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.

SOURCEfire®

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

SOURCEfire®

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.

SOURCEfire®

THE WORLD'S ONLY RELIABLE NEWSPAPER

COMPUTER VIRUS SPREADS TO HUMANS!

BAR GLASSES HELP YOU SEE STRAIGHT WHEN YOU'RE DRUNK!

Monday, August 22, 2011

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

SOURCEfire

Virus vs. Anti-Virus, 1990s Style

- Anti-Virus:
  - Definitions updated daily to weekly
  - Mutation rate: Busines hours response teams
  - Propagation: Sneakernet, BBS, Internet

SOURCEfire®

Virus vs. Anti-Virus, Today

- Viruses:
  - 2000: $5*10^4$
    2003: 10^5
    2008: 10^6
    Today: 10^7
  - Average in field lifetime: 2 to 3 *days*.

SOURCEfire®

Virus vs. Anti-Virus, Today

- Anti-Virus:
  - Definitions updated every 5 minutes
  - Mutation rate: Follow the sun response teams
  - Propagation: Internet-only

SOURCEfire®

How do AV firms know what viruses exist?

SOURCEfire®

GOOD OL' BOYS

Cooter's
Nashville, TN

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

SOURCEfire®

SOURCEfire®

- 1980's: Informal sample sharing alliances.

SOURCEfire®

- 1980's: Informal sample sharing alliances.
- 1990's: Informal sample sharing alliances.

SOURCEfire®

- 1980's: Informal sample sharing alliances.
- 1990's: Informal sample sharing alliances.
- 2000's: Informal sample sharing alliances.

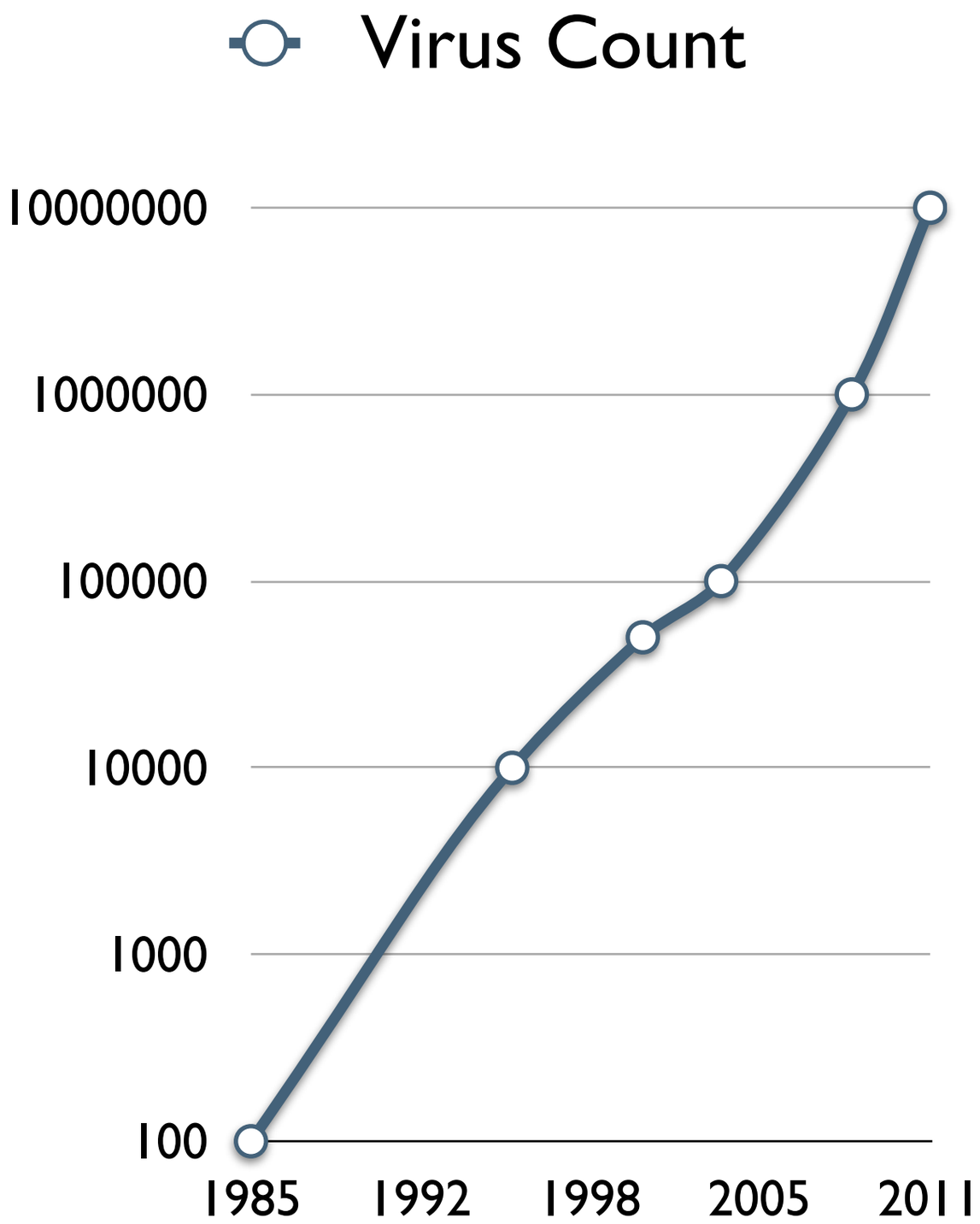
SOURCEfire®

- 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.

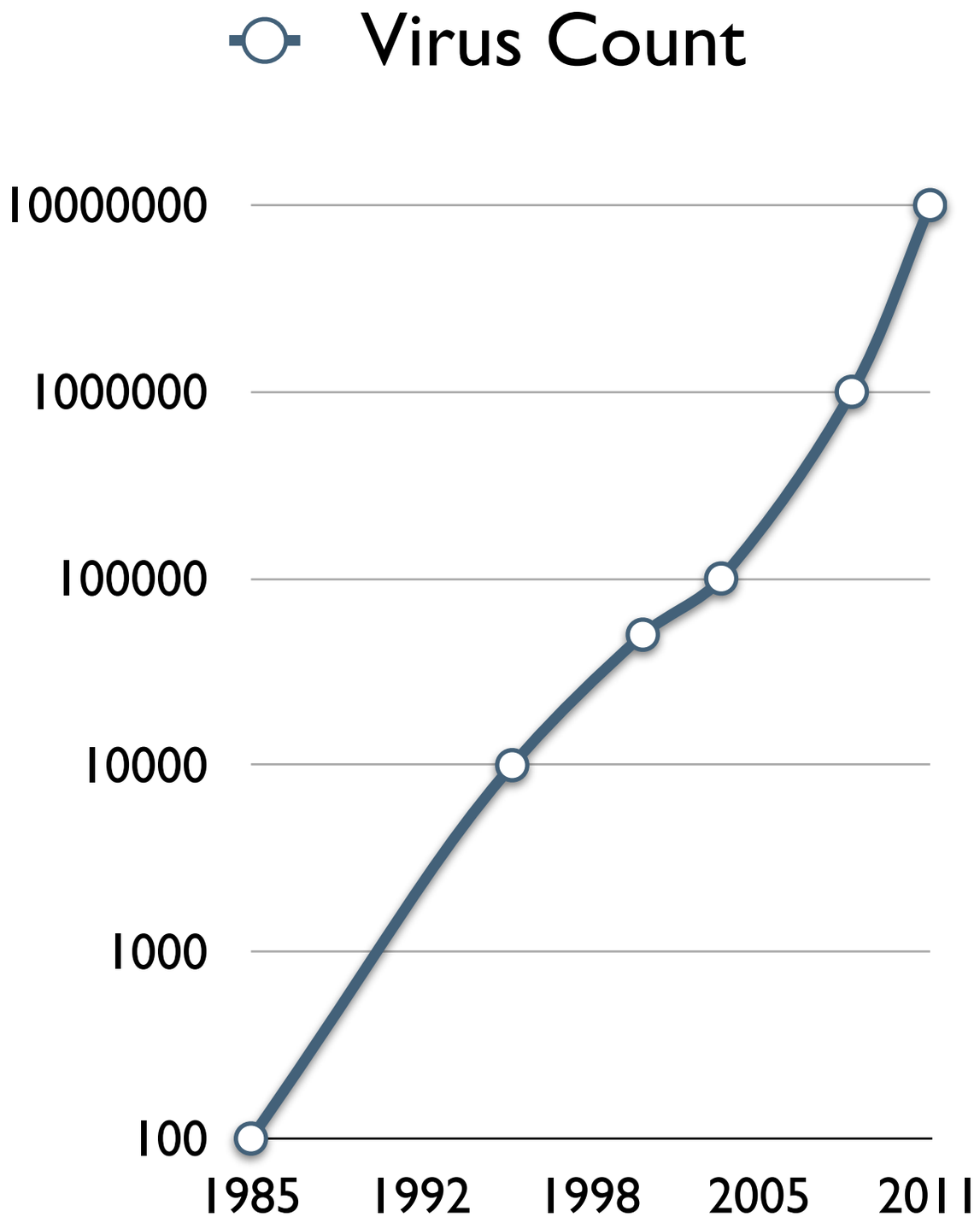
SOURCEfire®

- Virus Count

SOURCEfire®


Virus Count
10000000
1000000
100000
10000
1000
100
1985
1992
1998
2005
2011

SOURCEfire®



Intel

SOURCEfire®

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

SOURCEfire®

What can Cloud do for you?

(If you are building a security technology)

do?

- **What is the cloud?**

The cloud is a term used to describe the Internet. An
your hard drive in the cloud. Securely store your mus
and documents online and access them from anywhe

Source: Amazon's Cloud Player FAQ

SOURCEfire®

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

SOURCEfire®

What does Cloud AV Look like?

SOURCEfire®

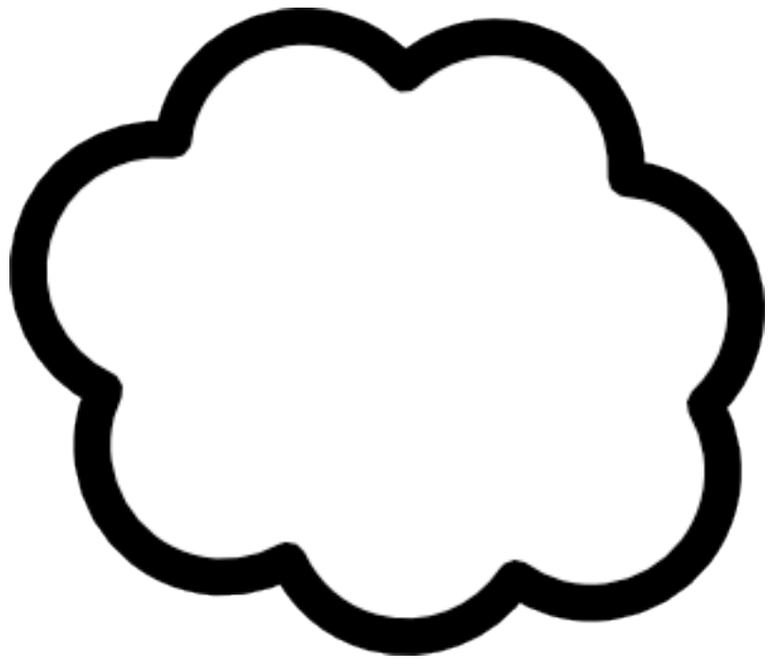
Conventional v. Cloud

SOURCEfire®

Conventional v. Cloud

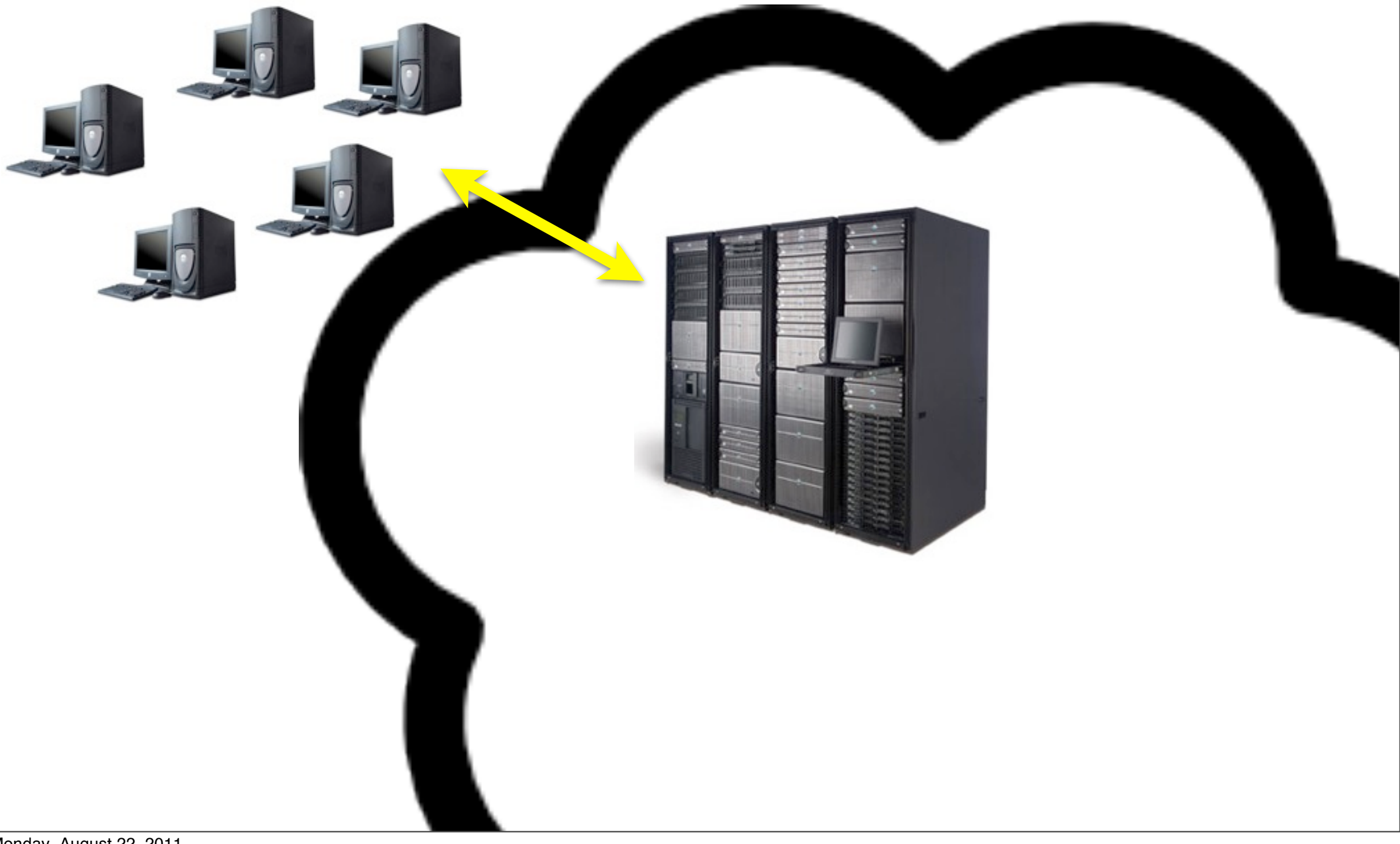
SOURCEfire®

Conventional v. Cloud



SOURCEfire®

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

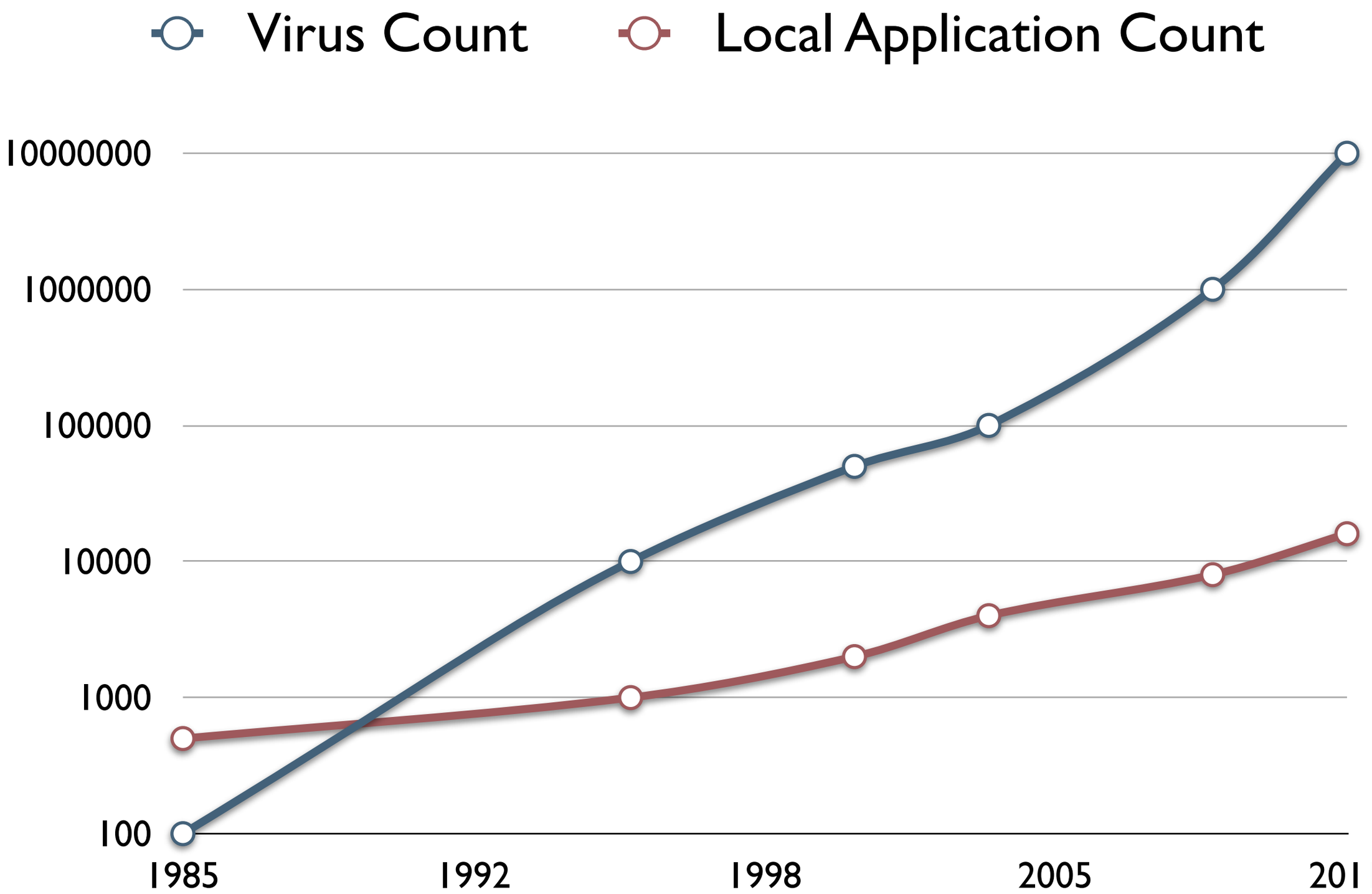
So why is this even possible?

SOURCEfire®

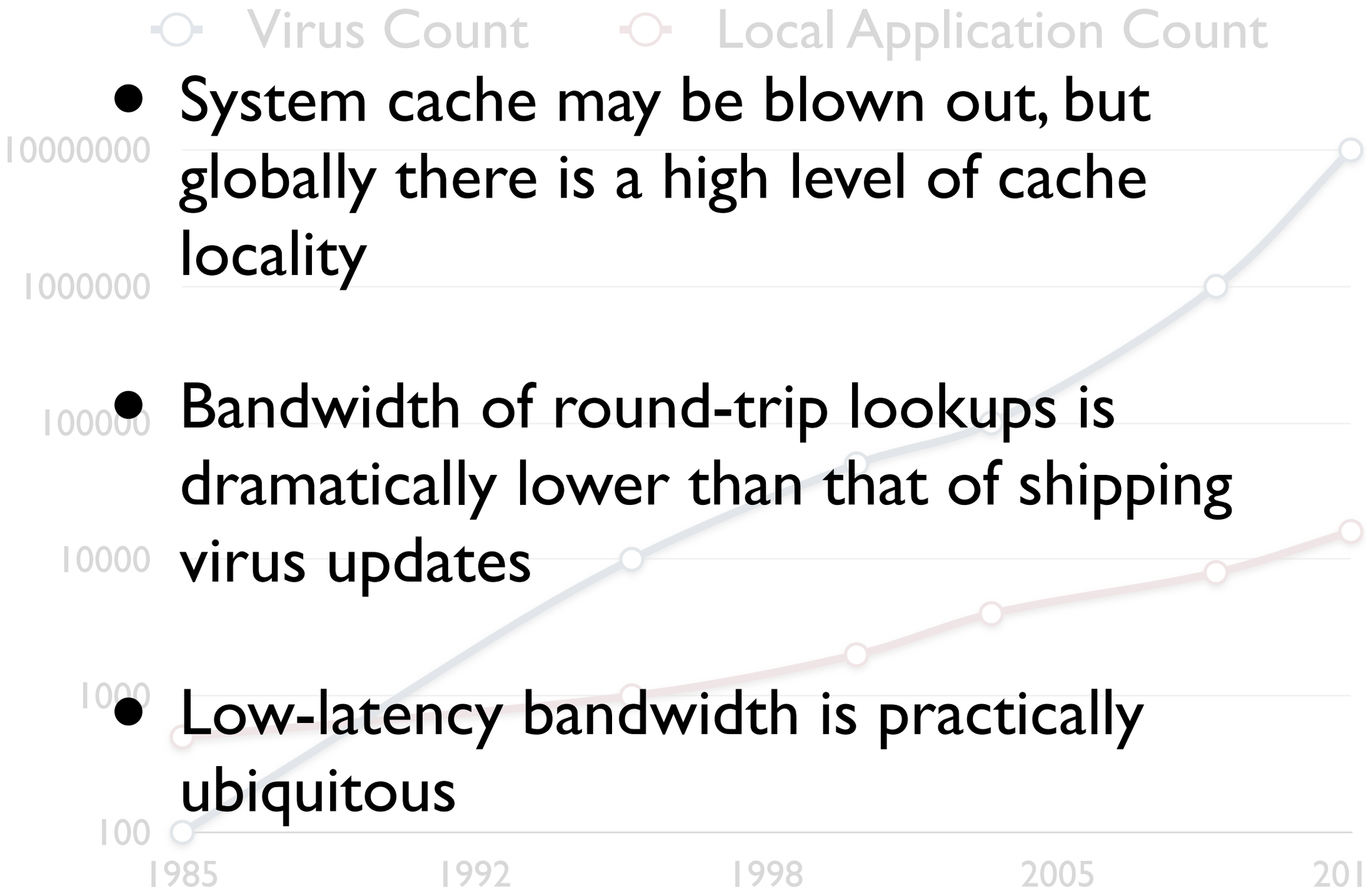
Virus Count

Local Application Count

SOURCEfire®


Virus Count
Local Application Count
10000000
1000000
100000
10000
1000
100
1985
1992
1998
2005
2011

SOURCEfire®



- 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

SOURCEfire®

What does this give you?

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

SOURCEfire®

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

SOURCEfire®

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)

SOURCEfire®

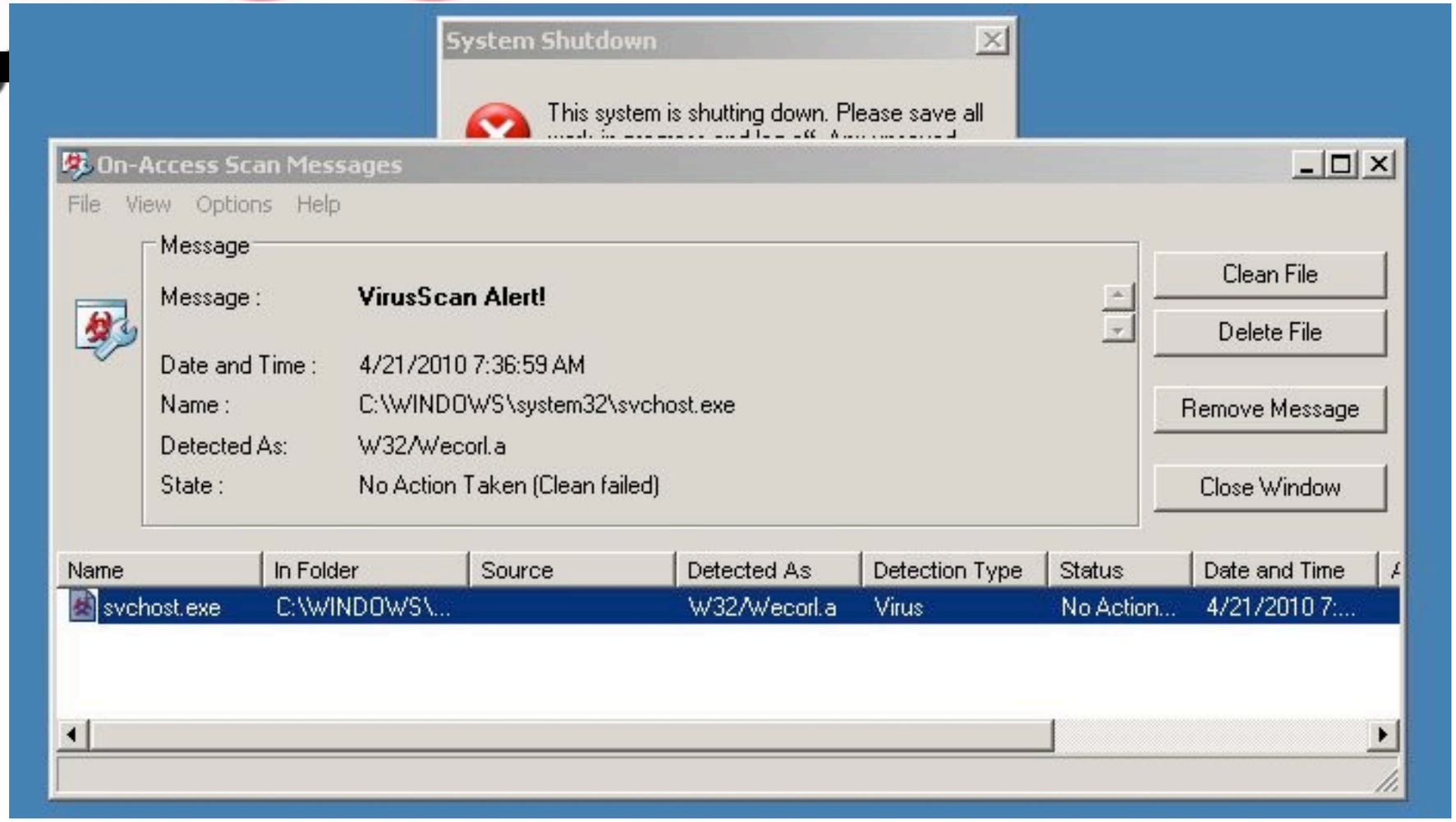
Novel Techniques

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

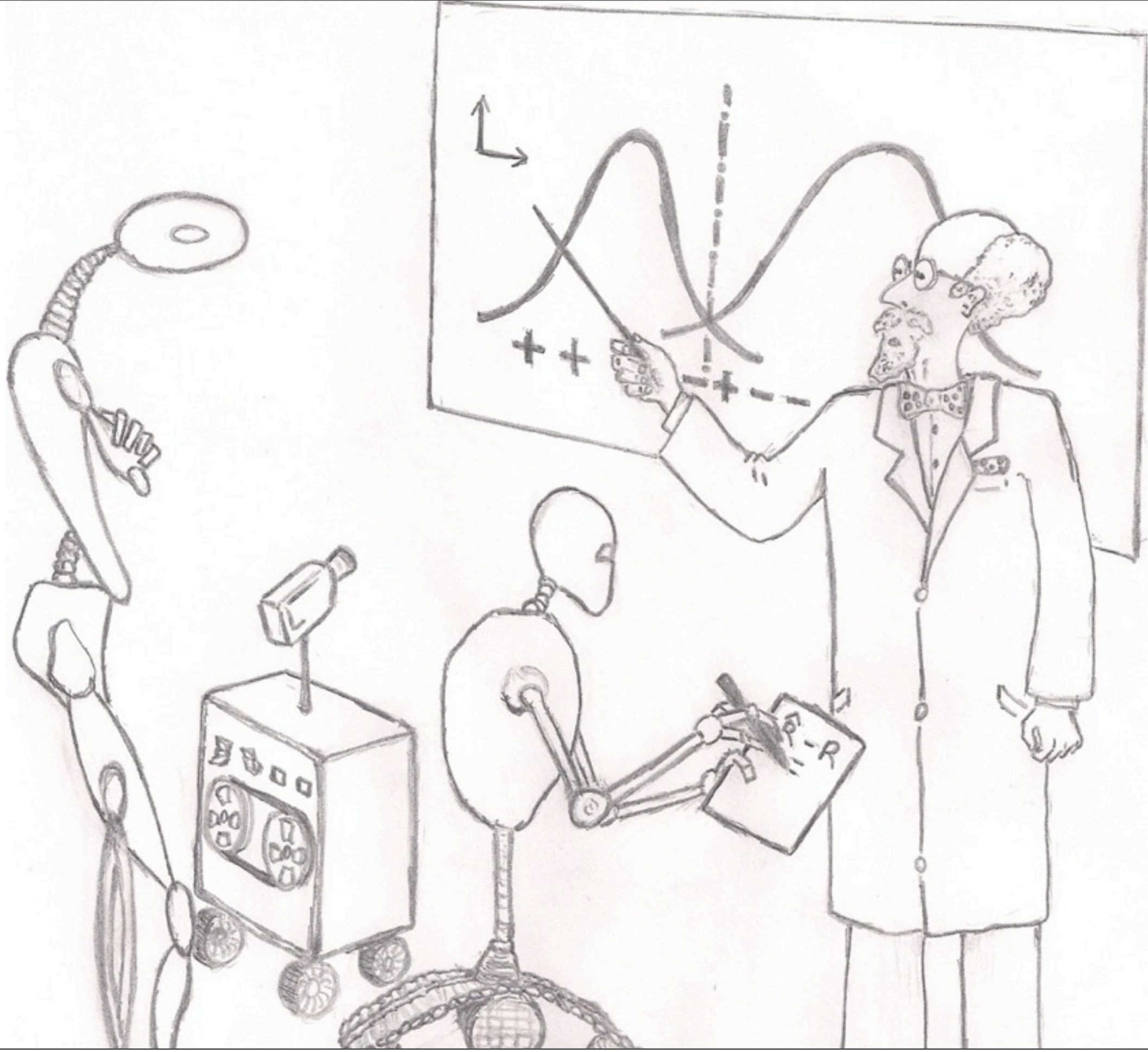
SOURCEfire®

SOURCEfire®

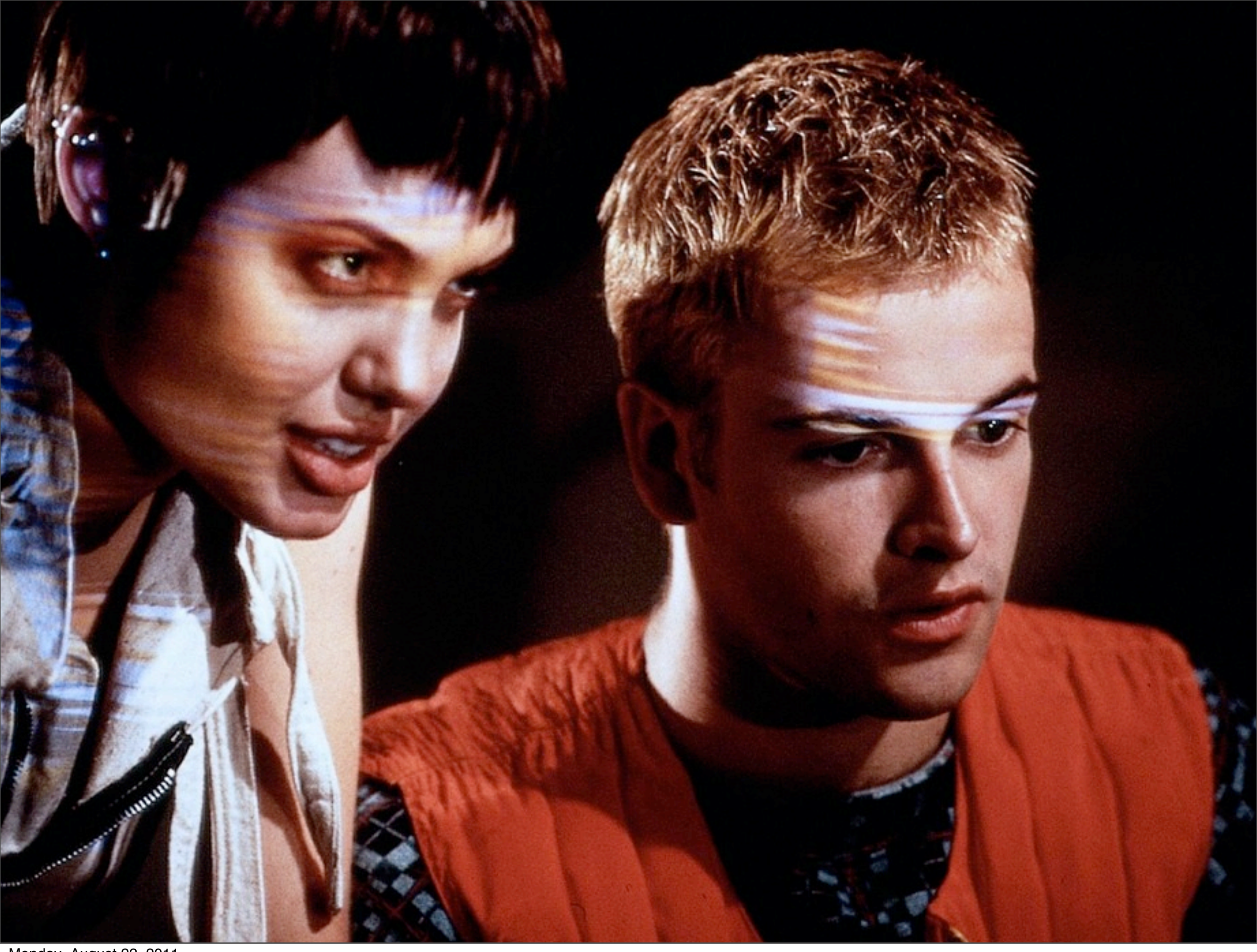
[redacted]



SOURCEfire®


+ +
- + - -
R - R


RETROSPECTIVE
THE BEST OF SUZANNE VEGA



Algorithm Design

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

SOURCEfire

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

SOURCEfire®

Monday, August 22, 2011

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

SOURCEfire®

Everything is a queue

And there are bad queues, and good queues

SOURCEfire®

Monday, August 22, 2011

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

SOURCEfire

Take only what you need

You can’t store everything online

SOURCEfire®

Current, stable, SoTA

- Multithreaded server
- Memcached layer
- MySQL/MSSQL/Oracle below
- Log files

SOURCEfire®

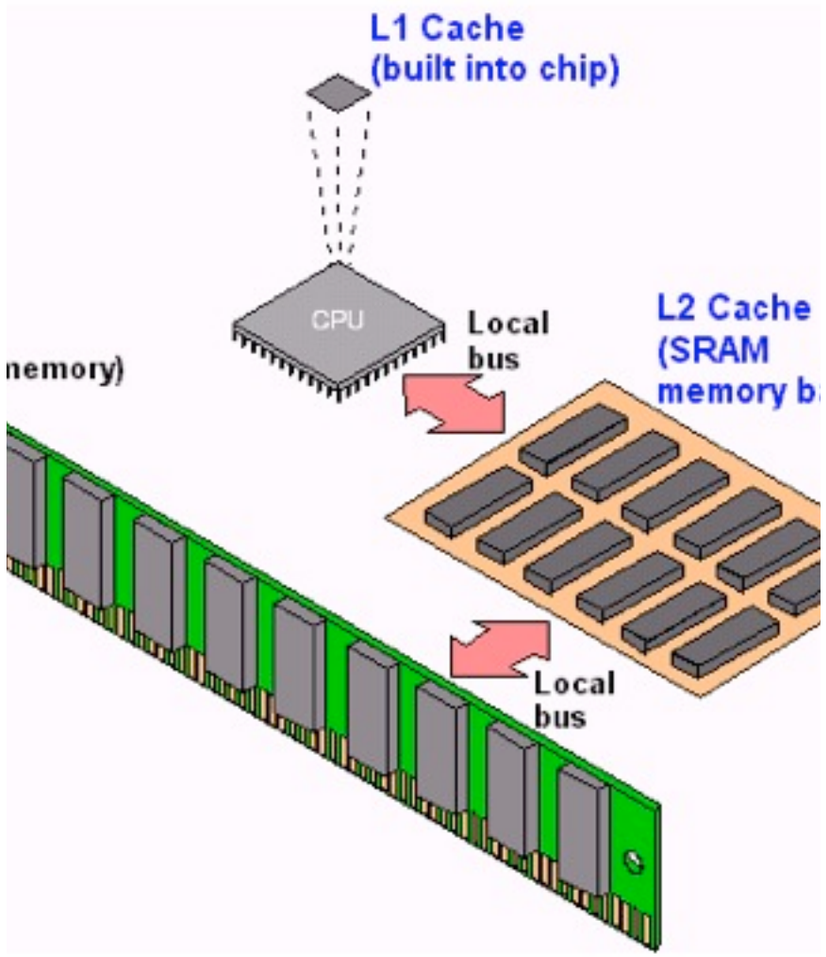
Current, non-stable, SoTA

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

SOURCEfire®

CPU Analogy

- Be VERY choosy about what data sits in L1, L2, L3, and disk, otherwise see Chernobyl slide



SOURCEfire

In Conclusion...

SOURCEfire®

Stop griping, start building.

SOURCEfire®

Cloud AV isn't just AV

It's a combination of...

Immunet

- Traditional catch-and-block
- Real-time analytics
- Retrospective repair
- Deep forensics

SOURCEfire®

But why just reinvent one acronym?

- HIDS/HIPS
- DLP
- 2FA (Duo Security)

SOURCEfire®

Questions?

SOURCEfire®

Contact Info

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