DETECTION OF AUTHENTICATION EVENTS INVOLVING STOLEN ENTERPRISE CREDENTIALS

Mijung Kim
Research Engineer
Micro Focus

Pratyusa K. Manadhata
Principal Researcher
Micro Focus
Motivation

Gone Phishing
Problem statement

Scalable, reliable, and timely detection of malicious authentication events
Challenges

- Base rate fallacy
- Similarity of good and bad events
A machine learning based solution

Data → Features → Models → Validation

Infrastructure
An authentication event

- Time of authentication
- Source device and source user
- Destination device and destination user
- Authentication type, orientation, logon type, outcome

*Hard to differentiate malicious from benign*
The context of an event

![Diagram showing the context of an event between a source device and a destination device. The diagram includes labels for Host + Network Events, Authentication Event, and CONTEXT.]]
Scalable, reliable, and timely classification of an authentication event’s context
EXPERIMENTAL RESULTS
Los Alamos National Labs data

- Collected from Los Alamos National Labs’ network over 58 days

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Users</td>
<td>12.4K</td>
</tr>
<tr>
<td>Devices</td>
<td>17.7K</td>
</tr>
<tr>
<td>Events (Authentication, DNS, Netflow, Process)</td>
<td>1.65B</td>
</tr>
<tr>
<td>Authentication events</td>
<td>1.05B</td>
</tr>
</tbody>
</table>

https://csr.lanl.gov/data/cyber1/
Malicious authentication events

749 events performed by a red team using stolen credentials

How to distinguish 749 malicious events from 1.05B events?
Data reduction for scalability

- 1.05 Billion
- 950M
- 100M
- 100M
- 950M
- 100M
- 740
- Malicious
- Benign
- Benign

Rules
Examples

- Filter out local events
- Focus on network authentication
- Focus on successful authentication
- ..

*Rule matching shouldn’t have false negatives, but false positives*
Feature extraction

- Given an authentication event at time T, extract features from:
  - Events on the source device in the time period (T-W)
  - Network events between the source and the destination
  - Events on the destination device in the time period (T + W)

- Feature identification via **domain expertise**
Example features

- **Authentication logs**
  - Failures/successes at the source and the destination

- **Netflow logs**
  - Connections per protocol, Number of bytes/packets on standard/non-standard ports, ..

- **DNS logs**
  - Frequency of DNS events at the source and the destination, ..
Model selection

- Model selection data
  - Randomly chosen 10K legitimate events and 3.5K compromised events
  - 5 fold replication of compromised events to handle class imbalance

- Training and test split: 75%:25% and 10 fold cross validation
Performance of different models

<table>
<thead>
<tr>
<th>Model</th>
<th>True Positive Rate</th>
<th>False Positive Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest</td>
<td>0.988</td>
<td>0.030</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>0.977</td>
<td>0.056</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>0.929</td>
<td>0.154</td>
</tr>
<tr>
<td>Multilayer Perceptron</td>
<td>0.973</td>
<td>0.076</td>
</tr>
<tr>
<td>SMO</td>
<td>0.951</td>
<td>0.135</td>
</tr>
</tbody>
</table>

Reporting 75:25 split results (10 fold CV results are similar)
An ‘end to end’ experiment

- Model generation
  - 8K benign and 2.5K malicious (5 fold replication)

- Parameter selection
  - 80M benign and 124 malicious

- Error estimation on Test data
  - 20M benign and 124 malicious
Precision-recall plots

- Better than ROC plots for imbalanced data sets
  - Even a very low FPR produces many FPs

- Precision
  - Fraction of true positives in events detected as malicious
  - $\frac{TP}{TP + FP}$

- Recall:
  - Fraction of malicious events detected
  - $\frac{TP}{TP + FN}$
Threshold selection

Threshold = 0.99
Precision = 0.19
Recall = 0.75
Test data results

In order to identify $\frac{3}{4}$th of the malicious events, the model will generate 52% false positives.

That is, 1 out of every 2 detections will be a false positive.
A note about false positives

- 1 false positive for each true positive may seem high
- But the number of true positives are very low
  - so the absolute number of false positives will be low.

- Test data: 120 true positives over 60 days.
Features from only authentication events

Threshold = 0.99
Precision = 0.3
Recall = 0.70

2 out of every 3 detections will be false positives.
MODEL GENERATION INFRASTRUCTURE
Model generation and prediction challenges

- Scalable feature computation and model learning
- Real time detection of compromised authentication events
- Performance issues
  - Feature extraction takes too long
Scale and performance assumptions

- **Data volume in a large enterprise**
  - 5 billion events/day (with 0.5 KB/event, 2.5 TB/day, without compression)
  - Higher number of events when including high volume sources such as Netflow

- **Streaming data in nature**

- **Analytics is continuous, not just on data at rest**
Event streaming framework

- Data aggregation every $W$ minutes
  - Aggregated data
  - Aggregated data
  - Aggregated data
- Feature computation every $S$ minutes
  - Features
  - Features
- Model update every $T$ hours
  - Model
- Event data streaming
  - Aggregated data streaming
  - Feature data streaming
  - Model streaming
Streaming malicious authentication detection

- **Authentication**
- **DNS**
- **Netflow**
- **Process**

Feature extraction every $S$ minutes

Model update every $T$ hours

- **Feature data**
- **Model**
- **Prediction**
- **Feedback**
User Interface

Ranked list of malicious events

Feature values for an authentication event
User Interface

Ranked list of malicious events

Details of malicious event
Feature values for an authentication event
### Feature values for an authentication event

<table>
<thead>
<tr>
<th>Time Issued</th>
<th>Destination Computer</th>
<th>Source User</th>
<th>Destination User</th>
<th>Auth Type</th>
<th>Status</th>
<th>Num of Successful auth events at Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct 28, 2017 04:37:59 PM</td>
<td>CS111</td>
<td>U66@DOM1</td>
<td>U750@DOM1</td>
<td>NTLM</td>
<td>Success</td>
<td>123</td>
</tr>
<tr>
<td>Oct 28, 2017 05:27:59 PM</td>
<td>C423</td>
<td>U66@DOM1</td>
<td>U66@DOM1</td>
<td>NTLM</td>
<td>Success</td>
<td>100</td>
</tr>
<tr>
<td>Oct 28, 2017 06:07:59 PM</td>
<td>C3888</td>
<td>U66@DOM1</td>
<td>U66@DOM1</td>
<td>NTLM</td>
<td>Success</td>
<td>50</td>
</tr>
<tr>
<td>Oct 28, 2017 06:47:59 PM</td>
<td>C14401</td>
<td>U66@DOM1</td>
<td>U66@DOM1</td>
<td>NTLM</td>
<td>Success</td>
<td>134</td>
</tr>
<tr>
<td>Oct 28, 2017 06:51:59 PM</td>
<td>C4403</td>
<td>U66@DOM1</td>
<td>U66@DOM1</td>
<td>NTLM</td>
<td>Success</td>
<td>99</td>
</tr>
<tr>
<td>Oct 29, 2017 02:03:59 AM</td>
<td>C798</td>
<td>U66@DOM1</td>
<td>U750@DOM1</td>
<td>NTLM</td>
<td>Success</td>
<td>134</td>
</tr>
<tr>
<td>Oct 29, 2017 02:37:50 AM</td>
<td>C1096</td>
<td>U66@DOM1</td>
<td>U66@DOM1</td>
<td>NTLM</td>
<td>Success</td>
<td>156</td>
</tr>
<tr>
<td>Oct 29, 2017 02:37:59 AM</td>
<td>C423</td>
<td>U66@DOM1</td>
<td>U750@DOM1</td>
<td>NTLM</td>
<td>Success</td>
<td>140</td>
</tr>
<tr>
<td>Oct 29, 2017 02:41:50 AM</td>
<td>C3888</td>
<td>U66@DOM1</td>
<td>U1299@DOM1</td>
<td>NTLM</td>
<td>Success</td>
<td>115</td>
</tr>
<tr>
<td>Oct 29, 2017 02:41:59 AM</td>
<td>C14401</td>
<td>U66@DOM1</td>
<td>U66@DOM1</td>
<td>NTLM</td>
<td>Success</td>
<td>134</td>
</tr>
<tr>
<td>Oct 29, 2017 02:41:59 AM</td>
<td>C4403</td>
<td>U66@DOM1</td>
<td>U733@DOM1</td>
<td>NTLM</td>
<td>Success</td>
<td>140</td>
</tr>
<tr>
<td>Oct 29, 2017 02:41:59 AM</td>
<td>C798</td>
<td>U66@DOM1</td>
<td>U714@DOM1</td>
<td>NTLM</td>
<td>Success</td>
<td>150</td>
</tr>
<tr>
<td>Oct 29, 2017 02:41:59 AM</td>
<td>C1096</td>
<td>U66@DOM1</td>
<td>U714@DOM1</td>
<td>NTLM</td>
<td>Success</td>
<td>151</td>
</tr>
</tbody>
</table>
Applying today’s lesson in your enterprise

- Start collecting event logs in your enterprise
  - Authentication logs
  - DNS logs, Netflow logs, ...

- Learn a classifier
  - Collect a labeled data set
  - Extract features
  - Learn a classifier and validate the classifier

- Apply the classifier to future authentication events
  - Flag the identified events for further examination
Related work

- Data set
  - https://csr.lanl.gov/data/cyber1/

- Data Breaches, Phishing, or Malware? Understanding the Risks of Stolen Credentials, Thomas et al., ACM Conference on Computer and Communications Security (CCS), Nov 2017, Dallas, TX.

THANK YOU!

manadhata@alumni.cmu.edu
m.kim@microfocus.com