Updates

• Abstracts are due on Friday
 Updates

• Abstracts are due on Friday
  • Via Canvas, at the beginning of class.
Updates

• Abstracts are due on Friday
  • Via Canvas, at the beginning of class.
  • Spelling, grammar and presentation matter.
Updates

• Abstracts are due on Friday
  • Via Canvas, at the beginning of class.
  • Spelling, grammar and presentation matter.
• The semester is blasting past!
Updates

• Abstracts are due on Friday
  • Via Canvas, at the beginning of class.
  • Spelling, grammar and presentation matter.
• The semester is blasting past!
  • Our final is on 12/13.
Updates

• Abstracts are due on Friday
• Via Canvas, at the beginning of class.
• Spelling, grammar and presentation matter.
• The semester is blasting past!
• Our final is on 12/13.
  • It will be cumulative - everything is fair game, but the second half of the semester will be emphasized.
### Where is Anomaly Detection Useful?

| System | Attack Density P(T) | Detector Flagging Pr(F) | Detector Accuracy Pr(F|T) | True Positives P(T|F) |
|--------|---------------------|-------------------------|--------------------------|----------------------|
| A      | 0.1                 |                         | 0.65                     |                      |
| B      | 0.001               |                         | 0.99                     |                      |
| C      | 0.1                 |                         | 0.99                     |                      |
| D      | 0.00001             |                         | 0.999999                 |                      |

\[
\text{Pr}(B|A) = \frac{\text{Pr}(A|B) \cdot \text{Pr}(B)}{\text{Pr}(A)}
\]
### Where is Anomaly Detection Useful?

| System | Attack Density P(T) | Detector Flagging Pr(F) | Detector Accuracy Pr(F|T) | True Positives P(T|F) |
|--------|---------------------|-------------------------|--------------------------|----------------------|
| A      | 0.1                 | 0.38                    | 0.65                     | 0.171                |
| B      | 0.001               | 0.01098                 | 0.99                     | 0.090164             |
| C      | 0.1                 | 0.108                   | 0.99                     | 0.911667             |
| D      | 0.00001             | 0.00002                 | 0.99999                  | 0.5                  |

\[
Pr(B|A) = \frac{Pr(A|B) \cdot Pr(B)}{Pr(A)}
\]
The Big One
The Big One
Malware

- Software with “malicious intentions” is generally categorized as malware.

  - A theoretical treatise on code that could reproduce itself.

- Countless real examples have followed:
Evolution of Malware

• Malware is generally classified into these categories:
  • **Virus** - generally included as part of an executable file, requires some assistance to infect.
  • **Worm** - similar to a virus, able to self propagate.
  • **Trojan** - infected software, generally do not spread.

• These are not “hard and fast” rules.
Ransomware

- Malware that denies access to a resource until payment is made.
- Encrypting your files, producing a lock screen with a secret password, etc…
- Often infects machines via email with dangerous attachments.
- …so a virus…
- …except for when it’s a worm…
WannaCry

• On May 12, 2017, hundreds of thousands of systems around the world were attacked.
  • NHS in UK, railways, factories, couriers and more…
• Took advantage of a vulnerability in Microsoft SMB protocol.
  • Patch available for two months.
• Vulnerability discovered by NSA…
• Damage estimated in the billions of dollars.
Detection and Evasion
Detection and Evasion

- Malware is most often detected statically:
Detection and Evasion

• Malware is most often detected statically:
  • MD5/SHA256 hashes are commonly used in commercial AVs
Detection and Evasion

- Malware is most often detected statically:
  - MD5/SHA256 hashes are commonly used in commercial AVs
- Tactics to evade such detection have become commonplace:
Detection and Evasion

• Malware is most often detected statically:
  • MD5/SHA256 hashes are commonly used in commercial AVs
• Tactics to evade such detection have become commonplace:
  • Encrypted Malware: Virus is encrypted, and each instance is encrypted with a different key.
Detection and Evasion

- Malware is most often detected statically:
  - MD5/SHA256 hashes are commonly used in commercial AVs
- Tactics to evade such detection have become commonplace:
  - **Encrypted** Malware: Virus is encrypted, and each instance is encrypted with a different key.
  - **Polymorphic** Malware: Encrypted, but the decryption routine is modified in every instance.
Detection and Evasion

• Malware is most often detected statically:
  • MD5/SHA256 hashes are commonly used in commercial AVs

• Tactics to evade such detection have become commonplace:
  • **Encrypted** Malware: Virus is encrypted, and each instance is encrypted with a different key.
  
  • **Polymorphic** Malware: Encrypted, but the decryption routine is modified in every instance.

  • **Metamorphic** Malware: Everything is entirely rewritten.
Detection and Evasion

• Malware is most often detected statically:
  • MD5/SHA256 hashes are commonly used in commercial AVs

• Tactics to evade such detection have become commonplace:
  • **Encrypted** Malware: Virus is encrypted, and each instance is encrypted with a different key.
  • **Polymorphic** Malware: Encrypted, but the decryption routine is modified in every instance.
  • **Metamorphic** Malware: Everything is entirely rewritten.

• Where does the arms race go from here?
A botnet is a network of software robots (bots) run on compromised machines which are administered by command and control (C&C) networks.

- Bot master - the owner/controller of a botnet

What is the advantage to this approach over the others?
• Worms, Trojan horses, backdoors, browser-bugs, etc...

• **Note**: the software on these systems is updated
• **Bot theft**: bot controllers penetrate/"steal" bots.
Statistics (controversial)

• The actual number of bots, the size of the botnets and the activity is highly controversial.
  
• As of 2012: millions of bots
  
• 1/4 of hosts are now part of bot-nets
  
• Growing fast (many more bots)

• **Assertion**: botnets are getting smaller(?!?)
  
• When they become large, they are more likely to be noticed and targeted for takedown.
Botnet Architecture

• An army of compromised hosts ("bots") coordinated via a command and control center (C&C).

“A botnet is comparable to compulsory military service for windows boxes”
-- Bjorn Stromberg
Typical Botnet
Typical Botnet
Typical Botnet

1. Compromise
Typical Botnet

2. Download
Typical Botnet

3. DNS Lookup
Typical Botnet

4. Join
Typical Botnet

5. Command
Typical Botnet
IRC

- 1988 - one-to-many or many-to-many chat (for BBS)
- Client/server -- TCP Port 6667
- Used to report on 1991 Soviet coup attempt
- Channels (sometimes password protected) are used to communicate between parties.
  - Invisible mode (no list, not known)
  - Invite only (must be invited to participate)
- Botnets rarely rely on IRC anymore.
  - Many ISPs block IRC these days.
P2P Botnets

- Bots that rely on centralized communications mechanisms such as IRC are generally easy to attack.
  - Single point of failure for the bad guys...
- Increasingly, botnets have turned to P2P-based architectures to avoid such weaknesses.
  - e.g., Slapper, Phatbot, Conficker
- What are the challenges for a botmaster relying on a P2P architecture?
P2P Botnets

• What advantages do defenders have in this situation?
  • How do communication patterns compare to IRC bots?
  • How do you tell between “legitimate” P2P traffic and that associated with bots?
Mobile devices offer new avenues for botnets.

• With the ability to communicate over multiple (5) interfaces, how does a provider defend against such multi-homed botnets?

• How does this change the game in terms of communications strategies for botmasters?
Campaign: DDoS
Campaign: DDoS
Campaign: DDoS

• Distributed Denial of Service (DDoS)

• With hundreds of thousands of malicious devices under their control, a botmaster can unleash massive torrents of traffic at a target.

• Examples: Unknown vs Estonia, Russia/Georgia, Anonymous vs Scientology, Unknown vs CNN, Unknown vs ...

• What’s the advantage of doing this from a botnet?
Stuxnet?
Stuxnet?

• What was Stuxnet?
• Classification?
Stuxnet?

• What was Stuxnet?
  • Classification?
• What was its goal?
  • How did it try to do this?
Stuxnet?

• What was Stuxnet?
• Classification?
• What was its goal?
• How did it try to do this?
• How was it delivered?
Stuxnet?

- What was Stuxnet?
- Classification?
- What was its goal?
- How did it try to do this?
- How was it delivered?
- Was it effective?
How are researchers learning?

• Honeypots are often used to attract, observe and eventually “dissect” bots.

• A number of recent efforts in this space have actually hijacked active botnets.

• Large portions of these networks have been monitored:
  • ... to learn about the targets of the botnet (and their success in exploiting them).
  • ... to learn about weaknesses in their architecture to use as a means of potentially interfering with the botnet.
  • ... to figure out whether deployed defenses are helping at all.
Campaign: Spam

- Spam: Unsolicited mass emailing, generally attempting to advertise a product (legitimate or otherwise).
  - In the past, has been as high as 90+% of email by volume.
  - Approximately 72% in 2014.
  - This is an economic problem... why?
- Botnets are an excellent platform for spam campaigns.
  - Massive bandwidth for sending messages
  - Many locations for hosting infrastructure.
Spamalytics

• Very little was previously known about the conversion rate of spam.

• Why not?

• **Methodology**: Hijack a botnet, watch what happens.

• Good methodology?

• Issues?
Spamalytics (cont)

- What was learned?
- What can we do in terms of defense?
Spamalytics (cont)

What was learned?

What can we do in terms of defense?

Spamalytics (cont)

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Item</th>
<th>Origin</th>
<th>Affiliate Programs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aracoma Drug</td>
<td>Orange bottle of tablets (pharma)</td>
<td>WV, USA</td>
<td>CIFr</td>
</tr>
<tr>
<td>Combitic Global Caplet Pvt. Ltd.</td>
<td>Blister-packed tablets (pharma)</td>
<td>Delhi, India</td>
<td>GlvMd</td>
</tr>
<tr>
<td>M.K. Choudhary</td>
<td>Blister-packed tablets (pharma)</td>
<td>Thane, India</td>
<td>OLPc</td>
</tr>
<tr>
<td>PPW</td>
<td>Blister-packed tablets (pharma)</td>
<td>Chennai, India</td>
<td>PhEx, Stmul, Trust, CIFr</td>
</tr>
<tr>
<td>K. Sekar</td>
<td>Blister-packed tablets (pharma)</td>
<td>Villupuram, India</td>
<td>WldPh</td>
</tr>
<tr>
<td>Rhine Inc.</td>
<td>Blister-packed tablets (pharma)</td>
<td>Thane, India</td>
<td>RxPrm, DrgRev</td>
</tr>
<tr>
<td>Supreme Suppliers</td>
<td>Blister-packed tablets (pharma)</td>
<td>Mumbai, India</td>
<td>Eva</td>
</tr>
<tr>
<td>Chen Hua</td>
<td>Small white plastic bottles (herbal)</td>
<td>Jiangmen, China</td>
<td>Stud</td>
</tr>
<tr>
<td>Etech Media Ltd</td>
<td>Novelty-sized supplement (herbal)</td>
<td>Christchurch, NZ</td>
<td>Staln</td>
</tr>
<tr>
<td>Herbal Health Fulfillment Warehouse</td>
<td>White plastic bottle (herbal)</td>
<td>MA, USA</td>
<td>Eva</td>
</tr>
<tr>
<td>MK Sales</td>
<td>White plastic bottle (herbal)</td>
<td>WA, USA</td>
<td>GlvMd</td>
</tr>
<tr>
<td>Riverton, Utah shipper</td>
<td>White plastic bottle (herbal)</td>
<td>UT, USA</td>
<td>DrMax, Grow</td>
</tr>
<tr>
<td>Guo Zhonglei</td>
<td>Foam-wrapped replica watch</td>
<td>Baoding, China</td>
<td>Dstn, UltRp</td>
</tr>
</tbody>
</table>

- What was learned?
- What can we do in terms of defense?
Spamalytics (cont)

- What was learned?
- What can we do in terms of defense?
Campaign: Click Fraud

- Click fraud is the revenue generated from clicking on paid-advertising links automatically, without any user desire or interest.

- Who are the adversaries here and what are they after?
  - Publisher (revenue)
  - Competitor (cost)

- Why are botnets used as part of these campaigns?
So What Do We Do?

- Given the magnitude of this problem, how do we fight it?
- We have area and problem... Think about solution and methodology!
- There are two places from which we can try to combat bots:
  - Local Network
  - At or above the ISP level
BotHunter: IDS Dialog Correlation

- Simple Approach: Why not just use an IDS looking for a single signature?
- Detection need not be based on a single event.
  - Knowing something about the structure of communication can potentially help us find our bot.
- So how do they do it?
Circle of Life

- Bots follow a very regular pattern: Scan, Infect, “Egg” Download, Communicate (C&C), Action.
- Why does this reduce false positives?
Simple Bayesian Calculation

• Just as an intuition...

• What is the probability of a false positive in a system?
  • $P(I|A) = 0.001$

• If we rely upon multiple independent indicators that are correlated in time:
  • $P(I|A) \times P(I|A) \times P(I|A) \times P(I|A) \times \ldots \times P(I|A)^n$

• We can reduce the number of false positives by not simply looking for single events.
Components

- BotHunter relies on SCADE and SLADE
- Inbound and outbound traffic scanning for phases 1 and 5
- Find suspicious payloads in intervening phases.

- Deployments:
  - Georgia Tech - four month deployment
  - SRI - one month deployment
Results

• True Positives:
  • Deploy 10 bots in a virtual network (Phatbot, RxBot, GTBot)
  • Overlay it with GT traffic.

• False Positives:
  • GT - Less than 1 per month
  • SRI - 1 in a single month

• Assumptions? Weaknesses?
Ransomware can be detected because it must…

… significantly alter files …

Great - how exactly do you do that?

UF-based startup has taken this concept from research to product!
• Ransomware can be detected because it must…
  
  r file:
  ly do
  s take
CryptoDrop

- Ransomware can be detected because it must...
  - ... significantly alter files ...
- Great - how exactly do you do that?
- UF-based startup has taken this concept from research to product!
From the Network - DNS

- Is this enough?
- What about all the networks that don’t deploy BotHunter?
- What about going after DNS instead?
• A botnet with a single IP address is easy to shut down.

• In response, many bots use Dynamic DNS and quickly move their hosting infrastructure between many IP addresses.

• What can be done now?
Domain Generation Algorithms

- To prevent takedown, bots can change the C&C domain they speak to each day.
- Ok, great. How do we coordinate this?
- HMAC(k, currentdomain) + .com/.org/.net
- Is random good enough?
What problem does the Pleiades paper try to solve?

How well does it do?

What does it miss?
Summary

• Botnets represent the current pinnacle of malware evolution.

• They can be reprogrammed infinitely! This makes them incredibly valuable for many kinds of attacks.

• Where are they not valuable?

• Techniques to identify and shut them down vary:
  • Organization: Detect the life-cycle.
  • ISP: Watch for DNS use, try and determine DGAs.