

# the evolution of collective intelligence

*adventures in information sharing*

wesyoung.me

# zombie hunting, the survival guide

- What is the REN-ISAC?
- Who do we work with?
- Why should I care?
- Challenges
- the Security Event System
- Collective intelligence
- BFG's...

# the REN-ISAC

- The REN-ISAC mission is to aid and promote cyber security operational protection and response within the higher education and research (R&E) communities.
- The mission is conducted within the context of a private community of trusted representatives at member institutions, and in service to the R&E community at-large.
- REN-ISAC serves as the R&E trusted partner for served networks, the formal ISAC community, and in other commercial, governmental, and private security information sharing relationships.

# What do we do?

- we're a CSIRT for .edu
- we send 10-12k notifications to all of north american .edu per month
- we provide community resources that allow our membership to communicate threat / experience data in a "safe space"
- create trusted interfaces between our membership and the rest of the world (leo, private industry, public resources, etc)
- we also build tools, participate in standards discussions.
- we hunt zombies.
- we go to conferences and meet-up's, and make relationships.
- we drink beer.

# Who do we (operationally) work with?

- Institutions of higher learning
- Law enforcement
- Industry groups
  - ISP's
  - Researchers
  - Policy Groups
- foreign AND domestic

# within our membership

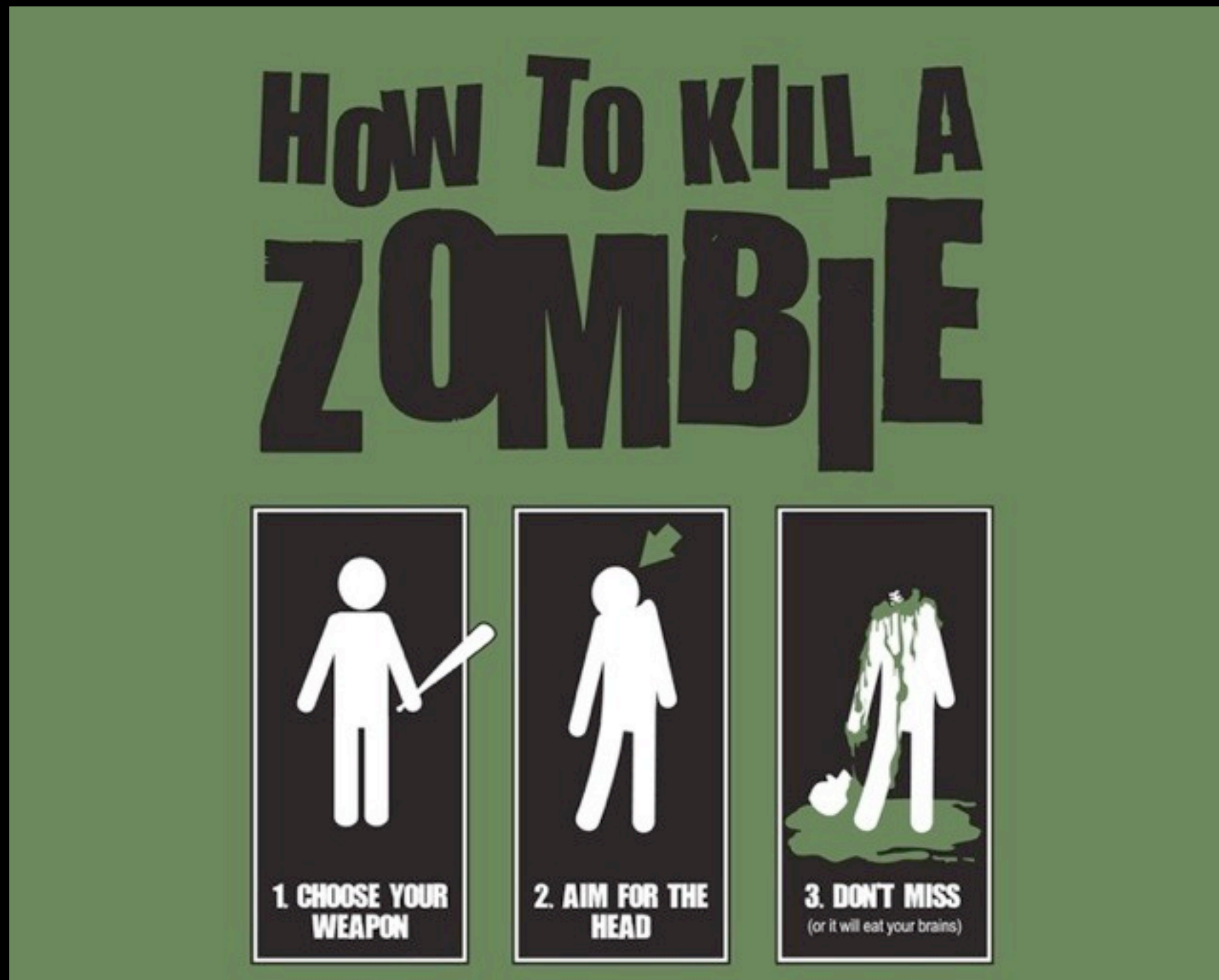
- 325+ Institutions (500+ 'distinct' campuses, state-systems, etc)
- 825+ individual members (role is firefighting with enterprise responsibility)
- They all pay \$700-900 a year
- Mostly North America (few scattered throughout other english speaking countries)
- lots of ipv4 allocations
- lots and lots of ipv6 allocations (in production for years)
- big bandwidth
  - typically a few hundred meg to multi-gig pipes
  - internet2 backbone -- 40-100 gig
- lots of different cultures, perceptions, ideals
- lots of diverse students (laptops coming and going from .kr, .cn, .us, .eu, .etc)
- firewalls... ha. yea right. Not enough beer for that rat-hole.
- Everyone and every institution is their own unique snowflake

# Challenges

- Competition for resources.
- Competitive advantage (sometimes we're competing ON security itself, in places where we shouldn't be).
- There's a delicate balance between free-market and global competitiveness, the question of where "security" falls in that equation hasn't readily been solved.
- Security doesn't respect national boundaries (neither do the bad guys, in-fact they use that to their advantage).
- Policy is hard (trying to get 325+ institutions to all sign the same agreement is a multi-year process, but it's legally possible and I can prove it).
- Tools cost money.
- Information exchange costs beer (there's a direct correlation between beers had and botnets taken down).
- Too many security geeks forget that we're on the same side here.. (good vs evil).
- Languages, formats (XML vs JSON... IODEF vs MAEC), ideals, political beliefs, self-interest, etc..
- Cultural differences, ethics, etc...
- It's really hard to measure if what we're doing is worth it.. (although, my paycheck right now depends on it being successful)

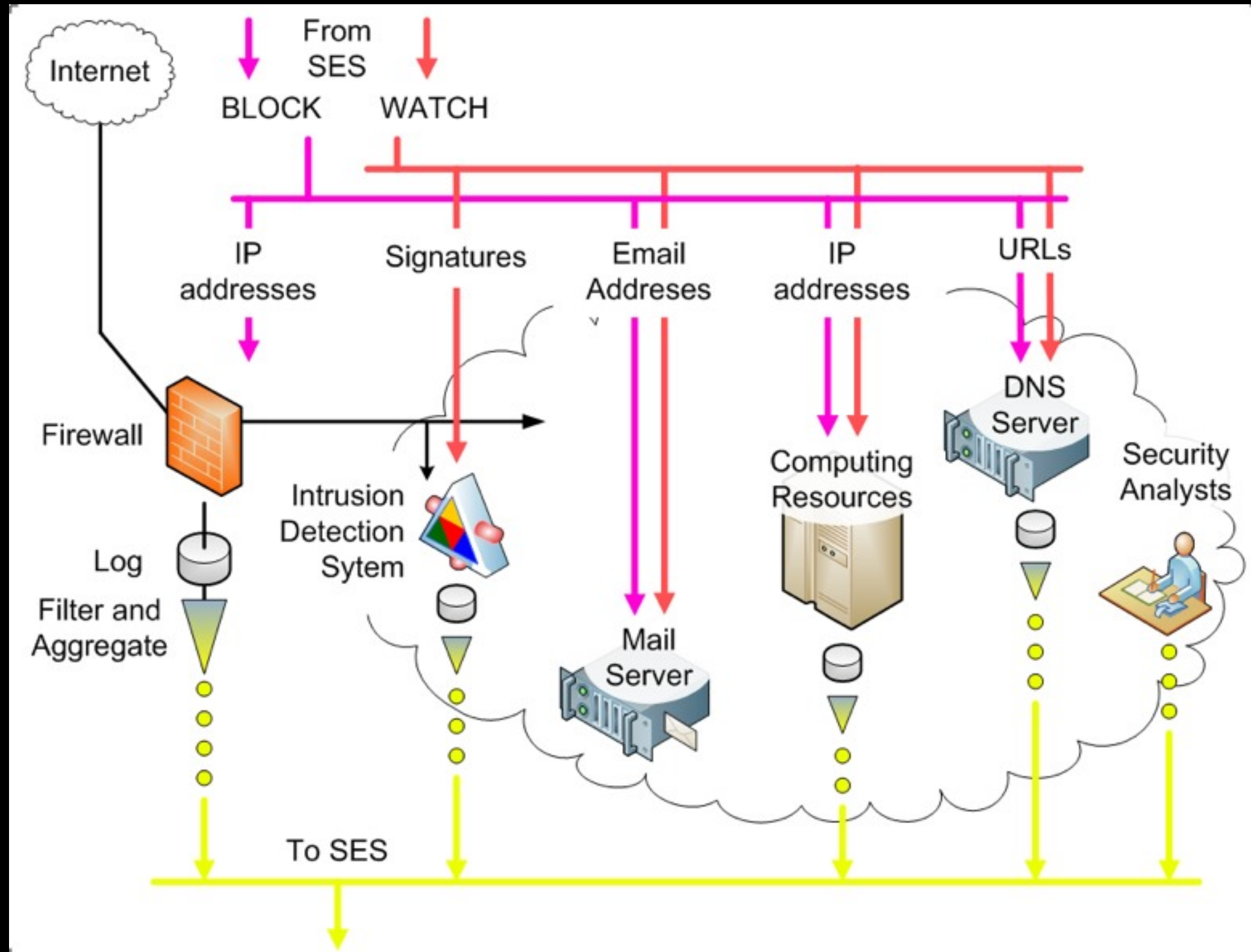
# How we share data...

(to kill zombies)





# the Security Event System



# SES.

- Site A makes a botnet controller submission (1.1.1.1/tcp/8080) via a web-interface.
- Site B pulls down the feed 5min later
- Site B throws feed in IDS
- Site B find zombies.
- Site B nukes zombies
- Site A just created more work for Site B by means of automation. It's as simple as that.

# SES v I (2008)

- SESvI uses open source components including Best Practical's Request Tracker for Incident Response (RTIR) for basic human interface and correlated event repository, Prelude Technologies Prelude Manager for raw event repository and correlation and libprelude API for automated client submission.
- Generates Intelligence Feeds (Block lists, watch lists, etc)
- Provide simple, automated correlation (this site scanned 10 Universities)
- Lower the barriers to entry when it comes to data-sharing
- We got something working in 18-months for \$120k (ish), no tools, just developing the process and glue-code.

# SES v1: Data Types

- IP address, representing just about any type of compromised host or source of threat, e.g. a botnet command and control (C&C) host or drone, a distributed denial-of-service (DDoS) attack source, a host scanning the Internet for vulnerable machines, etc.
- Classless Inter-Domain Routing(CIDR) block, representing a miscreant-heavy address range (e.g. Russian Business Network), and as descriptive information for IPv4 address-based records
- Autonomous System Number (ASN), as additional descriptive information
- Fully Qualified Domain Name (FQDN), representing for example, a botnet C&C, suspicious name server, other botnet infrastructure, or a consistently malicious domain
- URL representing for example, a malware download or phishing sites
- E-Mail address, for example, a phishing Reply-To address
- Malware descriptions, for example malware

# SES v1: Rollout (2009)

- “SES” has 10+ sites Sharing automated, machine generated data between 50 and 20,000 data-points per day per site.
- (SSH|Telnet|FTP|VNC|Pushdo|Darknet) Scanners.
- Near realtime (\* / 15... \* / 30min) in most cases, from live sensors as well as honeypots
- Leveraging Snort, Nepenthes, syslogs, Custom Darknet scripts via the current SES API (libprelude)
- We create a “correlated scanners” (multi-location) into a mitigation feed for sites to pull down.
- We also have a web page users can manually enter malicious domain-names, malware drop sites, botnet C&C into which produce various other mitigation feeds (stuff they’ve manually investigated).

# SES v1: Lessons Learned

- <http://bret.appspot.com/entry/how-friendfeed-uses-mysql>
- Database design, small, concise
- Database design to support “schema-less” data
- Standards-based, but don't tie to a single standard – make design decisions that accommodate multiple data representation standards in a single database
- Learn from other's successes and mistakes
- Community engagement for determining design priorities
- Feedback from a team of knowledgeable early adopters
- pilot pilot pilot with your community! they'll be the ones using it!

# SES v2: Collective Intelligence

- Locally correlated Events (typically malicious ip-infrastructure)
- Spamhaus DROP list (hijacked networks)
- Malwaredomains.com feed (malware hashes, malware domains, malware ip-infrastructure)
- Malwaredomainlist.com feed (malware urls, malware domains)
- DShield List(s) (scanning ip-infrastructure)
- Phishtank Data (phishing urls, phishing ip-infrastructure)
- Zeustracker data (binary urls, config urls, domains, ip-infrastructure)
- From each domain, you have massive potential intelligence from the name-servers involved with each domain.
- Whitelists (alexa top 10, 100, 1000, 10000, mirc servers.ini, etc)
- Locally discovered intel (potentially all of the above)
- 18-24 months, \$350k (ish)

# SES v2: Lessons Learned (2012)

- If you give people data, they will try to consume ALL OF IT! and quite possibly try to throw it into their firewalls...
- No one will read the doc until they block [www.netflix.com](http://www.netflix.com), even then... they will not read your doc (and they shouldn't need to, you're tools shouldn't suck that bad), even when it's tagged at the 40% confidence level
- If you don't iterate quickly, you're setting yourself up for failure (release early, release often, get feedback). Oh man is it painful, but it's the difference between getting something working and wasting years of your life...
- Organic growth is good, if you push new users who aren't ready to absorb the topic, you'll be left answering lots of questions
- Your "bleeding edge" users are your best friends, they'll help you flush out what's important and what needs to be documented.
- Your strongest metric should be how well your tool(s) / processes are adopted with little or no marketing, if people aren't using it, your tool sucks.
- If you wanna share data with people outside your local federation, start with a short term, two page MOU and the basics.
- Don't over-think it, usually your speculation and assumptions are WRONG! (usually)



# SESV2 Today

- SESv2 (aka CIF-v0.01) has been fully rolled out in the RI community
- Our community is pulling feeds into their real-time security infrastructure as well as integrating their IR applications into our REST API.
- Institutions choosing to preemptively block based on our feeds are identifying up to a 10x reduction in their incident count.
- We're leveraging our legal framework to data-share with partners in both the public and private space
- We're able to hand our partners an API key (post agreement) and give them access to data our community has tagged as "shareable"
- partners can simply install the client (as our constituents do) and pull / output the data in whatever format or directly into their application without the need to re-parse, etc..

and more  
importantly.....

- an entire [international] community has sprung up around this framework..

# Hurdles to Sharing Data

## (as I see them today)

- High barrier to entry
  - too much plain text
  - too much structure (the larger the protocol, the less people will use it)
  - not enough code
- Little or no legal [open] frameworks
  - To do this right, lawyers must sign off on it
  - very few lawyers understand the value
  - very few technical people understand how to communicate the value
- Cost
  - high barrier equals high cost
  - no legal framework means I need to talk to a lawyer; which takes time.. resources; etc.
- No Marketing
  - everyone has to know someone, or figure it out themselves
- Politics (reads “ignorance”)
- Thinking is hard. Makes @SecurityHulk’s head hurt.
- [it’s no wonder that] most data-sharing happens “underground”

# BFG: v3 (~2014)



- a fork of PreludeIDS (as defunct it seems), hbase, protocol buffers/thrift and ZeroMQ mixed with some FM and \$800k (less overhead)
- Throw your (obscure) data at it and “FM” (a technical term for ‘magic’) happens.
- The human intelligence SEM (my SEM can read your e-mail, twitter feed and blog)
- Inter-federation
  - we already work with lots of people, how do we transform that into something sustainable (legal frameworks, sharing agreements, etc)
- Teaching others how to hunt zombies more effectively.
- [http://www.ren-isac.net/ses/ses\\_news.html](http://www.ren-isac.net/ses/ses_news.html)

# BFG:

# Contributions thus far

- developed [google] protocol-buffers for IODEF (developing for thrift too)
  - tested and released perl bindings
  - integrated into CIF v1
  - ...and MAEC
  - ...and ICSG
  - ...and IDMEF
  - ...and whatever other format people want that lowers the barrier to sharing data *fast!*
- developing IODEF “v2” based on our lessons learned
- released [simple] legal framework for sharing data (U.S. based)
  - our legal counsel gave us signatory authority over the document
  - other’s counsel have ‘approved’ it
  - we put it out on github and licensed appropriately
  - IT’S LESS THAN SIX PAGES :)
- made every single line of code easily fork-able (github style!)

free.  
(as in beer)



# Project References

- [www.ren-isac.net/ses](http://www.ren-isac.net/ses)
- [collectiveintel.net](http://collectiveintel.net)



- The REN-ISAC is a federation of diverse research and education institutions concerned with operational computer and network security. What slowly started out with some people, some hacked up mailing lists, a wiki and some magic perl glue to share intelligence, quickly snowballed into a vast sea of data that no one could keep track of or use in their day to day operations.
- Over the last few years we've invested most of our development time and effort into building tools that lower the barrier to entry for our community to share data intelligently. These tools have not only been developed with our own CSIRT constituencies in mind, but also based on feedback from the international CSIRT community.
- This talk will focus on how our community went from a set of extremely raw tools to an automated end-to-end process of sharing data within a large heterogeneous community.
- + First we'll detail how institutions currently share data directly into each other's IR process with little or no human interaction.
- + We'll also discuss how we've enhanced various international standards that enable our constituency to further share data with law enforcement agencies as well as our trusted mitigation partners.
- + Additionally, this talk will review the most common data-sharing hurdles when partnering with external organizations, and why most global data-sharing ventures have failed to scale in this space. This will include things like data parsers, information sharing agreements and data formats.
- + And finally, we'll talk about how we plan to evolve this application into the big-data environment (hundreds of billions of things per day) over the next three years.
- Attendees should walk away with a real life set of tools and lessons learned, both technical and strategic, that they can use to scale internal intelligence operations past their own borders.