LongTail, an SSH Brute Force Analyzer

Eric Wedaa, Marist College
Poughkeepsie, New York
ACM Poughkeepsie Chapter,
Poughkeepsie NY Jan 16, 2017
Eric.Wedaa@marist.edu
Http://longtail.it.marist.edu/presentations
Data in this presentation is from Jan 4, 2017
LongTail, an SSH Honeypot and Brute Force Analyzer

LongTail is an SSH honeypot AND a set of programs to analyze the attacks and characterize attacking IP addresses into recognizable botnets.
Presentation Points

1) Introduction
2) SSH and Brute Force Attacks
3) LongTail HoneyPot
4) Level 0, 1, & 2 Analysis:
   - IP Address, Usernames and Passwords
5) Level 3 & 4 Analysis:
   - Attack Patterns and BotNets
6) Conclusion
Part 1, Introduction
I'm Eric Wedaa, and I've been a Unix sysadmin continuously since 1987 (Thank you SunQuest Information Systems).

I've done security work intermittently since 1996 (At Philips Semiconductors at that time).

These days I'm a CyberSecurity Consultant at Marist College in Poughkeepsie, New York and a SysAdmin at Ferguson Enterprises in Goshen, NY.
Marist College
Marist College

Marist College is in Poughkeepsie, New York along the scenic Hudson River.

- Marist was founded in 1905 by the Marist Brothers, and spun off as a private college in 1969.
- We have 6,526 undergrads and graduate students.
- The Princeton Review profiles Marist in its new book, "Colleges That Create Futures: 50 Schools That Launch Careers by Going Beyond the Classroom"
- Marist is best know for the Marist Poll, frequently used by NBC and the Wall Street Journal.
Why Write LongTail?

• Working at Marist I saw a HUGE number (100s/day through our IPS, 1000s/day to an unprotected site) of ssh brute force attempts.

• The attacks predominately came from China and Hong Kong against root.

• I decided to see if I could show, from a provable point of view, that these attacks were all coming from the same botnet or botnets.

• The analysis tools didn't exist to do this level of work, so I had to write my own.
Why Name it LongTail?

- A LongTail is a statistical term for a distribution pattern with a large number of values on the left, tapering down quickly to a large number of small values to the right. This pattern is seen frequently in LongTail graphs of usernames and passwords.

  ![Diagram of LongTail distribution]

Why Name it LongTail, part 2

Also, the Marist College mascot is the Red Fox, an animal known for it's...
Why Name It LongTail, part 2

Long Tail!
Who Else Has Looked At This?

• A Study of Passwords and Methods Used in Brute-Force SSH Attacks, by James P. Owens, Jr.
  – This is a Master's thesis from 2008, running a patched OpenSSH server similar to LongTail. They did not log the client software like LongTail does. A much smaller set of data was collected.
  – Over the course of approximately 13 weeks, the 3 honeypots were subjected to 399 separate attacks, consisting of more than 151,000 login attempts, originating from 349 IP addresses.
What is a Brute Force Attack?

- A brute force attack (also known as an “exhaustive key search”) is an attack that tries password after password after password looking for a match.

- Brute force attacks are free once the blackhats own a server.
Where do Passwords Come From?

- Passwords for attacks come from
  - Some of the “bad password lists”
  - Spell check dictionaries of different languages
  - The above words with “special” characters added
  - Words written in “l33t” speak
  - Random characters
  - Google searches for words
  - Google searches for passwords
Part 3, LongTail Honeypot

A general discussion of honeypots and LongTail's honeypot in particular
Existing Honeypots

- http://securehoney.net/ Last updated December 15, 2014
- http://kojoney.sourceforge.net/ Last updated in 2010
- And there are several others!
LongTail's Honeypot Design Requirements

1) Nobody can actually login to the system through the honeypot.

2) That it be as immune to fingerprinting (discussed later) as possible.

3) Be able to log data to a remote server for consolidation.

4) That it can log the originating IP address, username, password tried, time of attack, and client software used.
5) That it be easy or trivial to install and configure.

6) That it be portable across as many systems as possible. In particular the ability to run on Raspberry Pi or a Beaglebone was a desirable ability.
LongTail's Honeypot

- LongTail's honeypot is a modified version of openssh 6.7
- Source code was edited to ALWAYS fail a password.
- Source code was edited to log:
  - IP address
  - Password
  - Client software
  - Targeted Port Number
  - Source Port
  - Time of day
  - Account
Other Uses of Honeypots

- Besides using a honeypot to track attacks, honeypots can be used to test effectiveness of assorted firewalls, gateways, and intrusion protection systems.
How effective is our IPS?

Our IPS still lets stuff sneak through, but not very much at all!

<table>
<thead>
<tr>
<th>Time Frame</th>
<th>Number of Days</th>
<th>Total SSH attempts</th>
<th>Average Per Day</th>
<th>Std. Dev.</th>
<th>Median</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>erhp So Far Today</td>
<td>1</td>
<td>204</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>erhp This Month</td>
<td>3</td>
<td>950</td>
<td>316.67</td>
<td>189.31</td>
<td>265</td>
<td>570</td>
<td>115</td>
</tr>
<tr>
<td>erhp Last Month</td>
<td>31</td>
<td>881</td>
<td>28.42</td>
<td>105.08</td>
<td>0</td>
<td>458</td>
<td>0</td>
</tr>
<tr>
<td>erhp This Year</td>
<td>3</td>
<td>950</td>
<td>316.67</td>
<td>189.31</td>
<td>265</td>
<td>570</td>
<td>115</td>
</tr>
<tr>
<td>erhp Since Logging Started</td>
<td>690</td>
<td>66,916</td>
<td>96.98</td>
<td>492.33</td>
<td>4</td>
<td>5,383</td>
<td>0</td>
</tr>
<tr>
<td>erhp Normalized Since Logging Started</td>
<td>1</td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fingerprinting
What Is Fingerprinting

• Fingerprinting describes a method of contacting a server and determining it's “fingerprint”.

• The fingerprint can show it is a live system, or it can show that it is a known honeypot.

• If it's a honeypot, the blackhats move on.

• Fingerprinting is done by sending strings of data to a server, and then storing the return data and comparing it against already known fingerprints (like Nmap).
Immune to Fingerprinting?

- Any fingerprint taken will show that it is openSSH 6.7p1.
- Any fingerprint taken will show that it was compiled on a RHEL or CentOS (or whatever) server (based on handshakes and encryption methods).
- It looks “real” because it “is” real.
Part 3, LongTail Basic Analysis

A discussion of LongTail's level 0, 1, and 2 analysis.
Attack Analysis Levels

As part of this project I am proposing a new methodology of ssh brute force analysis.

Levels

By using levels of analysis, different reporting tools can be compared in a consistent manner, by comparing their analysis levels.

LongTail is the only tool that currently performs ALL levels of analysis.
Attack Analysis Levels

0) Where are the attacks coming from?
1) What accounts are being tried?
2) What passwords are being tried?
   • What account/password pairs are prevalent?
3) What attack patterns are there?
4) What botnets are out there?
Let's Look At The Live Site!

Http://longtail.it.marist.edu/honey -Quick Stats
Level 0, Where are the attacks coming from?
Quick Stats, Levels 0, 1 & 2

<table>
<thead>
<tr>
<th>Description</th>
<th>Today</th>
<th>Never seen Before Today</th>
<th>This Month</th>
<th>This Year</th>
<th>Since Logging Started</th>
</tr>
</thead>
<tbody>
<tr>
<td>Login Attempts</td>
<td>78,346</td>
<td>N/A</td>
<td>504,851</td>
<td>504,851</td>
<td>82,734,249</td>
</tr>
<tr>
<td>SSHPsycho Login Attempts [3]</td>
<td>0</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
<td>13,283,885</td>
</tr>
<tr>
<td>SSHPsycho-2 Login Attempts [7]</td>
<td>0</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
<td>27,003,767</td>
</tr>
<tr>
<td>Unique Passwords</td>
<td>27,836</td>
<td>2</td>
<td>41,138</td>
<td>41,138</td>
<td>1,414,381</td>
</tr>
<tr>
<td>Unique Usernames</td>
<td>312</td>
<td>1</td>
<td>420</td>
<td>420</td>
<td>28,956</td>
</tr>
<tr>
<td>Unique IPs</td>
<td>233</td>
<td>158</td>
<td>914</td>
<td>914</td>
<td>36,707</td>
</tr>
<tr>
<td>Active Honeypots</td>
<td>10</td>
<td></td>
<td>9</td>
<td>9</td>
<td>32</td>
</tr>
</tbody>
</table>
Graphs, levels 1 & 2

**Today**
- Current Top 20 Root Passwords

**Prior 7 Days**
- Prior 7 Days Top 20 Root Passwords
  - http://longtail.it.marist.edu 2017-01-04 02:44

**Current Top 20 Username "admin" Passwords**

**Prior 7 Days Top 20 Username "admin" Passwords**
- http://longtail.it.marist.edu 2017-01-04 02:44
Graphs, levels 1 & 2
Daily Trends, levels 1, 2, & 2a

- Top 20 Accounts
- Top 20 Root Passwords
- Top 20 Account "admin" Passwords
- Top Non 20 Root/Account "admin" Passwords
Top 20 Root Passwords

<table>
<thead>
<tr>
<th>Date</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017/01/03</td>
<td>174:root</td>
<td>165:12345</td>
<td>164:admin</td>
<td>152:123456</td>
<td>144:000000</td>
</tr>
</tbody>
</table>
Today's Activity

- Today's IP Address Activity
- Today's Usernames
- Today's Passwords
Today's IP Address Activity

- IP Addresses FIRST seen today
- IP Addresses
- Top 20 IP Addresses
- Attacks By Country, By IP Address
Today's Usernames

- Accounts FIRST seen today
- Accounts Tried
- Top 20 Accounts Tried
- Non "root"/"admin" pairs (Restricted Access)
- Top 20 Non "root"/"admin" pairs
Today's Passwords

- Passwords FIRST seen today
- "root" Passwords (Restricted Access)
- Top 20 "root" Passwords
- Account "admin" Passwords (Restricted Access)
- Top 20 Account "admin" Passwords
- Non "root"/"admin" Passwords (Restricted Access)
- Top 20 Non "root"/"admin" Passwords
This is the interesting part, how does LongTail group attacks into attack patterns and then into botnets.

This is where I perform analysis levels 3 & 4.
Attack Patterns

- An attack pattern is:
  - A list of accounts and passwords
  - From a single IP
  - Against a single IP
  - Where each attempt is within 180 seconds of the last attempt.
  - Sorted by account/password (without 'unique-ing') to counteract timing variations in reporting to rsyslog, multiple sends of the same pair, and attempted randomization by the botnet.
Sample Attack Pattern

root ->default<-
root ->freenas<-
root ->goflex<-
root ->live<-
root ->nosoup4u<-
root ->openelec<-
Attack Patterns
(Based of Jan 4, 2017, using only 2016 and 2017 data)

- The largest attack pattern so far is 261,276 entries.
- The median attack pattern so far is 406 entries, the mode is 15.
- Out of 132,238 attack patterns recorded, there are 35,132 attack patterns that repeat.
- There are 35,132 attack patterns that consist of more than one entry.
- There are 14,780 attack patterns that consist of more than three entries.
Grouping IPs Into Botnets

- After attack patterns are created, a word count of the attack pattern and an MD5 checksum of the attack pattern are created.

- While LongTail uses MD5 checksums, and stores the information required in several files, this demonstration uses 2 letter checksums for ease of understanding, and bundles all of the information into a single line in a table.
How Else Are Botnets The Same

- Botnets tend to attack the same accounts.
- Botnets tend to use the same client software.
- Botnets tend to use the same “styles” of passwords.
- Botnets tend to be geographically similar, either the same country or region (fromage_puant), or are widely distributed (big_botnet).
Characterization of Botnets

• Botnets do not seem to have strong characteristics of:
  – Time of day attacking
  – Day of week attacking

• Strong characteristics of botnets that LongTail has found are:
  – Accounts Tried
  – Client ssh software
  – Originating countries
  – Open Ports and Operating Systems
Botnets Found
(Sept 15, 2015)

BotNet Totals
(Not counting sshPsycho, sshPsycho-2, friends and associates)

Total number of botnets known: 127
Minimum attack size from all BotNets: 1
Average attack size from all BotNets: 97.04
Maximum attack size from all BotNets: 14,288
## Favorite Botnets (Sept 15, 2015)

<table>
<thead>
<tr>
<th>Name</th>
<th># of Bots</th>
<th>#/attempts</th>
<th>Min size</th>
<th>Avg. size</th>
<th>Max Size</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big_botnet</td>
<td>3,139</td>
<td>29,206</td>
<td>1</td>
<td>70</td>
<td>3.52</td>
<td>Bahrain, India, Brazil, others</td>
</tr>
<tr>
<td>Pink Roses</td>
<td>379</td>
<td>16,785</td>
<td>1</td>
<td>11.71</td>
<td>73</td>
<td>Widely distributed</td>
</tr>
<tr>
<td>Fromage Puant</td>
<td>64</td>
<td>76,997</td>
<td>1</td>
<td>553.79</td>
<td>14,288</td>
<td>France, others</td>
</tr>
<tr>
<td>Dead China</td>
<td>45</td>
<td>41,185</td>
<td>1</td>
<td>219.07</td>
<td>1,678</td>
<td>China</td>
</tr>
<tr>
<td>15-08-04-botnet-40</td>
<td>38</td>
<td>258</td>
<td>1</td>
<td>3.97</td>
<td>6</td>
<td>Bahrain, etc</td>
</tr>
<tr>
<td>Small bots 3</td>
<td>36</td>
<td>1,355</td>
<td>1</td>
<td>9.34</td>
<td>27</td>
<td>US, Vietnam, others</td>
</tr>
</tbody>
</table>
Finding Command And Control

- LongTail has a website where the analysis is displayed.
- Webpages are clearly named as including passwords and usernames.
- Most of these pages are restricted to Marist only, but the links are indexed by Google.
- IP Addresses that perform GETs against only those webpages but do not attempt to get any other pages are HIGHLY suspect.
By scanning the Apache httpd access logs, LongTail can determine which IP addresses are only having “403” page requests without ever performing a “200” page request.
LongTail Log Analysis @ Marist College / IP Addresses Looking For Password Files Only

This page is updated daily. Last updated on Wed Jan 4 03:55:01 EST 2017

Please see the notes in the source code for this script for details on how this analysis is done.

<table>
<thead>
<tr>
<th>IP Address</th>
<th>Country</th>
<th>Number of ssh login attempts from this IP address</th>
<th>Number Of Password Pages Requested</th>
<th>Count of all non-403 pages requested</th>
<th>Last date seen</th>
</tr>
</thead>
<tbody>
<tr>
<td>136.243.17.161</td>
<td>Germany</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>2017/Jan/01 08:41:34</td>
</tr>
<tr>
<td>5.9.112.6</td>
<td>Germany</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>2016/Dec/30 15:59:01</td>
</tr>
<tr>
<td>58.246.148.214</td>
<td>China</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2016/Oct/10 14:49:47</td>
</tr>
<tr>
<td>185.158.135.205</td>
<td>undefined</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2016/Oct/09 04:51:35</td>
</tr>
<tr>
<td>79.155.83.117</td>
<td>Spain</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2016/Nov/29 13:42:19</td>
</tr>
<tr>
<td>77.59.131.225</td>
<td>Switzerland</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2016/Oct/13 13:00:22</td>
</tr>
</tbody>
</table>
Summary and Conclusion