

How I (mostly) freed control of my data from evil corporate overlords

Charles N Wyble

Known Element Enterprises

<http://www.knownelement.com>

Presented to UUASC

March 8th 2010

Who am I?

- Charles N Wyble
 - My own little corner of the net is at <http://www.knownelement.com>
 - I have a blog at <http://blog.knownelement.com>
 - And for the ADD generation I have a microblog at <http://mblog.knownelement.com> (mirrors to twitter @charlesnw)

What have I done?

- Lots and lots of system/security/network operations work.
 - Large scale e-commerce operations
 - Disney parks.com and associated properties (30 million uniques per month, 2 billion per year revenue)
 - Evite.com (300 million uniques per month, 2 million e-mails per day)
 - Various random credit card processors, ad companies
 - Lots of contract/consulting/startup work
 - I'm available for hire

What am I up to?

- Building a brand (known element enterprises) ((micro)blog/www/wiki/e-mail/xmpp/sip)
- Striving to own my data
 - Where is my data?
 - Who has access to my data?
- Striving to share my data
 - How/where/when I want with those that I wish to

Why am I doing this?

- Lots and lots of centralized service failures (RIM, Microsoft, Amazon)
 - Links to many on my wiki
- I build systems that allow my clients to own their data, I should drink my own champagne.

How am I going about it?

- Map out where existing data is and determine what can be migrated (ready)
- Evaluate replacement systems (aim)
- Assemble hardware/software infrastructure for hosting (keep aiming)
- Migrate data (trigger pull)
- Host data in sustainable fashion (target down)

Dude wheres my data?

- There are two major types of data
 - Generated Data
 - Created Data

Dude where is my Generated Data?

Data that we generate (we won't be covering this but I leave it as a **HIGHLY** recommended exercise for the audience). It's in many places:

- health data
- financial data
- travel data
- search data
- requests for directions and many other types of information
- Web/mail/chat server logs

Anyone interested in a future presentation on hacking at layer 8? :) (hint google/facebook is the new trust anchor. Own the data, own the world)

Dude.... check out that cloud man.

Data that is created by me. Every day we create a lot of data.

- Textual
 - E-mail
 - Blogs
 - Tweets
 - Social networking posts
- Audio
 - Skype/sip
 - Voice notes
- Other
 - Pictures
 - Videos

Evaluate Replacement Systems (existing proprietary systems)

- News (google news/google reader)
- Photos (flickr)
- Microblogging (twitter)
- Blogging (blogspot)
- Issue tracking and software configuration/project management (github, lighthouse)
- Invocing clients (freshbooks)
- URL shortener (tinyurl/bit.ly)
- Knowledge/data management (google docs and text files/notes on my bb)
- Centralized login (facebook connect/claimid)
- Colloboration (webex/skype)
- CRM (salesforce.com)
- E-mail (gmail)
- Note taking (end note/microsoft one note)
- Fleet tracking (latitude)

Evaluate Replacement Systems (new floss systems)

- Applications I moved to
 - News (rss2email, Dashboard, Tattler)
 - Photos (gallery2)
 - Microblogging (status.net)
 - Blogging (wordpress)
 - Issue tracking and software configuration/project management (git redmine)
 - Invocing clients (Argentum)
 - URL shortener (Casimir)
 - Knowledge/data management (mediawiki)
 - Centralized login (Active Directory/RADIUS/phpMyID)
 - Colloboration (BigBlueButton)

Physical Infrastructure

- Current setup
 - 1 dell optiplex (main-server) hosts all production services
 - 1 whitebox system (dev-server) to test changes before going to prod
 - Motorola DSL modem in bridge mode → Cisco 1841 router → Cisco 3548 switch
 - 2 external USB drives for backups (rotated)
 - 2 APC UPS units (one for dev-server, one for main-server)
- Future plans
 - 1 dell optiplex (mythfe-livingroom) secondary server (dns/mysql/apache)
 - Ec2 instance (activated if internet link fails)
 - Redundant routers/switches (I am pursuing various cisco certs so need this anyway)
 - Whole house UPS/generator
- More information on the current setup, and future plans available on the wiki. Feel free to ask questions as well.

Software Infrastructure

- Operating System: Ubuntu 9.10 server
- E-mail (postfix/dovecot)
 - E-mail is an interesting beast
 - Lots of issues hosting your own due to bad actors and unpatched windows systems :(
- Voice/data/text communication (bigbluebutton)
 - Jabber
 - Freeswitch/Asterisk
- WWW
- LDAP
-

Migrate Data

- Some copy/paste
- Some api based stuff
- Gave me a chance to massively reorganize

Host Data in a Sustainable Fashion

- Backups
- Security
- Monitoring

Backups

- s3
 - cost effective
 - off site
 - easy to implement
 - you wanted to actually restore those backups? :)
- local storage (usb)
 - cost effective
 - want off site and rotation? just buy a few drives
 - easy to implement
 - restores very nicely
- replication
 - o mysql
 - o dns
 - o apache
 - o linux ha

Security

- Software/logical
 - snort/securita
 - greensql
 - logwatch
 - openvas
 - awstats
- Physical
 - alarm system
 - dogs
 - bolting the gear to the rack
 - Other stuff

Monitoring

- Internal
 - OpsView
 - Netdisco
 - Rancid
 - Apt-get install snort logwatch (nice daily summaries of system activity)
- External
 - Nothing at the moment. Evaluating Pingdom and Keynote

The mostly part

- Still using skype but have a migration path
- Still using linkedin but have a migration path (foaf/web of trust)
- Still using bing.com for search, g00g maps for directions.
 - Expensive to solve in a centralized fashion.
 - Cheap to solve in a distributed fashion
- Still using the completely insecure public GSM network (have a migration path - socialwifi.net)

Resources

- Lots more info at http://wiki.knownelement.com/index.php/Data_Ownership
- Keep an eye on www.knownelement.com as I find, implement and use even more data ownership systems.

I'll be back

- I'll have some more to talk about
 - Working on some really cool stuff in the area of Linux networking (wired and wireless) Trying to find an answer to the question "Is it possible to run a network 100% on Linux based systems"
 - Working on some security research (honeynets/honeypots/parallel cracking and fuzzing)