Outline

• The Cooperation of CZ.NIC and III
• Statistics Result of Malware
• Malware Family Classiffication
• Malicious Domain Analysis
The Cooperation of CZ.NIC and III

- Implement HaaS Project from 1. June 2016 to 31. May 2018
- CZ.NIC provides malware samples to III
Statistics

**honeypot type**

- **52%** dionaea
- **48%** dionaea

**OS**

- Renesas SH
- x86-64
- PowerPC or cisco 4500
- ARM
- MIPS
- ASCII text executable
- Intel 80386

**Hard coded IP**

<table>
<thead>
<tr>
<th>IP</th>
<th>Count</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>192.168.1.101</td>
<td>383</td>
<td>Private</td>
</tr>
<tr>
<td>62.4.24.135</td>
<td>32</td>
<td>French</td>
</tr>
<tr>
<td>62.4.24.135</td>
<td>27</td>
<td>French</td>
</tr>
<tr>
<td>163.172.18.61</td>
<td></td>
<td>UK</td>
</tr>
<tr>
<td>8.8.8.8</td>
<td></td>
<td>USA google DNS</td>
</tr>
</tbody>
</table>
Map the category with VT information

- Extract Indicators of Compromise (IoC) from malware samples and obtain more information through VirusTotal (VT)

[ref] https://www.virustotal.com/#/home/upload
### File Identification

#### MD5

#### SHA1

#### SHA256

#### Ssdeep

24: vcyEcorpMvPWHs4rlHNUGA/dWJa3Snu3eTsm:vdEorq+M4rONUGdUa3C8R

#### File size

1.6 KB (1600 bytes)

#### File type

unknown

#### Magic literal

Bourne-Again shell script text executable

#### TriID

Linux/UNIX shell script (100.0%)

### VirusTotal Metadata

#### First submission

2017-04-22 04:42:53 UTC (November, 1 week ago)

#### Last submission

2017-04-22 04:42:53 UTC (November, 1 week ago)
### Virus Total Analysis

- **SHA256:**
- **File name:** Bins.sh
- **Detection rate:** 22 / 56
- **Analysis date:** 2017-05-18 02:10:30 UTC (October, 1 week ago)

#### Antivirus Results

<table>
<thead>
<tr>
<th>Antivirus</th>
<th>Result</th>
<th>Updated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad-Aware</td>
<td>Trojan.Downloader.BashAgent.TX</td>
<td>20170518</td>
</tr>
<tr>
<td>AegisLab</td>
<td>Troj.Downloader.Shell OC</td>
<td>20170518</td>
</tr>
<tr>
<td>ALYac</td>
<td>Trojan.Downloader.BashAgent.TX</td>
<td>20170518</td>
</tr>
<tr>
<td>Arcabit</td>
<td>Trojan.Downloader.BashAgent.TX</td>
<td>20170518</td>
</tr>
<tr>
<td>Avast</td>
<td>BV:Downloader-1B [Dnr]</td>
<td>20170518</td>
</tr>
<tr>
<td>AVG</td>
<td>Linux/Downloader.CP</td>
<td>20170518</td>
</tr>
<tr>
<td>BitDefender</td>
<td>Trojan.Downloader.BashAgent.TX</td>
<td>20170517</td>
</tr>
<tr>
<td>Cyren</td>
<td>Trojan.MAQB-3</td>
<td>20170518</td>
</tr>
<tr>
<td>DrWeb</td>
<td>Linux.DownLoader.275</td>
<td>20170518</td>
</tr>
</tbody>
</table>
However...
SHA256:  
File name: Index.html
Detection rate: 0 / 56  
Analysis date: 2018-01-24 09:36:59 UTC (2 months ago)

File identification

MD5
SHA1
SHA256
Sadeep 48: nSZLa5BNWvcmYV8KbgnrerxPerorqQklMxPzI6mX8mEtHmczmIhUeSUsoXXnRwL:
MLYBNyKYYrerxPerorqOfxLtSmX8mz
File size 2.9 KB (3001 bytes)
File type HTML
Magic literal HTML document text
TrID HyperText Markup Language with DOCTYPE (80.8%)
HyperText Markup Language (19.3%)
Tags Html

VirusTotal metadata
First submission 2018-01-24 09:36:59 UTC (2 months ago)
Last submission 2018-01-24 09:36:59 UTC (2 months ago)
File name Index.html
Malware Families Classifier
Malware Family Classification

- Extract n-gram of binary bytes and transform them to computed features via TF-IDF.
- Adapt Symantec naming rules as malware families label.
- Show roughly malware families distribution in the way of projection by TSNE.
- Find out which classification methods can get best performance of accuracy rate.

[ref] https://www.symantec.com/security-center/virusnaming
Get malware labels from VT

Top-10 malware in cowrie with Symantec naming rules

- Downloader.Trojan - 475
- Linux.Chikdos.Blgen2 - 432
- Trojan.Gen.NPE - 295
- SecurityRisk.gen1 - 232
- Linux.Dofloo - 196
- Trojan.Gen.6 - 180
- Linux.Backdoor.Kaiten - 114
- Linux.Gafgyt - 57
- Linux.Lightaidra - 55
- Trojan.Gen.NPE.2 - 53

VirusTotal-Symantec naming rules

Top-10 malware
Get malware labels from VT

Top-10 malware in dionaea with Symantec naming rules

- W32.Downadup.B - 1529
- Trojan.Gen.2 - 109
- W32.Rahack.W - 102
- W32.Downadup - 101
- Trojan.Gen - 96
- Backdoor.Trojan - 77
- W32.Rahack.H - 58
- W32.Pinfi.B - 31
- W32.IRCBot - 23
- SMG.Heur!gen - 19

71.3%
Get malware labels from VT

Top-10 malware in cowrie + dionaea with Symantec naming rules

- W32.Downadup.B - 1529
- Downloader.Trojan - 475
- Linux.Chikdos.B!gen2 - 432
- Trojan.Gen.NPE - 298
- SecurityRisk.gen1 - 244
- Trojan.Gen.6 - 187
- Trojan.Gen.2 - 121
- Linux.Backdoor.Kaiten - 114
- W32.Rahack.W /102

6.6%
8.1%
11.7%
12.8%
41.3%
TSNE with 1-gram TF-IDF model
Category merging- 
Merge to 6 malware

Top-6 malware in cowrie + dionaea with Symantec naming rules

- W32.Downadup.B - 1529
- Trojan -1195
- Linux.Chikdos.B!gen2 -432
- SecurityRisk.gen1 -244
- Linux.Dofloo -196
- W32.Rahack.W /102
TSNE with 1-gram TF-IDF model
TSNE with 2-gram TF-IDF model
TSNE with 3-gram TF-IDF model
Classifiers comparison

N-grams comparison (accuracy)

<table>
<thead>
<tr>
<th>Classifier</th>
<th>1-gram</th>
<th>2-gram</th>
<th>3-gram</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomforest</td>
<td>0.962</td>
<td>0.954</td>
<td>0.96</td>
</tr>
<tr>
<td>ExtraTree</td>
<td>0.951</td>
<td>0.956</td>
<td>0.962</td>
</tr>
<tr>
<td>DecisionTree</td>
<td>0.96</td>
<td>0.956</td>
<td>0.959</td>
</tr>
<tr>
<td>K-neighbors</td>
<td>0.962</td>
<td>0.859</td>
<td>0.867</td>
</tr>
<tr>
<td>QDA</td>
<td>0.956</td>
<td>0.924</td>
<td>0.956</td>
</tr>
<tr>
<td>SVM</td>
<td>0.932</td>
<td>0.932</td>
<td>0.959</td>
</tr>
<tr>
<td>XGBoost</td>
<td>0.948</td>
<td>0.932</td>
<td>0.959</td>
</tr>
</tbody>
</table>
About the Classifier

- Using only binary bytes information can help to distinguish different types of malware.
- Based on the features of 1-gram from binaries, we can classify malwares to their belonging families with high accuracy. The detection rate is over 95%.
- We can classify the unknown malwares into their belonging families.
Malicious Domain Analysis (1/3)

• Step 1: Run dynamic analysis for each malware sample.
• Step 2: Extract the URL list from network traffic logs
• Step 3: Extend the C2 via Ziffer system and get further understanding about each URL.
Malicious Domain Analysis (2/3)

• Explore Malware Distribution of each malicious domain name. The ZifferSystem will output the Malware Distribution graph and distribution information.
Malicious Domain Analysis (3/3)

- 334 malicious domain groups have been generated.
- 17 malicious domain groups were exist in our threat intelligence platform.
- An example of a malicious domain group:
THANK YOU