The Cloud-y Future of Security Technologies

Adam J. O’Donnell, Ph.D.
Director, Cloud Engineering
Immunet, Inc.
The Cloud-y Future of Security Technologies

Adam J. O’Donnell, Ph.D.
Chief Architect, Cloud Technology Group
Sourcefire, Inc.
About Immunet

• Founded in mid-2008 to build next-gen AV
• Funding through Altos Ventures, TechOperators in Nov 2009
• Acquired by SourceFire Dec 2010, announced Jan 2011
About me

• Founded in late-1978 to build next-gen of the family line

• Funding through Guardent, consulting, and NSF GRFP @ Drexel University

• Acquired by Cloudmark in 2005, started Immunet full-time when funded in 2009.
COMPUTER VIRUS SPREADS TO HUMANS!
Virus vs. Anti-Virus, 1980s Style

• Viruses:
  • Count: $10^2$
  • Mutation rate: What mutations?
  • Propagation: sneakernet
Virus vs. Anti-Virus, 1980s Style

- Anti-Virus:
  - Low definition count, updated monthly
  - Mutation rate: What mutations?
  - Propagation: USPS
Virus vs. Anti-Virus, 1990s Style

- Viruses:
  - Count: $10^{3-4}$
  - Mutation rate: Fairly low
  - Propagation: Sneakernet, BBS, Internet
Virus vs. Anti-Virus, 1990s Style

- Anti-Virus:
  - Definitions updated daily to weekly
  - Mutation rate: Business hours response teams
  - Propagation: Sneakernet, BBS, Internet
Virus vs. Anti-Virus, Today

- Viruses:
  - 2000: $5 \times 10^4$
  - 2003: $10^5$
  - 2008: $10^6$
  - Today: $10^7$

- Average in field lifetime: 2 to 3 days.
Virus vs. Anti-Virus, Today

• Anti-Virus:
  • Definitions updated every 5 minutes
  • Mutation rate: Follow the sun response teams
  • Propagation: Internet-only
How do AV firms know what viruses exist?
Sample Sharing Alliances

- Informal groups of AV researchers at firms that agree to share, on a hourly or daily basis, drops of new malware
- Based upon who you know and what samples you regularly have
• 1980’s: Informal sample sharing alliances.
• 1980’s: Informal sample sharing alliances.

• 1990’s: Informal sample sharing alliances.
• 1980’s: Informal sample sharing alliances.
• 1990’s: Informal sample sharing alliances.
• 2000’s: Informal sample sharing alliances.
• 1980’s: Informal sample sharing alliances.
• 1990’s: Informal sample sharing alliances.
• 2000’s: Informal sample sharing alliances.
• 2010’s: Informal sample sharing alliances, some centrally collected logs from the big boys.
Virus Count
End result?

• Analyst teams are overwhelmed with stopping threats days after they disappeared from circulation.

• Current, real world, in field efficacy of AV products is approximately 43% for new malware for generic detections.
What can Cloud do for you?

(If you are building a security technology)
• What is the cloud?
The cloud is a term used to describe the Internet. An store your music and documents online and access them from anywhere.
The Cloud is...

• Services where data is held and computation is done server-side and presentation is done client-side

• Business models built around pricing as a function of service usage
What does Cloud AV Look like?
Conventional v. Cloud
Conventional v. Cloud
Conventional v. Cloud
• From a high level it is similar to what lives on the desktop

• Accepts crypto hashes, fuzzy hashes, machine learning feature vectors and spits out “good/bad”
• Multi-tier data storage (cache, database, flat files)

• Allows for analysis of events on a global scale, rather than system local
So why is this even possible?
Virus Count vs. Local Application Count

- Virus Count
- Local Application Count

1985: 100
1992: 1000
1998: 10000
2005: 100000
2011: 1000000

Monday, August 22, 2011
• System cache may be blown out, but globally there is a high level of cache locality

• Bandwidth of round-trip lookups is dramatically lower than that of shipping virus updates

• Low-latency bandwidth is practically ubiquitous
What does this give you?

- Intelligence
- Accuracy
- Data for and ability to apply novel techniques
Intelligence

• Continuous collection of who saw what, when, and in what context

• Can request additional data on any file that is suspicious or requires further analysis

• Extracted from your community, not what is passed around by sample vendors
Accuracy

• Closes the gap between when a signature is first published and when it is available to the client

• Optimize around real metrics (not guesses) about in-field efficacy based upon lookups from end users

• Crowdsourced whitelisting and blacklisting (more on that in a bit)
Novel Techniques

- Global prevalence tracking
- Real data for machine learning
- Retrospective conviction
- APT hunting
Algorithm Design

or, just because it isn’t $O(n^x)$, doesn’t mean it’s fast.
Bad Algorithms

- $O(x^n)$, where $x, n$ are any of the following:
  - User count
  - Rule count
  - Anything that may grow as the system gets older
Good Algorithms

• Anything $O(1)$
  • Use hash tables extensively

• If $O(x^n)$
  • $x, n$ should be constants, such as the number of features examined in an executable

• Or, do it offline / out of band
Everything is a queue
And there are bad queues, and good queues
Good Queues

- Shoot for G/D/n, with service rates defined by aforementioned O(1) algorithms

- Thank you, Harish Sethu @ Drexel University, for making me take Queueing Theory
Take only what you need

You can’t store everything online
Current, stable, SoTA

- Multithreaded server
- Memcached layer
- MySQL/MSSQL/Oracle below
- Log files
Current, non-stable, SoTA

- Asynchronous server
- Memcached layer
- NoSQL: Redis / MongoDB / Riak / Membase / Cassandra, pick your poison
- Log files
CPU Analogy

- Be VERY choosy about what data sits in L1, L2, L3, and disk, otherwise see Chernobyl slide
In Conclusion...
Stop griping, start building.
Cloud AV isn’t just AV

It’s a combination of...
• Traditional catch-and-block
• Real-time analytics
• Retrospective repair
• Deep forensics
But why just reinvent one acronym?

- HIDS/HIPS
- DLP
- 2FA (Duo Security)
Questions?
Contact Info

Adam J. O’Donnell, Ph.D.
Chief Architect, Cloud Technology Group
Sourcefire, Inc.
aodonnell@sourcefire.com