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 of different?
- How does their support and confidence values evolve?
- How much time does a prediction rule leave for reaction?

CSIRT-MU

Experiment Setup

SABU Alert Sharing Platform

- Originated in academic networks of Czech Republic
- Contributors from academia, public and private sectors
- `https://sabu.cesnet.cz/en/start`

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


```
{
   "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
      }
   ]
}
```


CSIRT-MU

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
   => Organization_A.dionaea1:Attempt.Exploit:445
   #SUPP: 0.00551 #CONF: 0.9

Organization_A.dionaea2:Recon.Scanning:139
   => Organization_A.dionaea2:Attempt.Exploit:445
   #SUPP: 0.00613 #CONF: 0.83333
```

Top-10 sequential rules – support and confidence

Rule	Input		Output	Support	Confidence
1	Org_A.tarpit:Recon.Scanning:2323, Org_A.nemea.hoststats:Recon.Scanning::None	⇒	Org_A.tarpit:Recon.Scanning:23	0.00438	0.88386
2	Org_A.nemea.bruteforce:Attempt.Login:23	⇒	Org_A.tarpit:Recon.Scanning:23	0.00824	0.53465
3	Org_A.nemea.hoststats:Recon.Scanning:None	⇒	Org_A.hoststats:Recon.Scanning:None	0.01987	0.68214
4	Org_A.tarpit:Recon.Scanning:2323	⇒	Org_A.tarpit:Recon.Scanning:23	0.06655	0.70099
5	Org_A.tarpit:Recon.Scanning:2222	⇒	Org_A.tarpit:Recon.Scanning:22	0.00834	0.58155
6	Org_A.tarpit:Recon.Scanning:2323, Org_A.hoststats:Recon.Scanning:None	⇒	Org_A.tarpit:Recon.Scanning:23	0.00487	0.89071
7	Org_A.nemea.hoststats:Recon.Scanning:None, Org_B.nemea.hoststats:Recon.Scanning:None	⇒	Org_A.hoststats:Recon.Scanning:None	0.00544	0.80088
8	Org_A.hoststats:Recon.Scanning:None, Org_A.tarpit:Recon.Scanning:443	⇒	Org_A.tarpit:Recon.Scanning:80	0.00289	0.90000
9	Org_A.hoststats:Recon.Scanning:None, Org_B.nemea.hoststats:Recon.Scanning:None	⇒	Org_A.nemea.hoststats:Recon.Scanning: None	0.00411	0.60284
10	Org_A.tarpit:Recon.Scanning:2323, Org_A.hoststats:Recon.Scanning:None, Org_A.nemea.hoststats:Recon.Scanning:None	⇒	Org_A.tarpit:Recon.Scanning:23	0.00266	0.83962

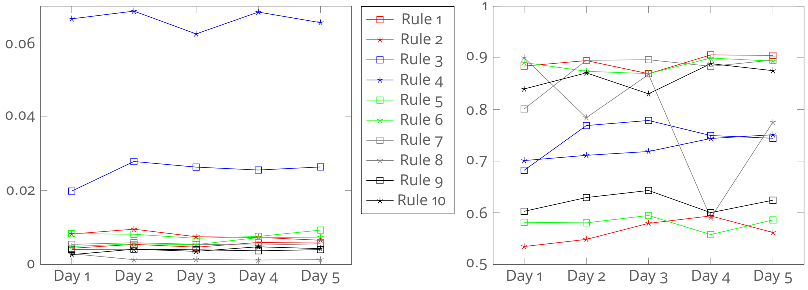
CSIRT-MU

Support and confidence values of Top-10 sequential rules during the experiment

Rule	**Day 1** (133,785 seq.)		**Day 2** (129,180 seq.)		**Day 3** (137,364 seq.)		**Day 4** (140,093 seq.)		**Day 5** (140,844 seq.)	
	Supp.	**Conf.**	**Supp.**	**Conf.**	**Supp.**	**Conf.**	**Supp.**	**Conf.**	**Supp.**	**Conf.**
1	0.00438	0.88386	0.00544	0.89453	0.00468	0.86909	0.00595	0.90554	0.00580	0.90476
2	0.00824	0.53465	0.00955	0.54844	0.00750	0.57953	0.00733	0.59387	0.00655	0.56178
3	0.01987	0.68214	0.02789	0.76877	0.02637	0.77863	0.02558	0.74947	0.02641	0.74415
4	0.06655	0.70099	0.06864	0.71114	0.06246	0.71855	0.06838	0.74378	0.06551	0.75104
5	0.00834	0.58155	0.00818	0.58045	0.00708	0.59474	0.00758	0.55777	0.00930	0.58606
6	0.00487	0.89071	0.00557	0.87378	0.00537	0.86925	0.00727	0.89938	0.00739	0.89356
7	0.00544	0.80088	0.00587	0.89504	0.00546	0.89618	0.00524	0.88341	0.00559	0.89545
8	0.00289	0.9	0.00129	0.78403	0.00138	0.86758	0.00119	0.59011	0.00130	0.77542
9	0.00411	0.60284	0.00414	0.62941	0.00397	0.64311	0.00369	0.60023	0.00401	0.62431
10	0.00266	0.83962	0.00412	0.87070	0.00355	0.83022	0.00478	0.88859	0.00427	0.875

CSIRT-MU

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)

Rule	Input		Output	Min. △t	Avg. △t
1	Org_A.tarpit:Recon.Scanning:2323, Org_A.nemea.hoststats:Recon.Scanning::None	⇒	Org_A.tarpit:Recon.Scanning:23	12	1,530
2	Org_A.nemea.bruteforce:Attempt.Login:23	⇒	Org_A.tarpit:Recon.Scanning:23	121	7,539
3	Org_A.nemea.hoststats:Recon.Scanning:None	⇒	Org_A.hoststats:Recon.Scanning:None	1	401
4	Org_A.tarpit:Recon.Scanning:2323	⇒	Org_A.tarpit:Recon.Scanning:23	901	5,882
5	Org_A.tarpit:Recon.Scanning:2222	⇒	Org_A.tarpit:Recon.Scanning:22	914	7,041
6	Org_A.tarpit:Recon.Scanning:2323, Org_A.hoststats:Recon.Scanning:None	⇒	Org_A.tarpit:Recon.Scanning:23	21	2,019
7	Org_A.nemea.hoststats:Recon.Scanning:None, Org_B.nemea.hoststats:Recon.Scanning:None	⇒	Org_A.hoststats:Recon.Scanning:None	4	735
8	Org_A.hoststats:Recon.Scanning:None, Org_A.tarpit:Recon.Scanning:443	⇒	Org_A.tarpit:Recon.Scanning:80	35	22,754
9	Org_A.hoststats:Recon.Scanning:None, Org_B.nemea.hoststats:Recon.Scanning:None	⇒	Org_A.nemea.hoststats:Recon.Scanning: None	1	2,698
10	Org_A.tarpit:Recon.Scanning:2323, Org_A.hoststats:Recon.Scanning:None, Org_A.nemea.hoststats:Recon.Scanning:None	⇒	Org_A.tarpit:Recon.Scanning:23	12	1,528

CSIRT-MU

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

CSIRT-MU

THANK YOU FOR YOUR ATTENTION!

csirt.muni.cz

@csirtmu

Martin Husák

husakm@ics.muni.cz