Lecture Outline

• Reminder: guest lecture Friday by Bill Marczak
  – Zoom link w/ password emailed out tonight
  – If you encounter difficulties, rendezvous via Piazza

• Finish botnet discussion: Pay-per-Install (PPI)

• Project presentations & reports

• Anonymity:
  – Brief look at Tor’s evolution
    • Plus a “teachable moment”
  – Anonymizing data (packet traces)
Pay-Per-Install (PPI)
So far, Torpig has been distributed to its victims as part of Mebroot. Mebroot is a rootkit that takes control of a machine by replacing the system’s Master Boot Record (MBR). This allows Mebroot

Mebroot has no malicious capability per se. Instead, it provides a generic platform that other modules can leverage to perform their malicious actions. In particular, Mebroot provides functionality to manage (install, uninstall, and activate) such additional modules. Immediately after the initial reboot, Mebroot contacts the Mebroot C&C server to obtain malicious modules (5). These modules are
Insights from the Inside:
A View of Botnet Management from Infiltration

Chia Yuan Cho§, Juan Caballero†§, Chris Grier§, Vern Paxson‡§, and Dawn Song§
§UC Berkeley †Carnegie Mellon University ‡ICSI

Abstract

Recent work has leveraged botnet infiltration techniques to track the activities of bots over time, particularly with regard to spam campaigns. Building on our previous success in reverse-engineering C&C protocols, we have conducted a 4-month infiltration of the MegaD botnet, beginning in October 2009. Our

2009. While much of our measurement drew upon our earlier work in reverse-engineering MegaD’s C&C protocol [11] and the cryptographic routines that obfuscate it [12], we also developed additional methods for gathering information about the botnet. We discovered that we could use “Google hacking” to locate additional C&C servers based on fingerprinting the web pages they sup-

An inside view of FireEye’s takedown. On Nov. 6, 2009, FireEye launched a coordinated effort to take down MegaD. The takedown was widely lauded as successful since MegaD’s spam trickled to a halt. However, 16 days later its share of the world’s spam exceeded its 4% pre-takedown level and by Dec. 13 it had climbed to 17% [6].
Installs4Sale.net - надежный сервис по загрузкам, достойный доверия

ПРЕИМУЩЕСТВА

- Быстро осуществляем отгрузку практически в любой регион. Принимаем заказы на ми克斯 стран по вашему выбору.
- Для постоянных клиентов действуют скидки и бонусы в виде дополнительного объема загрузок.
- Договоренности по всем данным и получать индивидуальные условия вы можете в our работе.

КОНТАКТЫ

- 560869831
- 550525933
- info [at] installs4sale.net
GangstaBucks.com - it pays on time!
We pay for all installs!

Join our ranks and by tomorrow you could get your first payout!
Prices are USD *per thousand installs*.
Project Presentations: Logistics

• Held last two weeks of regular Semester
  – I’ll finalize assignments by this weekend
• Aim for ~30 minutes of material
• Split presentation w/ partner ~50/50
• Schedule practice talk w/ me 3+ days prior
  – Should be fully drafted and timed
• Post short context summary to Piazza the morning before
  – Assume the class has read it
Project Presentations: Content

• Introduction framing apt for audience
  – A thoughtful tour of the problem space
    • This is the #1 value take-away for your fellow students
  – What you tackled, why it’s significant
  – Assume audience has read your Piazza summary

• Sketch of related work sufficient to appreciate contribution
  – Will also address some “why didn’t you try X?” questions
  – Frame how other researchers have undertaken evaluations in this space
Project Presentations: Content, con’t

• Your **strategy** for pursuing your research
  – Explain technical undertaking / challenges
  – Explain evaluation methodology

• Frame the “**data**”
  – What does it cover
  – What does it *not* cover
  – What you know about quality/representativeness
  – If you’re doing a security analysis, the “data” is your visibility into what you’re analyzing
    • E.g. source code, black-box binaries, papers
Project Presentations: Content, con’t

• What unexpected issues arose?
  – Emphasize **lessons learned**, not just surprises
    • Can provide **valuable take-aways** for other work

• Preliminary results
  – Bring out what is significant
  – Persuade us
  – **Be thoughtful in data presentation** (see below)
  – **Illuminate** limitations

• What remains
  – For your work
  – Implications / open questions for future work
Presenting Effectively: Slides

- Think creatively
- Make judicious use of color
- Avoid serif fonts
- Avoid overly busy slides
- Avoid “wall of bullets” on slide after slide
- Use animations to engage your audience
  - Keep them from peeking ahead, deciding they got it, and tuning out
  - Focus their attention by emphasizing current discussion point / downplaying non-points
Presenting Effectively: Voice

• Do not read your slides
  – ProTip: short phrases force fill-ins
• Do not read your speaker notes
  – ProTip: try not having any (you won’t have any!)
• Find & deliver genuine energy/enthusiasm
• Vary your tone
  – Glitches are an opportunity, not a problem: respond in the moment
• Find a conversational pace
• (Don’t worry about audience eye contact!)
Project Reports

• Treat the class Projects web page as a CFP
  – CFP = Call For Papers
  – Formatting, deadline requirements are serious
  – Read and deliver on the Writing Pointers
    • https://www.icir.org/vern/cs261n/writing.html

• “Be thoughtful in data presentation (see below)”
Huge amount of “real estate” to convey just one number
(b) Total traffic for 62.34.164.84.
Particularly meaningless to connect categorical points with lines. (“Instances” likely hugely overcounts polymorphic malware)
Plots only 7 (x,y) points … which are discretized. Wasted Y-axis real estate. Not clear why 8-hour bins are appropriate.
Hugely misleading compressed X-axis
Data glitch in second point visually dominates presentation. Straight-line interpolation on log-linear plot can be highly misleading.
Left-hand plot completely dominated by \texttt{IN.failed}.
Right-hand plot just shows that all of \texttt{IN.total} was \texttt{IN.failed}.
Terrible use of “real estate”. Can’t tell anything about details other than single spike in the lowest bin. Much better to use logarithmic X-axis.

Figure 1: Histogram: Given an IP address, how many blacklisted domains use that IP address as a name server?
Conveys just 6 points. Avoid using lines to connect log-scaled values!
Figure 6: Total memory usage of traffic monitoring system.
Misleading Y-axis: highly unlikely that changes between 7.8% and 8.25% are actually at all interesting
<table>
<thead>
<tr>
<th>Trace</th>
<th>% Speculatively executed tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>FB2009</td>
<td>1.22</td>
</tr>
<tr>
<td>FB2010</td>
<td>2.04</td>
</tr>
<tr>
<td>CC.b</td>
<td>1.01</td>
</tr>
<tr>
<td>CC.e</td>
<td>1.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trace</th>
<th>% of tasks that straggled even when they executed locally</th>
</tr>
</thead>
<tbody>
<tr>
<td>FB2009</td>
<td>26.4</td>
</tr>
<tr>
<td>FB2010</td>
<td>39.2</td>
</tr>
<tr>
<td>CC.b</td>
<td>55</td>
</tr>
<tr>
<td>CC.e</td>
<td>56</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trace</th>
<th>% of speculatively executed tasks that were killed</th>
</tr>
</thead>
<tbody>
<tr>
<td>FB2009</td>
<td>77.9</td>
</tr>
<tr>
<td>FB2010</td>
<td>88.6</td>
</tr>
<tr>
<td>CC.b</td>
<td>74.4</td>
</tr>
<tr>
<td>CC.e</td>
<td>48.8</td>
</tr>
</tbody>
</table>

**Large horizontal gaps make it visually a pain to read.** Why 4 tables and not one table with 5 columns?
dfs.datanode.bytes_read

254
27
20
13
14
23
0
14
23334290.78
46668581.56
Highly distracting central gap
Just what do the authors want us to take away from this?
(a) Number of Followers for Cluster Users and a Random Sample of Users
Distribution for categorical data not meaningful

Fig. 7: CDF of user account languages for users in size 2 clusters
An example of good use of plot “real estate”