28th ANNUAL FIRST CONFERENCE SEOUL JUNE 12 - 17, 2016

GETTING TO THE SOUL OF INCIDENT RESPONSE
Current AVs only detect 70% of web shells
web shell classification at scale
About me

Thomas Kastner, MSc.

nimbussec

- Fast detection of hacked websites

SBA Research

- Research center for information security
Definition: web shell

A web shell is a **script** that can be uploaded to a **web server** to enable **remote administration** of the machine. ...

A web shell can be written in **any language** that the target web server supports. The most commonly observed web shells are written in languages that are widely supported, such as **PHP** and **ASP**. Perl, Ruby, Python, and Unix shell scripts are also used. ...

US-CERT TA15-314A
Detection is difficult

Due to the potential simplicity and ease of modification of web shells, they can be difficult to detect. For example, anti-virus products sometimes produce poor results in detecting web shells. …
Detection is difficult

Webshell detection rate

- AV 1: 60.65%
- AV 2: 60.69%
- AV 3: 71.42%
NeoPI & Statistics
NeoPI & Statistics

Statistical methods to detect obfuscated content

- Index of Coincidence
- Entropy
- Compression
- Longest Word
- Poison words
Real World Data

2013

AVs 50%  NeoPI 70%
Real World Data

2013

True positive rate

False positive rate

90%

70%

5%

30%
What we learned so far

NeoPI was meant for human analysts

Search for thresholds via grid search

- Marketing called it already machine learning
Support Vector Machine
<?php
    eval(gzip('...')
?>
<?php

eval(base64_decode(
"bWFrZSBuaW1idXNlYyBncmVhdCBpbg=="));

function ( function ( string ) );
?>
function ( function ( string ) );
function ( function
  ( function ( function ( string
    ( string )
  string ) )
) )

131_6 = 55
313_6 = 117
132_6 = 56
324_6 = 124
244_6 = 100
445_6 = 173
Real World Data

2015

AVs  65%  SVM  85%
Real World Data

99.9%

Accuracy

5-fold cross-validation

2015
Real World Data

2015

MY NOSE WILL GROW NOW!
Accuracy paradox

Benign samples: 2,500,000

Malicious samples: 3,000

\[
\text{Acc} = \frac{TN + TP}{TN + TP + FN + FP}
\]

\[
\text{Acc} = \frac{2500000 + 0}{2500000 + 0 + 3000 + 0} = 99.8\%
\]
What we learned

Forget accuracy (and most other metrics)

- use TPR and FPR instead
k-Nearest Neighbor
$k = 3$
k-Nearest Neighbors

Distance metrics:

- Euclidean distance
- Hamming distance

Select $k$ by hand or via heuristic

Take distance into account

- weight = $1 / \text{distance}$
SVM

k-NN

malicious

suspicious

benign
Benchmark

Intel Core i5-3340M CPU @ 2.70GHz
2x8GB DDR3/1600 SODIMM
(SAMSUNG SSD SM841 256GB)

35,000 files
12,000 infected
Real World Data

2016

- True positive rate: 83%
- False positive rate: 0.03%
- Positive rate: 0%
- False positive rate: 95%
What we learned

16,500,000 classifications / day

Treat training set (web shells) like code

- Version control
- Unit tests
What we learned

Quality of training set

<table>
<thead>
<tr>
<th></th>
<th>1 human</th>
<th>2 humans</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPR</td>
<td>91%</td>
<td>95%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1 human</th>
<th>2 humans</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPR</td>
<td>2%</td>
<td>0.03%</td>
</tr>
</tbody>
</table>
Statistics and heuristics
Support Vector Machine
k-Nearest Neighbor

2013  2014  2016
2013: 50% to 70%
2016: 70% to 95%
Future improvements

Different strategies for tokenization

New machine learning algorithms

- Deep Learning
- Neural Networks

New frameworks

- Tensor Flow
- DSSTNE
we are currently here

Grossglockner: 3,798 m (12,461 ft)