TOWARDS PREDICTING CYBER ATTACKS USING INFORMATION EXCHANGE AND DATA MINING

Tuesday 26th June, 2018

Martin Husák
Jaroslav Kašpar

CSIRT-MU
Introduction

Information Exchange
- From collaborative intrusion detection to sharing expertise
- Numerous alert sharing platforms and communities

Predictions and Early Warnings
- Common attackers follow certain patterns
- Attack progression – from reconnaissance to intrusion
- Address space patterns – large scans, worm infections, etc.
- Leveraging such knowledge is a subject of research
Approach

Data Mining

- Sequential rule mining
- TopKRules algorithm implemented in SPMF library
- Top-10 sequential rules mined every day for one week

Research Question?

- Comparison of mined rules – are they the same or different?
- How does their support and confidence values evolve?
- How much time does a prediction rule leave for reaction?
Experiment Setup

SABU Alert Sharing Platform

- Originated in academic networks of Czech Republic
- Contributors from academia, public and private sectors

Dataset

- 1,100,000 alerts collected over 1 week from 22 organizations
- Honeypots and network-based IDS as alert sources
- 220,000 alerts per day
- 130,000 attack sequences per day
Example of an Alert

```json
{
    "Format": "IDEA0",
    "ID": "3ad275e3-559a-45c0-8299-6807148ce157",
    "DetectTime": "2014-03-22T10:12:56Z",
    "Category": ["Recon.Scanning"],
    "ConnCount": 633,
    "Description": "Ping scan",
    "Source": [
        {
            "IP4": ["93.184.216.119"],
            "Proto": ["icmp"]
        }
    ],
    "Target": [
        {
            "Proto": ["icmp"],
            "IP4": ["93.184.216.0/24"],
            "Anonymised": true
        }
    ]
}
```
Illustrative Results

**SSH Brute-forcing in multiple networks**

Organization_A.kippo:Attempt.Login:22,
Organization_B.cowrie:Attempt.Login:22

=> Organization_C.kippo:Attempt/Login:22
#SUPP: 0.00367 #CONF: 0.54545

**Network scanning followed by exploitation**

Organization_A.dionaea1:Recon.Scanning:139

#SUPP: 0.00551 #CONF: 0.9

Organization_A.dionaea2:Recon.Scanning:139

#SUPP: 0.00613 #CONF: 0.83333
# Top-10 sequential rules - support and confidence

<table>
<thead>
<tr>
<th>Rule</th>
<th>Input</th>
<th>Output</th>
<th>Support</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Org_A.tarpit:Recon.Scanning:2323, Org_A.nemea.hoststats:Recon.Scanning::None</td>
<td>Org_A.tarpit:Recon.Scanning:23</td>
<td>0.00438</td>
<td>0.88386</td>
</tr>
<tr>
<td>2</td>
<td>Org_A.nemea.bruteforce:Attempt.Login:23</td>
<td>Org_A.tarpit:Recon.Scanning:23</td>
<td>0.00824</td>
<td>0.53465</td>
</tr>
<tr>
<td>3</td>
<td>Org_A.nemea.hoststats:Recon.Scanning:None</td>
<td>Org_A.hoststats:Recon.Scanning:None</td>
<td>0.01987</td>
<td>0.68214</td>
</tr>
<tr>
<td>4</td>
<td>Org_A.tarpit:Recon.Scanning:2323</td>
<td>Org_A.tarpit:Recon.Scanning:23</td>
<td>0.06655</td>
<td>0.70099</td>
</tr>
<tr>
<td>5</td>
<td>Org_A.tarpit:Recon.Scanning:2222</td>
<td>Org_A.tarpit:Recon.Scanning:22</td>
<td>0.00834</td>
<td>0.58155</td>
</tr>
<tr>
<td>7</td>
<td>Org_A.nemea.hoststats:Recon.Scanning:None, Org_B.nemea.hoststats:Recon.Scanning:None</td>
<td>Org_A.hoststats:Recon.Scanning:None</td>
<td>0.00544</td>
<td>0.80088</td>
</tr>
<tr>
<td>9</td>
<td>Org_A.hoststats:Recon.Scanning:None, Org_B.nemea.hoststats:Recon.Scanning:None</td>
<td>Org_A.nemea.hoststats:Recon.Scanning:None</td>
<td>0.00411</td>
<td>0.60284</td>
</tr>
</tbody>
</table>
Support and confidence values of Top-10 sequential rules during the experiment

<table>
<thead>
<tr>
<th>Rule</th>
<th>Day 1 (133,785 seq.)</th>
<th>Day 2 (129,180 seq.)</th>
<th>Day 3 (137,364 seq.)</th>
<th>Day 4 (140,093 seq.)</th>
<th>Day 5 (140,844 seq.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00438</td>
<td>0.88386</td>
<td>0.00544</td>
<td>0.89453</td>
<td>0.00468</td>
</tr>
<tr>
<td>2</td>
<td>0.00824</td>
<td>0.53465</td>
<td>0.00955</td>
<td>0.54844</td>
<td>0.00750</td>
</tr>
<tr>
<td>3</td>
<td>0.01987</td>
<td>0.68214</td>
<td>0.02789</td>
<td>0.76877</td>
<td>0.02637</td>
</tr>
<tr>
<td>4</td>
<td>0.06655</td>
<td>0.70099</td>
<td>0.06864</td>
<td>0.71114</td>
<td>0.06246</td>
</tr>
<tr>
<td>5</td>
<td>0.00834</td>
<td>0.58155</td>
<td>0.00818</td>
<td>0.58045</td>
<td>0.00708</td>
</tr>
<tr>
<td>6</td>
<td>0.00487</td>
<td>0.89071</td>
<td>0.00557</td>
<td>0.87378</td>
<td>0.00537</td>
</tr>
<tr>
<td>7</td>
<td>0.00544</td>
<td>0.80088</td>
<td>0.00587</td>
<td>0.89504</td>
<td>0.00546</td>
</tr>
<tr>
<td>8</td>
<td>0.00289</td>
<td>0.9</td>
<td>0.00129</td>
<td>0.78403</td>
<td>0.00138</td>
</tr>
<tr>
<td>9</td>
<td>0.00411</td>
<td>0.60284</td>
<td>0.00414</td>
<td>0.62941</td>
<td>0.00397</td>
</tr>
<tr>
<td>10</td>
<td>0.00266</td>
<td>0.83962</td>
<td>0.00412</td>
<td>0.87070</td>
<td>0.00355</td>
</tr>
</tbody>
</table>
Evolution of support (left) and confidence (right) values in sequential rules in consecutive day
### Top-10 sequential rules – minimal and average time differences (in seconds)

<table>
<thead>
<tr>
<th>Rule</th>
<th>Input</th>
<th>Output</th>
<th>Min. $\Delta t$</th>
<th>Avg. $\Delta t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Org_A.nemea.hoststats:Recon.Scanning::None</td>
<td>Org_A.hoststats:Recon.Scanning::None</td>
<td>1</td>
<td>401</td>
</tr>
<tr>
<td>7</td>
<td>Org_A.nemea.hoststats:Recon.Scanning::None, Org_B.nemea.hoststats:Recon.Scanning::None</td>
<td>Org_A.hoststats:Recon.Scanning::None</td>
<td>4</td>
<td>735</td>
</tr>
<tr>
<td>9</td>
<td>Org_A.hoststats:Recon.Scanning::None, Org_B.nemea.hoststats:Recon.Scanning::None</td>
<td>Org_A.nemea.hoststats:Recon.Scanning::None</td>
<td>1</td>
<td>2,698</td>
</tr>
</tbody>
</table>
Conclusion and Future Work

Conclusion

- Examination of real-world security alerts and possibility of attack prediction in collaborative environment
- Mined sequential rules are stable over time
- Many rules are unfit for practical use – proper (manual) filtering is recommended
- The rules leave enough time to react (often in order of minutes)

Future Work

- Further development of the prediction component of SABU
- Visualization of the mined rules
THANK YOU FOR YOUR ATTENTION!

(csirt.muni.cz)

@csirtmu

Martin Husák
husakm@ics.muni.cz