the next ten years.
starting with the last ten years.
the REN-ISAC

- ren-isac.net/about/index.html
- goo.gl/chH2B
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
what we tried...

• webpage with wget scripting of plain text to firewalls, ids’s and dns servers (2004-2008)

• 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)

• our library did the conversion to your $DEVICE for you (magic)
lessons learned

• plain text was too hard to parse and build into applications (no API)

• XML didn’t scale well (for lots of tiny little messages).

• JSON was too loose to to enforce any kind of “morals” on the data, serialization still took it’s toll...
next version.
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 ms counts)
- easy integration with anything that “carries a message” (zmq, http, smtp, etc...)
XML vs PB

• http://www.oreillynet.com/xml/blog/2008/07/google_hates_xml.html
now for the philosophical stuff...

what we’re working to solve over the next ten years
pardon the social media buzz words here, it’s not what you think...

i promise.
it is now 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/without an actual API)
it is now 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.

• 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)
social network #fail

• they do not allow their hubs to inter-operate with other networks (and therefore 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.. :(
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 someone like you

• the google plus is a social platform for security peeps who have no desire for Facebook’s shenanigans
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?
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
- it seems like we’re actually just inhibiting the data-sharing process (we add no value, we’re just in the way).
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 big data?
cost.
using completely made-up numbers.

- Program X is $1,000,000 million dollars per year (in 2012 dollars, before QE-infinity hits ;))

- primarily in FTE cost. Let says there are 2 pdf's a day for a total of 730 [pdf’s] a year.

- $1,000,000 / 730 = $1,369.86 per pdf to produce
using completely made-up numbers.

- Program X has 350 participant organizations
- It takes each org an FTE 30-60min to digest, triage and act on 2 PDF’s per day
- Each org pays that FTE ~$60k USD (again, 2012 dollars), roughly $30/hr (fuzzy math)
- $30/hr * 350 = $10,500 / day to absorb that information
using completely made-up numbers.

- $10,500 \times 313 \text{ days (≈ business days)} = \$3,232,500 \text{ / year (high end)} \text{ to participate in that program by the industry}
cost.

• we’ve invested 5 years and ~$1.5m into this framework, to reduce the cost for everyone.
Feed Globe

This globe displays all the botnet IPs discovered by feeds on CIF. If you are curious on how to query CIF you can visit my blog post. Or the help file on the project site.

List of Feed providers on feed.josephha.com:
- spambase.org
- ars トラック abuse.ch
- alienvisit.com
- malwaredomains.com
- morganresearchgroup.org - cymru
- anik.org
- defender.alex.sk
- malware.com.br
- malwareblacklist.com
- thesexexpert.com
- malwaredomainlist.com
- malware.com
- panic hit me dump
- phishshark.com
- shadowserver.org
- spytracker.abuse.ch
- infiltrated.net

Credit to Wes, and the CIF Team for making a wonderful product.
are you doing automated data sharing yet?
collectiveintel.net
free.
(as in beer)