this is the only slide that matters...

i'm wes.. i'm here to help you build tools
and
share data.

collectiveintel.net

intro to [effective] data-sharing, 101.

information sharing is really all about

- messaging
- storage
- analytics
- communication
- economics
- warfare
- scale
- trust
- people
- culture

goals for today

- i hate it when...
- the state of the union
- vision

imagine a “box”

- that worked like google personalized search
- that not just read, but understood your email and favorite blogs
- that communicated seamlessly with your network infrastructure (firewalls, IDS, name-servers, etc) in real-time
- that could be peered with your close, trusted, partners “box” to exchange information

now imagine that "box"

- leveraged existing science intelligence analytics (bio, chem, etc) to analyze the data
- mitigated things before you've even had your first cup of coffee...
- could handle trillions of 'events' per day (netflow, passive dns, log flow, etc)

and... was an ‘open’ framework

imagine watson on
netflow...

whois?

survival guide

- what is the REN-ISAC?
- who do we work with?
- challenges
- the Security Event System (SES)
- Collective Intelligence Framework (CIF)
- BFG's...
- the next ten years

Please... INTERRUPT ME!

we have 120min together... and the doors are locked, for your safety.... :)

whois ?

- 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 the CSIRT for north american .edu
- we send 10-12k notifications per month to .edu
- 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 [lots of] beer.

within our membership

- 325+ Institutions (500+ ‘distinct’ campuses, state-systems, etc)
- 825+ individual members (role is firefighting with enterprise responsibility)
- 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 -- ~200 universities, 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.
- Everyone and every institution is their own unique snowflake

who do we work with?

- Institutions of higher learning
- Law enforcement
- Industry groups
  - ISP's
  - Researchers
  - Policy Groups
  - the APWG (standards development, road-show's, bar nights, etc)
- domestic and foreign

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.

challenges

- 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)

challenges

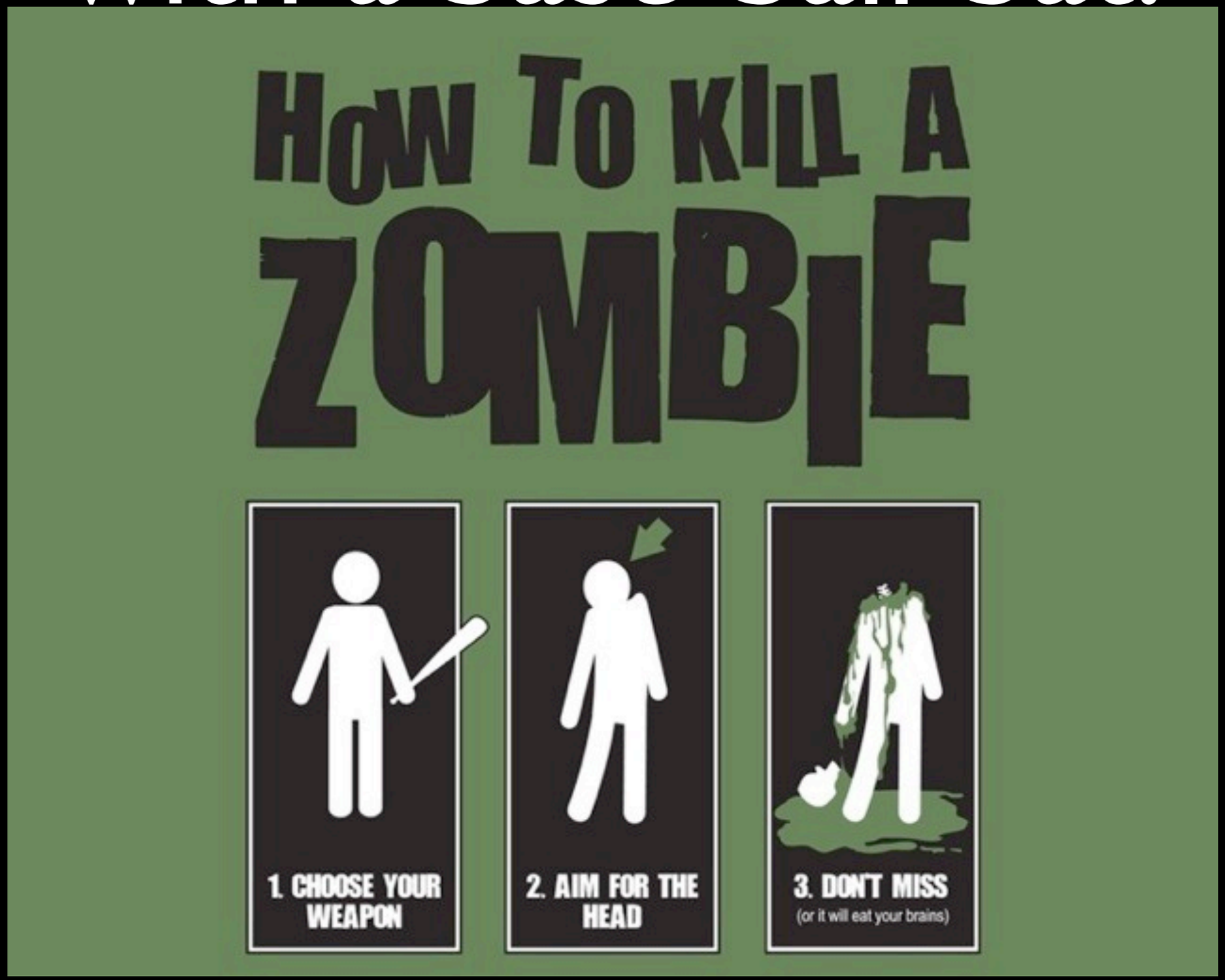
- High barrier to entry
  - too much plain text
  - too much structure (the larger the protocol, the less people will use it)
  - not enough code
- Little or no legal [open] frameworks
  - To do this right, lawyers must sign off on it
  - very few lawyers understand the value
  - very few technical people understand how to communicate the value

challenges

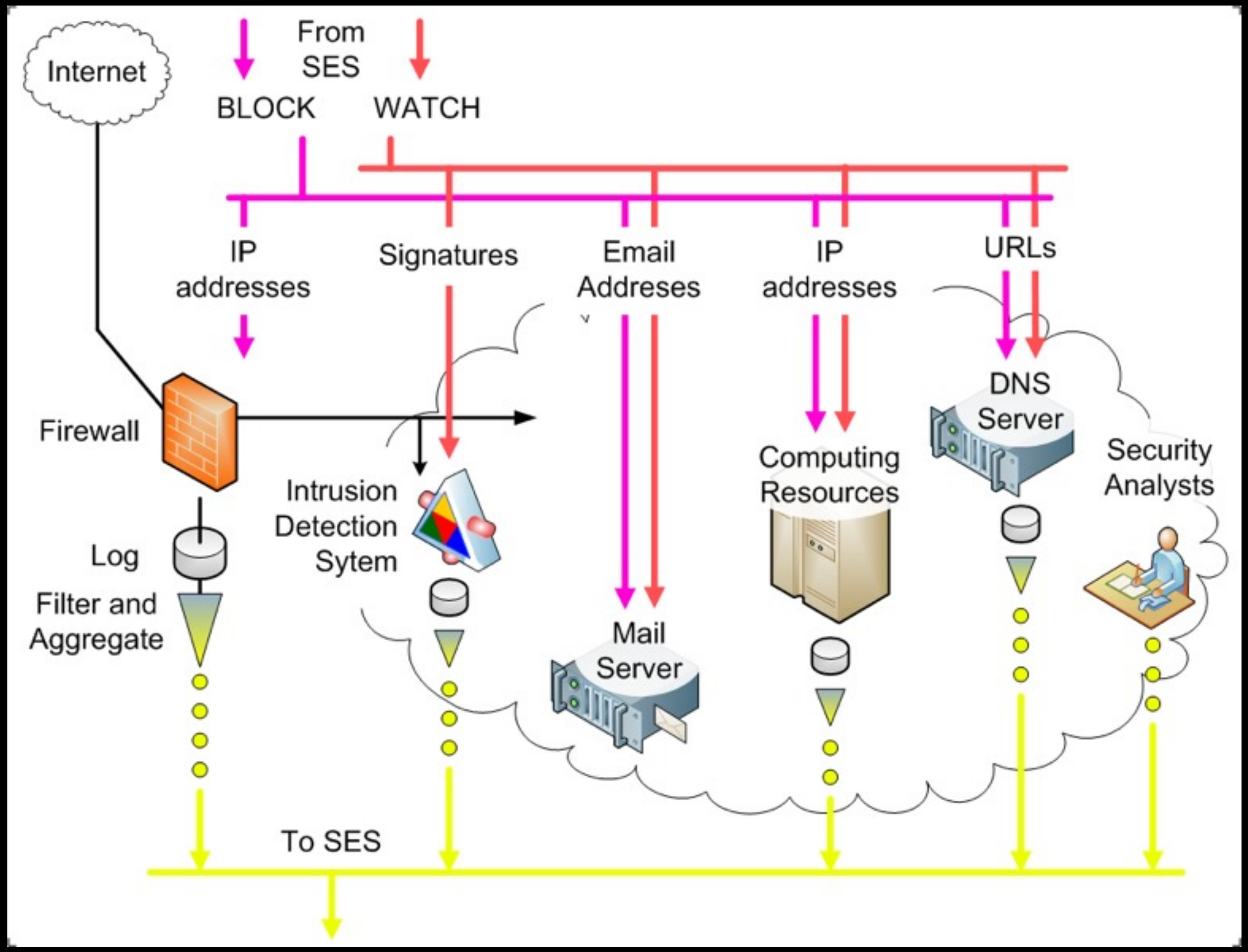
- Cost
  - high barrier equals high cost
  - no legal framework means I need to talk to a layer; which takes time.. resources; etc.
- No Marketing
  - everyone has to know someone, or figure it out themselves
- Politics (reads “ignorance”)
- Thinking is hard. Makes @SecurityHulk’s head hurt.
- [it’s no wonder that] most data-sharing happens “underground”

how do you go about sharing information?

we aim for the head, with a base-ball-bat.



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 (a simple denial of service?) for Site B by means of automation.

what we tried...

- webpage with wget scripting of plain text to firewalls, ids's and dns servers (2008)
- RT+IR+Prelude with wget scripting of plain text to firewalls, ids's and dns servers (2009)
- people had to do their own conversion to the $DEVICE rules

what else we tried...

- RT+IR+Prelude+CIF+XML with a “special client-side library” to $DEVICE (2011)
- RT+IR+Prelude+CIF+JSON with a “special client-side library” to $DEVICE (2012)
- RT+IR+Prelude+CIF+Protocol buffers with a “special client-side library” to $DEVICE (2013)

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!

Collective Intelligence

(SESv2, CIFv0)

- 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, 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 (google customer service approach is usually best)

SES v2: Lessons Learned (2012)

- 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-engineer something, usually your speculation and assumptions are WRONG.

SESv2 Today

- operational in the RI community for over a year
- 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 (and soon internationally)
- 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..

lessons learned

- plain text was too hard to parse and build into applications (no API)
- XML was too fat to parse quickly (scale--)
- JSON was too loose to to enforce any kind of “morals” on the data, serialization still took it’s toll...

protocol buffers

- Protocol buffers are Google's language-neutral, platform-neutral, extensible mechanism for serializing structured data – think XML, but smaller, faster, and simpler.
- developers.google.com/protocol-buffers

protocol buffers

- enforce the format
- portable across programming languages
- 10x-100x smaller than XML/JSON
- faster encode/decode (with scale, every second counts)
- easy integration with anything that "carries a message" (zmq, http, smtp, etc...)

SESv2.1 (CIF v1)

- entering the final stage of release candidates
- community tested (testing? ... providing feedback, keeping my inbox full)

New Tab | CIF Client Extension Options | CIF Query

Collective Intelligence Framework | Query Server | Submit Data | Settings | [Top of Page] v1.2

Run a New Query

Server: ses-qa

Query: google.com

Submit Log Query: [+]

QUERY RESULTS

google.com

Results for **google.com**

Server Name: ses-qa
Feed Restriction: RESTRICTED
Time: 2013-02-04T18:01:04Z
Export: Text Table CSV

restriction	address	protocol/ports	detecttime	impact	severity	confidence	description	Incident Meta Data (Expand/Collapse all)	Additional Data (Expand/Collapse all)	alternativeid [restriction]
PRIVILEGED	ns1.google.com		2011-07-16T21:00:47Z	suspicious nameserver	medium	10.625	unknown_html	Related Event Show Data	Show Data	http://support.clean-mx.de/clean-mx/viruses.php?id=911451 [LIMITED]
PRIVILEGED	ns2.google.com		2011-07-16T21:00:47Z	suspicious nameserver	medium	10.625	unknown_html	Related Event Show Data	Show Data	http://support.clean-mx.de/clean-mx/viruses.php?id=911451 [LIMITED]
PRIVILEGED	ns4.google.com		2011-07-16T21:00:47Z	suspicious nameserver	medium	10.625	unknown_html	Related Event Show Data	Show Data	http://support.clean-mx.de/clean-mx/viruses.php?id=911451 [LIMITED]
PRIVILEGED	ns3.google.com		2011-07-16T21:00:47Z	suspicious nameserver	medium	10.625	unknown_html	Related Event Show Data	Show Data	http://support.clean-mx.de/clean-mx/viruses.php?id=911451 [LIMITED]
PRIVILEGED	ns2.google.com		2011-08-	suspicious	medium	10.625	unknown_html_rfi_php	Related Event	Show Data	http://support.clean-mx.de/clean-

New Tab | CIF Client Extension Options | CIF Query

Collective Intelligence Framework | Query Server | Submit Data | Settings | [Top of Page] v1.2

Results for **google.com**

Server Name: ses-qa
Feed Restriction: RESTRICTED
Time: 2013-02-04T18:01:04Z
Export: Text Table CSV

restriction	address	protocol/ports	detecttime	impact	severity	confidence	description	Incident Meta Data (Expand/Collapse all)	Additional Data (Expand/Collapse all)	alternativeid [restriction]
PRIVILEGED	ns1.google.com		2011-07-16T21:00:47Z	suspicious nameserver	medium	10.625	unknown_html	Related Event Show Data	Show Data	http://support.clean-mx.de/clean-mx/viruses.php?id=911451 [LIMITED]
PRIVILEGED	ns2.google.com		2011-07-16T21:00:47Z	suspicious nameserver	medium	10.625	unknown_html	Related Event Show Data	Show Data	http://support.clean-mx.de/clean-mx/viruses.php?id=911451 [LIMITED]
PRIVILEGED	ns4.google.com		2011-07-16T21:00:47Z	suspicious nameserver	medium	10.625	unknown_html	Related Event Show Data	Show Data	http://support.clean-mx.de/clean-mx/viruses.php?id=911451 [LIMITED]
PRIVILEGED	ns3.google.com		2011-07-16T21:00:47Z	suspicious nameserver	medium	10.625	unknown_html	Related Event Show Data	Show Data	http://support.clean-mx.de/clean-mx/viruses.php?id=911451 [LIMITED]
PRIVILEGED	ns2.google.com		2011-08-14T11:59:19Z	suspicious nameserver	medium	10.625	unknown_html_rfi_php	Related Event Show Data	Show Data	http://support.clean-mx.de/clean-mx/viruses.php?id=967566 [LIMITED]
PRIVILEGED	ns3.google.com		2011-08-14T11:59:19Z	suspicious nameserver	medium	10.625	unknown_html_rfi_php	Related Event Show Data	Show Data	http://support.clean-mx.de/clean-mx/viruses.php?id=967566 [LIMITED]
PRIVILEGED	ns4.google.com		2011-08-14T11:59:19Z	suspicious nameserver	medium	10.625	unknown_html_rfi_php	Related Event Show Data	Show Data	http://support.clean-mx.de/clean-mx/viruses.php?id=967566 [LIMITED]
PRIVILEGED	ns1.google.com		2011-08-14T11:59:19Z	suspicious nameserver	medium	10.625	unknown_html_rfi_php	Related Event Show Data	Show Data	http://support.clean-mx.de/clean-mx/viruses.php?id=967566 [LIMITED]
PRIVILEGED	ns2.google.com		2011-09-02T12:00:21Z	suspicious nameserver	medium	10.625	tr%2fagent.892928.8	Related Event Show Data	Show Data	http://support.clean-mx.de/clean-mx/viruses.php?id=988753 [LIMITED]
PRIVILEGED	ns3.google.com		2011-09-02T12:00:21Z	suspicious nameserver	medium	10.625	tr%2fagent.892928.8	Related Event Show Data	Show Data	http://support.clean-mx.de/clean-mx/viruses.php?id=988753 [LIMITED]
PRIVILEGED	ns4.google.com		2011-09-	suspicious	medium	10.625	tr%2fagent.892928.8	Related Event	Show Data	http://support.clean-mx.de/clean-

New Tab | CIF Client Extension Options | Submit Data

Collective Intelligence Framework | Query Server | Submit Data | Settings

Document the Observation

General Details

Server: ses-qa

Data: badsite.biz,google.com,1.2.3.4

multiple datapoints (max: 1000) may be entered as comma-separated values. they will all share the same classifications below

Impact: Malware / Exploit

Refer to the impact taxonomy here

Description: unknown

description (e.g. zues c&c)

Confidence: Somewhat Confident

Refer to the confidence taxonomy here

Severity: Low

Refer to the severity taxonomy here

Protocol: N/A

Portlist:

e.g. 21,22,80-89

Add Alternative ID

Sharing Information

Groups:
- [x] everyone
- [ ] general.ren-isac.net
- [ ] xsec.ren-isac.net

PARSED ITEMS

Hostname/IP: 1.2.3.4
Hostname/IP: badsite.biz
Hostname/IP: google.com

Demo's

- Query/Submit
- Installation
- Federated Sharing

BFG (~2014)

- a "fork" of PreludeIDS (the idea, not the code itself), hbase, protocol buffers/thrift and ZeroMQ mixed with some 'FM' (a technical term for 'magic'), some machine learning goo and $800k USD via the NSF.
- Throw your (obscure) data at it and 'FM' happens.
- The human intelligence SEM (my SEM can read your e-mail, twitter feed and blog and tell my network what to do in real-time)
- Inter-federation
  - we already work with lots of people, how do we transform that into something sustainable (legal frameworks, sharing agreements, etc)

BFG: Contributions thus far

- developed [google] protocol-buffers for IODEF
- tested and released perl bindings
  - integrated into CIF v1 (spring 2013)
  - ...and MAEC
  - ...and ICSG
  - ...and IDMEF
  - ...and whatever other format people want that lowers the barrier to sharing data *fast*!

BFG: Contributions thus far

- developing "IODEF v2" based on our lessons learned
- released [simple] legal framework for sharing data (U.S. based, also used internationally)
  - our legal counsel gave us signatory authority over the document
  - other's counsel have 'approved' it / executed on 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 (in true git style)

now that you know how to share data,
here's what [i think] the next ten years
look like...

the past...

- large mailing lists of people you've met at the bar and are willing [/mandated to?] share data with
- web-portals you can share data via a wiki
- web-portals you can download a pdf from
- web-portals you can download structured data from (with/with-out an actual API)

the past...

- trust is controlled by how much the group is willing to share with itself
- the larger the group, the lower the overall trust measure
- there are hard ceilings to data-sharing in this model
- these are all problems we have today

what social networks have re-taught us... again.

- they do not allow their hubs to inter-operate with other networks (and therefor other hubs...)
- AOL made IM easy, Jabber re-invented it and took it over (till everyone moved to fb)
- we saw this movie play out over the last decade+... Prodigy is no longer with us.. :(

what social networks have re-taught us... again.

- build your platform so data-hubs can grow organically within your "social graph" (or org)
- allow those hubs to be self-selective to whom they will share what types of data to (not everyone is created equal)

the next ten years

- yes, this is a ten year problem
- AOL didn't "realize" they were a "media company" till the early to mid 2000's
- it took that long for the browser market to solve this "federation" problem and gain adoption.
- it took that long for web2 to take hold

the next ten years

- teh Facebook is a social platform for connecting you with your friends
- the LinkedIn is a social platform for connecting you with your friends who have $$ and would be dumb enough to hire you
- the google plus is a social platform for security peeps who have no desire for Facebook's shenanigans (if Ferg was here, he'd +1 this)

the next ten years

- the APWG is a social platform for connecting e-crime researchers
- the US-CERT is a social platform for connecting .gov with each-other and private industry
- the REN-ISAC is a social platform for connecting edu's with other edu's

the next ten years

- what happens if there was a google+ feature that allowed you to specifically share something with a target group?
- what happens if you could dump structured +encrypted data in that sharing window (ever played with scrambls?)?
- what happens if there was an API into that platform...?

with technology like this

- why does the REN-ISAC need to exist at all?
- why do you need to be “a member” of something to share information?

where we fail

- most [international] information sharing communities are great aggregators of internally shared information
- most cross-hub action happens by those who are in many communities
- we’re actually just inhibiting the data-sharing process

you should be thinking...

- if you're not already doing automated data-sharing, why not?
- should you be focused on designing a new standard? or evolving something that already works?
- what does your architecture look like in ten years if you're standardizing around XML or JSON (or even Protocol Buffers)?

you should be thinking...

- yourself as a platform for trusted relationship building (reads: do you have a bar night at your cons?)
- how to enable your community to individually share data with the rest of the world, not just with itself
- is your business model focused on sharing data? or facilitating relationships..?

the new Science of Networks

- people are hubs
- the should be enabled as such

solve problems, don't invent them

- what tools exist to help solve this problem?
- are your partners thinking in terms of 'big data'?
- are you thinking in terms of bigger data (trillions++)?

criminals

- develop tools
- extract value/wealth from others
- which enables them to build more sophisticated tools
- which enables them to extract even more value/wealth from others

espionage

- “tax-payer” funded (usually)
- provides for the development of sophisticated tools
- which allows ‘states’ to extract value/wealth from other ‘states’ (or economic-hit-man their wealth, which is the same thing)
- which allows for more sophisticated tools

#irony

"intelligence" (a/v, ids, threat feeders, etc) companies build tools so they can extract value from customers to protect them from other people who are trying to extract value/wealth from their customers

...so they can build even more sophisticated tools that protect us from the other more sophisticated tools that someone else (states, criminals, etc) are now building.

how big is your business?

can you compete with that and expect to remain successful?

as a nation, can we compete on security and expect to remain prosperous?

(ps, i'm a libertarian)

which brings me back
to the beginning...

intro to [effective]
~~data-sharing~~, economics 101.

free.

(as in beer)

Skynet
DEVELOPMENT TEAM