSION SEMANTICS

Sherif Saad, Issa Traore
Information Security and Object Technology Lab
University of Victoria
ECE Department
Overview

- Introduction
- Related Work
- Approach
- Experiment
- Conclusion
Introduction: Definition

- Attack Scenario:
  - Elicits the steps and actions taken by the intruder to breach/compromise the system.
  - Also known as attack plan.

- Attack Pattern:
  - A collection of malicious actions that together represent a pattern.
Introduction: Motivation

- Extracting attack scenario is needed to:
  - Elicit the attack and extract useful attack intelligence,
  - Identify the compromised resources,
  - Spot the system vulnerabilities,
  - Determine the intruder objectives and the attack severity.
Introduction: Research problem

- IDSs generate low level intrusion alerts that describe individual attack event.
- IDS are not designed to recognize attack plans or discover multistage attack scenarios.
- IDSs tend to generate massive amount of alerts with high rate of redundant alerts and false positives.
- False negatives, which correspond to the attacks missed by the IDS.
Related Works

- Statistical and Clustering Approaches
  - Use alerts similarity and statistical characteristics.
  - Can handle large amount of IDS alerts.
  - Can reconstruct novel and unknown attack scenarios.
  - Cannot detect causality between individual attacks
  - Limited to simple attack scenarios and attack patterns.
  - Reconstruct false attack scenarios.
Related Works (ctd.)

- Knowledge Based Approaches
  - Use hard coded knowledge and rely on explicit knowledge.
  - Hard to maintain and update the knowledge-base
  - Can reconstruct complex and multistage attack scenarios, but not novel attacks scenarios.
  - Cannot handle false negatives and missing attack steps.
  - Cannot detect hidden and implicit relations between attacks.
Research Objectives

- Develop a new approach that:
  - Handle large amount of IDS alerts.
  - Handle complex multistage attack scenarios.
  - Automatically reconstruct novel and unknown attack scenarios with high accuracy.
  - Minimize the affect of missing attack steps.
IDS sensors use different formats and vocabularies to describe the alerts.

IDMEF provides common alert message structure (syntax) not semantic.

IDS alerts message attributes are symbolic data. It is hard to measure similarity or distance between symbolic data.
Approach: Intrusion Ontology

Fig. 1: Intrusion Ontology Screenshot
Approach: Semantic Relevance
Approach: Semantic Clustering

- Use the ontology to measure the semantic relevance between different alert messages.

\[
\text{sem}_{\text{rel}}(x, y) = \frac{\sum_{r \in R_{xy}} w(r)}{\text{cardinality}(R)}
\]

- Using the semantic relevance we build the alert correlation graph (ACG).

- Analyze the ACG to extract all maximum cliques in ACG. We consider every maximum clique in the ACG as a candidate attack scenario/pattern.
## Semantic Clustering Example

<table>
<thead>
<tr>
<th>ID</th>
<th>Source</th>
<th>Target</th>
<th>Attack</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>201.134.12.11</td>
<td>172.16.112.10</td>
<td>Scan</td>
</tr>
<tr>
<td>$a_2$</td>
<td>201.134.12.11</td>
<td>172.16.116.44</td>
<td>Scan</td>
</tr>
<tr>
<td>$a_3$</td>
<td>135.13.216.191</td>
<td>172.16.113.84</td>
<td>Scan</td>
</tr>
<tr>
<td>$a_4$</td>
<td>201.134.12.11</td>
<td>172.16.112.10</td>
<td>BufferOverflow</td>
</tr>
<tr>
<td>$a_5$</td>
<td>135.13.216.191</td>
<td>172.16.116.44</td>
<td>Scan</td>
</tr>
<tr>
<td>$a_6$</td>
<td>201.134.12.11</td>
<td>172.16.112.10</td>
<td>RootAccess</td>
</tr>
<tr>
<td>$a_7$</td>
<td>135.13.216.191</td>
<td>172.16.116.44</td>
<td>TelnetAccess</td>
</tr>
</tbody>
</table>
Semantic Clustering Example

Alerts Correlation Graph  All Maximum Cliques in ACG
The Impact class in the ontology contains the set of attack prerequisites and consequences.

The causality relation between two attack instances \(a\) and \(b\) is a value between 0 and 1 given by

\[
\text{causality}(a, b) = \frac{|A \cap B|}{|A \cup B|}
\]

The sequence of attacks in the attack scenario is based on the causality between attack instance in the scenario.
Experiments

- Datasets
  - DARPA 2000 dataset from MIT Lincoln Laboratory.
  - The Treasure Hunt dataset.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>LLDDOS1.0</th>
<th>Hunt-DMZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alerts</td>
<td>2170</td>
<td>671848</td>
</tr>
<tr>
<td>Intrusions</td>
<td>16</td>
<td>49</td>
</tr>
<tr>
<td>Sources</td>
<td>273</td>
<td>28</td>
</tr>
<tr>
<td>Targets</td>
<td>738</td>
<td>37</td>
</tr>
<tr>
<td>Duration</td>
<td>$\approx$ 100 minutes</td>
<td>$\approx$ 893 minutes</td>
</tr>
</tbody>
</table>
Evaluation Metrics

- Two performance metrics:
  - Completeness: the ratio between the number of correctly correlated alerts by the number of related alerts (i.e. that belong to the same attack scenario).
  - Soundness: the ratio between the number of correctly correlated alerts by the number of correlated alerts.
# Experiments Results

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Correlated alerts</th>
<th>True alerts</th>
<th>Related alerts</th>
<th>Completeness</th>
<th>Soundness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web Exploit 1</td>
<td>503337</td>
<td>503337</td>
<td>503337</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Web Exploit 2</td>
<td>101071</td>
<td>100758</td>
<td>100758</td>
<td>100.00%</td>
<td>99.69%</td>
</tr>
<tr>
<td>Protocol Exploit</td>
<td>1730</td>
<td>1701</td>
<td>1705</td>
<td>99.77%</td>
<td>98.32%</td>
</tr>
<tr>
<td>Reconnaissance</td>
<td>3097</td>
<td>2973</td>
<td>3053</td>
<td>97.38%</td>
<td>96.00%</td>
</tr>
<tr>
<td>Privilege Escalation</td>
<td>20191</td>
<td>19981</td>
<td>20103</td>
<td>99.39%</td>
<td>98.96%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Approach</th>
<th>Completeness</th>
<th>Soundness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ning et al</td>
<td>93.96%</td>
<td>93.96%</td>
</tr>
<tr>
<td>Liu et al</td>
<td>87.12%</td>
<td>86.27%</td>
</tr>
<tr>
<td>Al-Mamory and Zhang</td>
<td>86.5%</td>
<td>100%</td>
</tr>
<tr>
<td>Li et al</td>
<td>92.2%</td>
<td>not provided</td>
</tr>
<tr>
<td>Our Approach</td>
<td>100%</td>
<td>99.70%</td>
</tr>
</tbody>
</table>
Conclusion & Future Work

- The use of semantic correlation and ontology:
  - Allow us to develop a better alert correlation and attack scenario reconstruction technique.
  - Enable interoperability between heterogeneous IDS sensors.
  - Improve the knowledge-base maintenance.
  - Eliminate the need of hard-coded rules.
Conclusion & Future Work

Our future work will focus on:

- **False negative**: improve the attack causality analysis to predict missing attack steps.
- **False positive**: develop an ontology-based rule induction to reduce the false positive alerts.
Thanks

Questions??