The FreeBSD.org cluster refit

Simon L. B. Nielsen
Hat: FreeBSD.org clusteradm team
BSDCan 2013

Table of Content

- Introduction
- What does admins team do
- The old cluster
- The new cluster
- November 2012 Security Incident
- Questions?

Introduction

Introduction

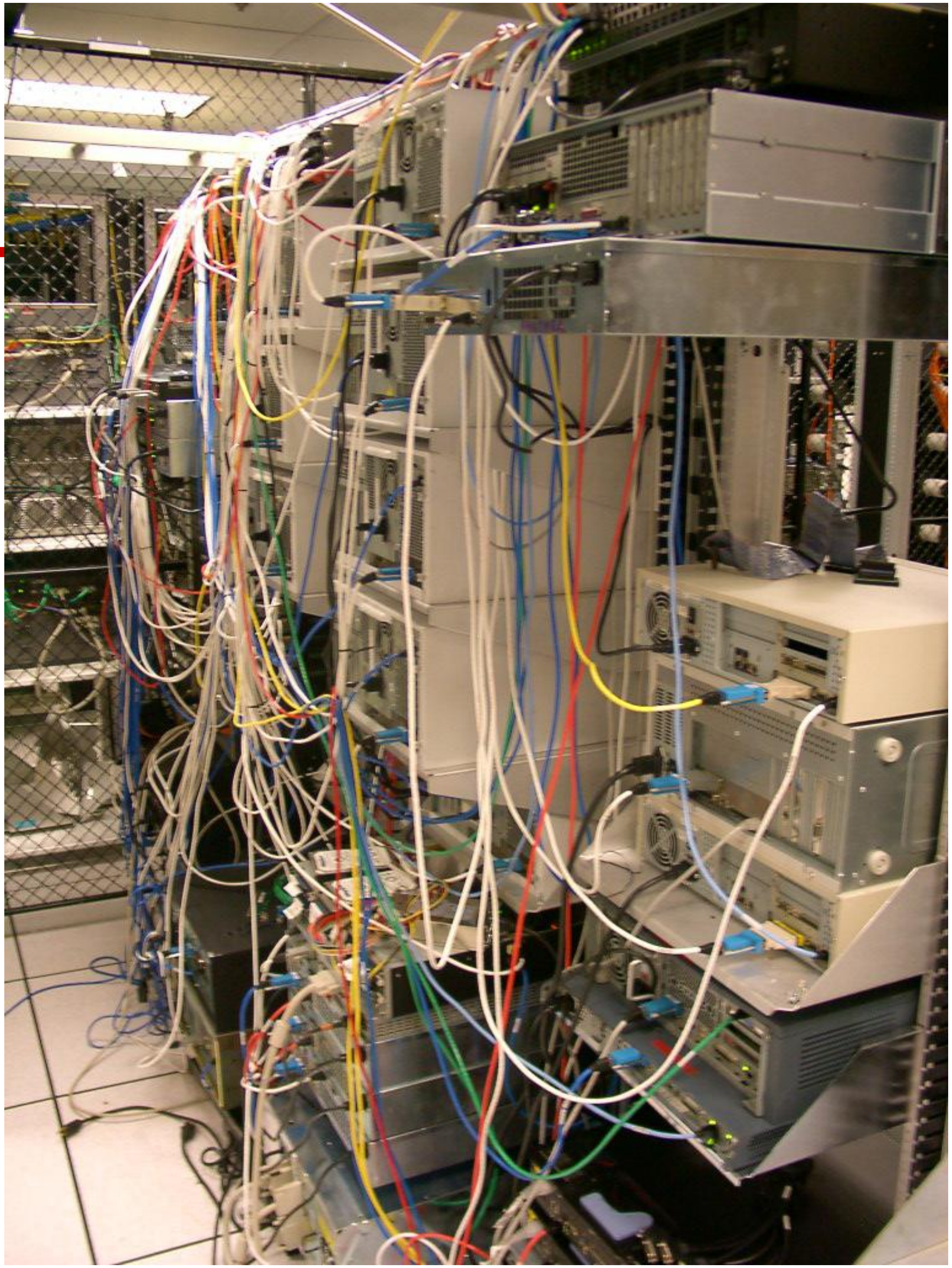
- *Cluster* is just collection of systems
- Primarily cluster at Yahoo!
  - Secondary at ISC, NYI, BME, Sentex
- Work done by many people
- Part of run by subteam, most noticeable
  - clusteradm (9 members)
  - accounts (2 members)
  - postmaster (5 members)
  - Overlap between teams
  - People have real lives
- Most, not all, admins are committers

History at Yahoo!

- SC5, Santa Clara site 5
- SP1, SpacePark #1
- YSV, Yahoo Corp Sunnyvale

- SC5 -> SP1 move ; 2006-11
  - peter@ did forklift
  - Systems moved to new cabinets
  - Forklift move of cabinets SC5 -> SP1
- SP1 -> YSV ; 2012-05 - 2013-05
  - Project evil - sbruno@!
  - Reinstall of everything from scratch
  - A few physical system moved

SC5
2003-09



FreeBSD.org authentication

- Primary authentication is SSH key
- Secondary authentication is PGP
  - Mostly used when primary authentication fails
- Normal users, no passwords... mostly
- Kerberos used for su(8) password store
  - Some use of more traditional Kerberos
- Standalone systems (wiki, forum etc)

What does admins team do

What does admins team do

- Make sure the FreeBSD.org project can function
- Provide support services, which are useful for enough people, and manageable

Admins run services - public

- Email
  - Inbound / outbound SMTP
  - Spam filtering
  - Mail forwarding for @FreeBSD.org
  - Mailing lists (Mailman)
- Web
  - www.FreeBSD.org
    - web build
    - CGIs...
  - people.FreeBSD.org
  - wiki.FreeBSD.org
  - cvsweb / svnweb / p4web
  - portaudit / vuxml

Admins run services (continued)

- Version Control Systems (primary systems)
  - Subversion
  - CVS
  - Perforce
  - Summer of Code SVN (day-to-day by soc-admin@)
- GNATS (day-to-day by bugmeister@)

Admins run services (continued)

- Master mirroring infrastructure
  - ftp-master
  - cvsup-master
  - portsnap-master
  - freebsd-update-master
- Some public mirrors
  - Subversion (all)
  - freebsd-update (only 1 of 4)
  - portsnap (only 1 of 6)

Admins run services (continued)

- Authoritative DNS - FreeBSD.org etc.
  - externally ISC SNS
  - DNSSEC
- Shell server (freefall)
- Developer reference systems
- NFS /home
- Administration support tools
  - admbugs (bugzilla)
  - monitoring (nagios)
  - inventory / tracking (rackmon)

Admins run infrastructure services

- Network
  - Switches (L2 only)
  - Routers (including BGP)
  - Firewalls
  - Inter-site VPN (to NYI, ISC etc.)
  - IPv4 and IPv6 (where possible)
- Authentication
  - ssh keys
  - Kerberos (fancy password store)
- Directory service (LDAP)
  - Previously NIS
- DNS resolvers / recursive DNS
  - DNSSEC validation

Admins run infrastructure services (continued)

- NTP
- Serial console infrastructure
  - Console servers (Cyclades, OpenGear etc)
  - conserver
- Remote power controllers
- Backups
- audit
  - auditdistd
- Netbooting infrastructure
  - System install
  - Emergency recovery

"Hosted" services

- Portsbuild (portmgr@)
    - pointyhat
    - build nodes
- git-beta.FreeBSD.org (uqs@)
- portsmon.FreeBSD.org (linimon@)
- freebsd-update (security-officer@)
- portsnap (security-officer@)
- foundation.FreeBSD.org
- Redports (decke@, portmgr@)
- Core team election (des@)
- Coverity Prevent (uqs@, philip@)

What FreeBSD.org does not run

- FTP mirrors
- Most cvsup mirrors
- Most www mirrors
- Most *cc*.FreeBSD.org DNS
- tinderbox.FreeBSD.org (des@)
- forums.FreeBSD.org (forum-admins@)
- portscout.FreeBSD.org (zi@)

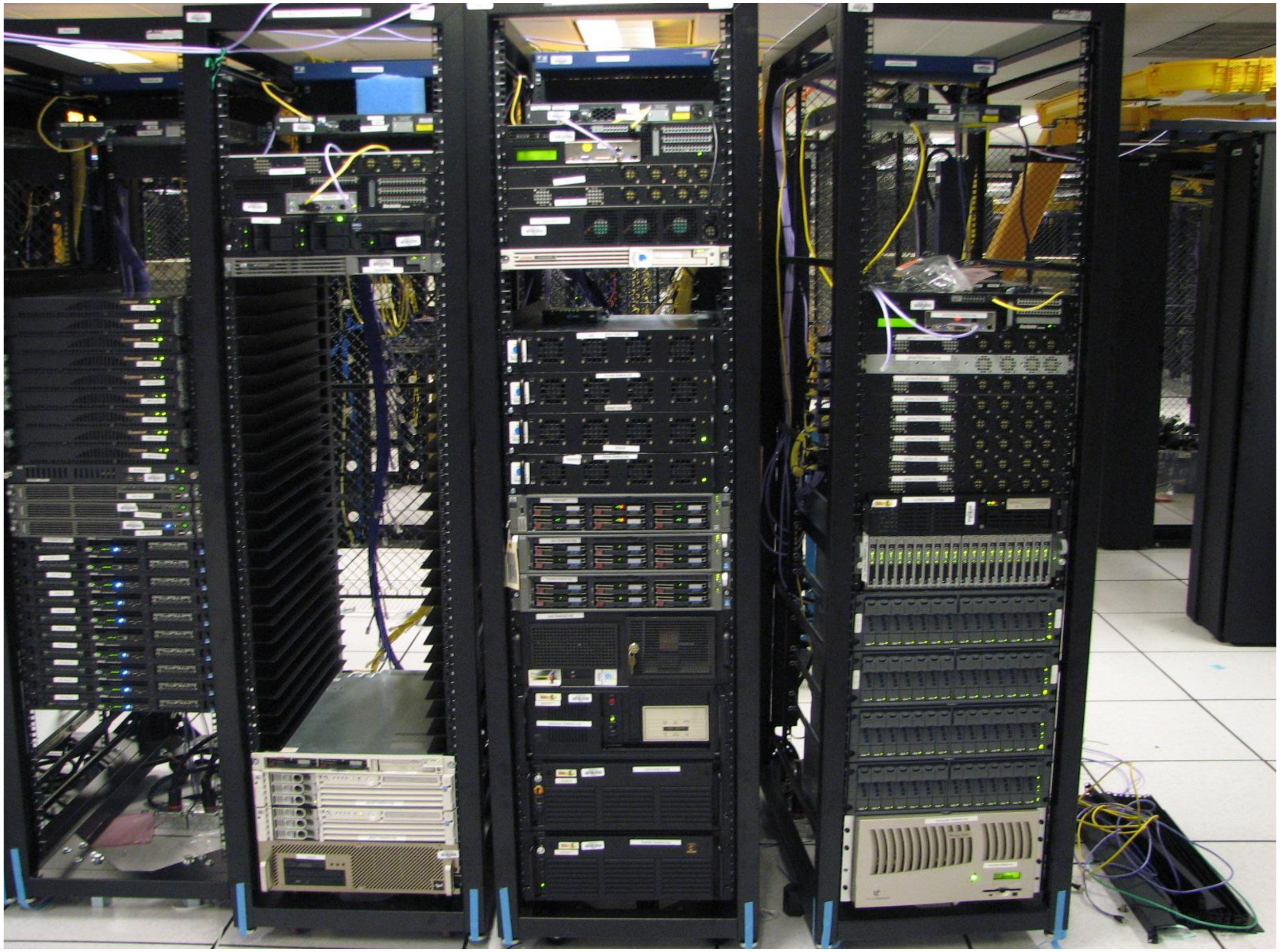
Services may be onboarded later.

The old cluster (at SP1 / SC5)

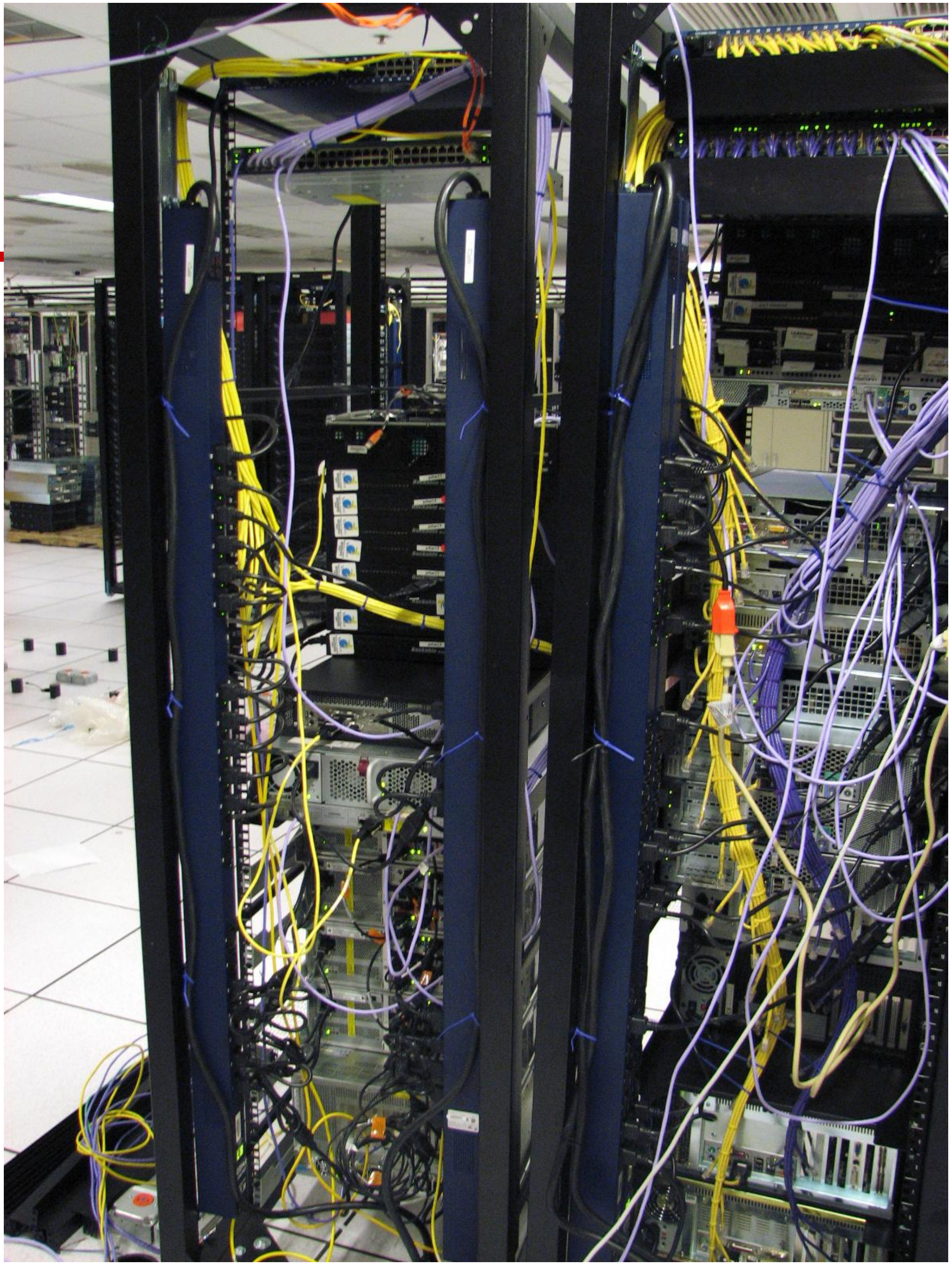
The old cluster

- Flat network, for everything
  - Shell server (freefall)
  - Web server (www)
  - Package building servers (pointyhat, and more)
  - Infrastructure (DNS, NIS etc)
  - CVS, Perforce etc. servers
- NetApp filer for NFS /home
- NIS for accounts (no passwords)
- NFS share for ssh keys
- Kerberos for su (to root etc.)
- Firewall was Yahoo! route filters
- IPv6 via tunnel from ISC

SC5, pre move, 2006-11
Close to later SP1



SC5, cabinets 2006-11



Single points of failure (old cluster)

- ssh-keys on dumpster NFS share
- /home on dumpster NFS share
- freefall, hub, repoman NFS cross mounts
- Single conserver
- One big security domain

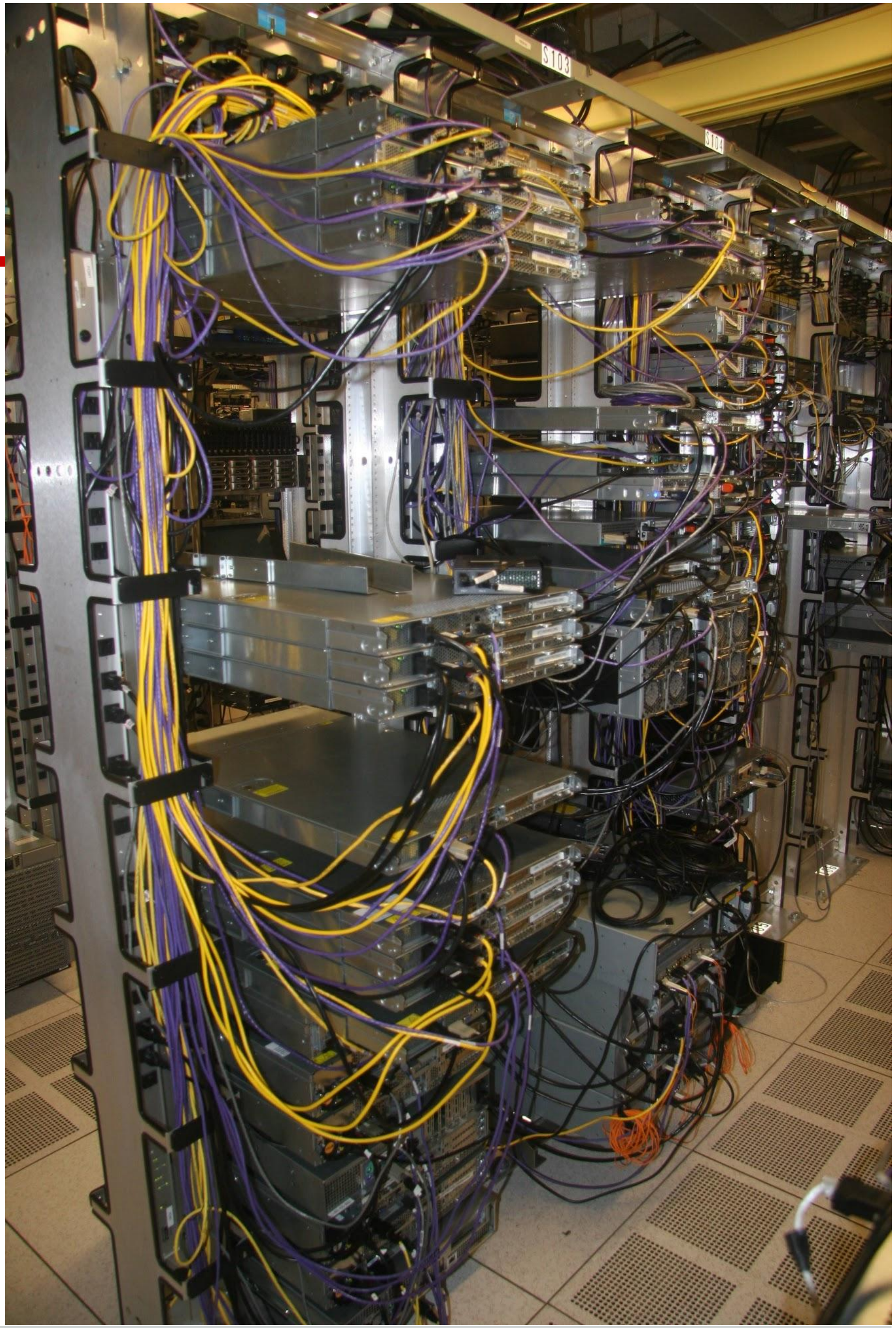
The new cluster (at YSV)

YSV introduction / goals

- Flexible networking
  - Own firewalls
  - Own switches
- As much FreeBSD as possible
- As much separation as possible
  - VLANs
  - Jails, jails, jails, and jails... and jails
- As little NFS as possible
- NIS... bye bye
  - LDAP to replace NIS
- Most hardware donated by Yahoo!
- FreeBSD Foundation helps when needed

YSV
2013-04-30

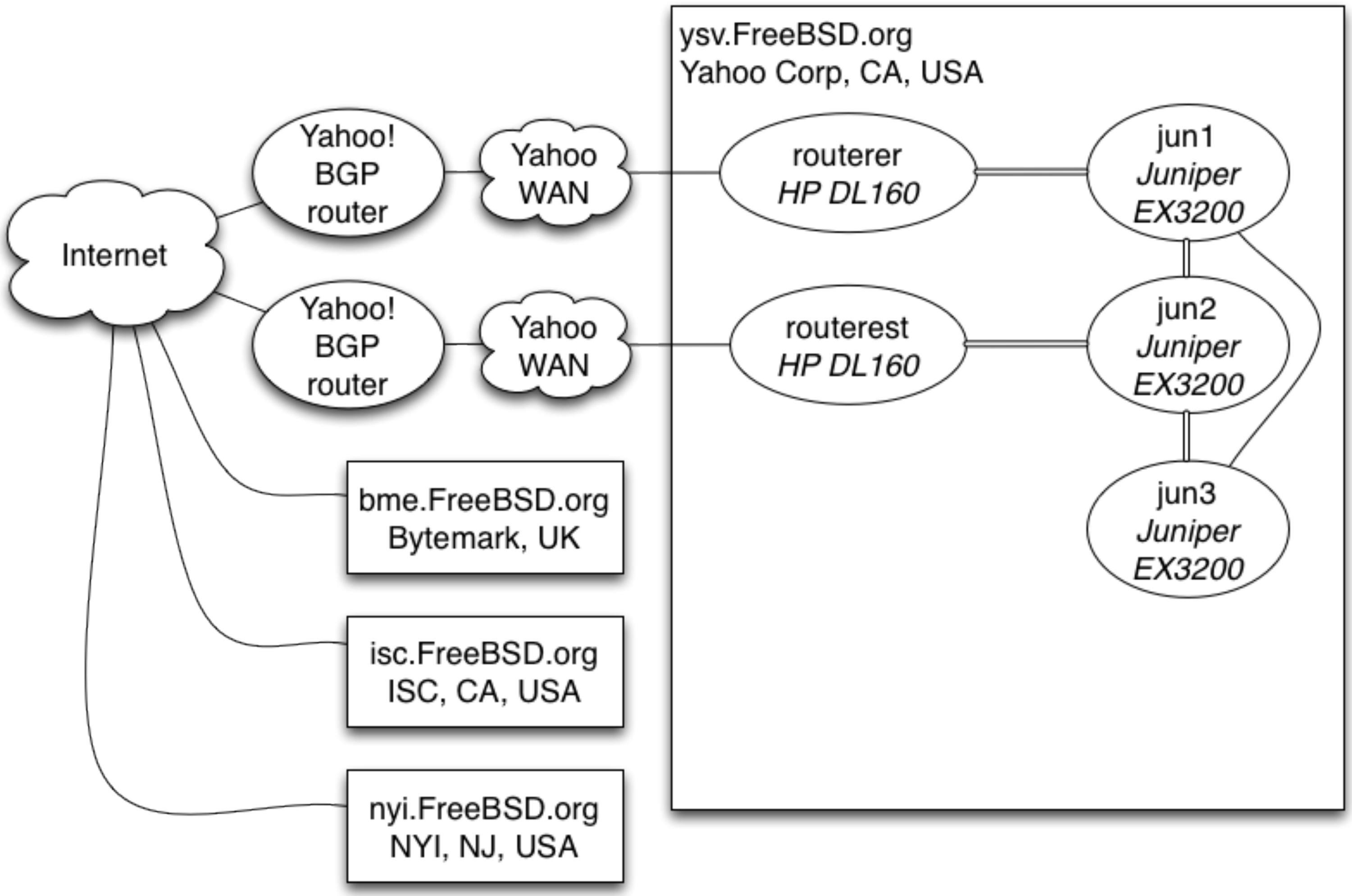
YSV
2013-04-30



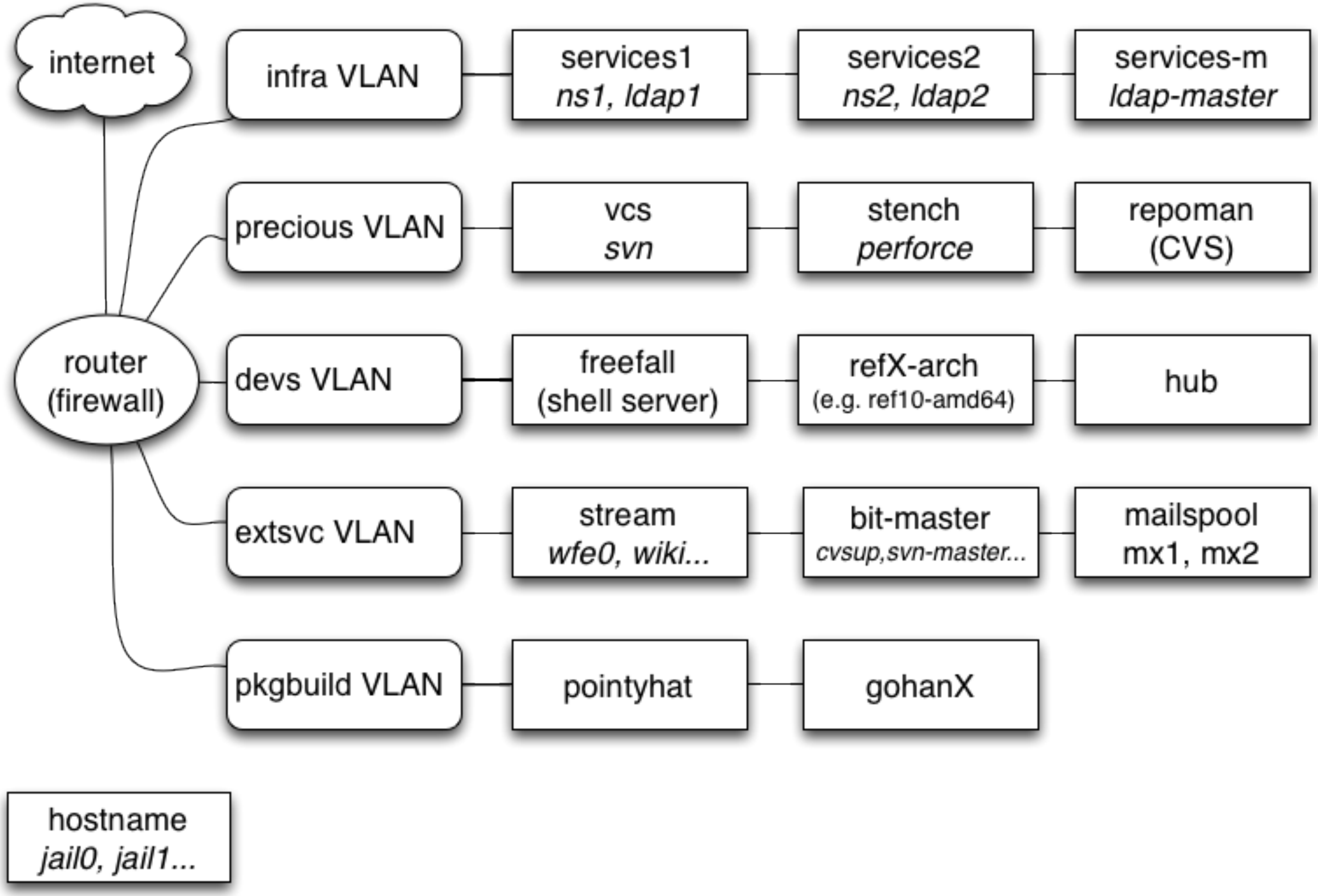
YSV Network

- BGP uplink...
- v4 + v6 (native)
  - v6 only hosts / jails
- FreeBSD.org firewalls
  - 2 firewalls, different racks
  - pf
  - carp
  - openbpgd
  - IPsec VPN
  - FreeBSD 10-CURRENT
- FreeBSD.org-managed switches
  - many VLANs
    - peter vs. simon compromise

YSV, network uplink



YSV, logical network



internet

router (firewall)

infra VLAN — services1 *ns1, ldap1* — services2 *ns2, ldap2* — services-m *ldap-master*

precious VLAN — vcs *svn* — stench *perforce* — repoman (CVS)

devs VLAN — freefall (shell server) — refX-arch (e.g. ref10-amd64) — hub

extsvc VLAN — stream *wfe0, wiki...* — bit-master *cvsup,svn-master...* — mailspool mx1, mx2

pkgbuild VLAN — pointyhat — gohanX

hostname *jail0, jail1...*

Jails

- ezjail
- Use on ZFS and UFS
- jail_interface in rc.conf
- shared stuff dumped in basejail/etc/
  - resolv.conf
  - periodic.conf
- Sendmail for null mailer
- Many v6-only jails
  - Fewer than we like (distfiles etc)
- nullfs RO cross mounts for data sharing

Web serving

- Varnish frontend
- nginx for HTTPS
  - Send all traffic to varnish
- Backend jails:
  - Static pages
  - wiki
  - svnweb (viewvc)
  - p4web
  - CGI...
  - Mailman
- Most backend servers are Apache
- Separate web build jail
  - nullfs RO into static serving jail

New user directory, NIS

- NIS worked OK
- Quirks
  - Large NIS groups = many UDP packets
    - small packet loss = lot of noise
  - Only unix authentication
  - No built in nice replication
  - No built in handling of multiple sites
- Security...

New user directory, LDAP

- Flexible schema
- Built in replication
- Built in integrity protection (SSL)
- Widely supported

FreeBSD.org LDAP user

```
uid: simon
cn: Simon L. B. Nielsen
loginShell: /bin/tcsh
uidNumber: 982
gidNumber: 493
homeDirectory: /home/simon
gecos: Simon L. B. Nielsen
...
```

FreeBSD.org LDAP user (conti)

```
...
adminShell: /bin/tcsh
sshPublicKey: ssh-rsa AAAA...
adminPublicKey: ssh-rsa AAAA...
objectClass: account
objectClass: adminAccount
objectClass: freebsdAccount
objectClass: posixAccount
objectClass: soAccount
```

New user directory, LDAP

1. nss_ldap
   - The obvious choice
   - Increases lookup latency significantly
     - Even with nscd
   - Total dependency on LDAP server working
   - Limited flexibility in login policy
   - Does not handle ssh keys
   - Depend on ports installed on all systems
2. Generate passwd / group locally
   - Perl script, requires LDAP modules everywhere
   - Script used CVS pserver...
3. Now update.sh

update.sh - why?

- LDAP server can be down
- Clients only need base system tools
- Can be very flexible in defining policies for host

update.sh - the new wheel!

- Each server has a role
  - admin
  - developers
  - ...
- Builds and merges
  - master.passwd
  - group
  - /etc/ssh-keys/
  - /root/.k5login
- Distribution tarballs created and signed
  - Strongly freebsd-update / portsnap inspired
- Made available via plain HTTP
- Clients run update.sh every 10 minutes

Why not puppet, CFEngine etc.

- Limited experience with puppet etc. in clusteradm
- Expected it would take too long to set up
- Most likely do it in the future

November 2012 Security Incident

Slides mostly by Peter Wemm

Introduction

- In November 2012 an SSH key was used to gain cluster access from a developer's personal machine.
- Attackers used package build infrastructure as a foothold.
- Was quickly identified and shut down but took a long time to validate and rebuild.

Initial SSH Key theft

- They obtained a developer ssh key
  - No passphrase
- Every machine in the cluster trusted it
- Had access to a multiple remote sites with password-less sudo

What happened

- David Wolfskill noticed while they explored.
    - We all owe him for catching this early
- The attackers didn't capitalize
- Were able to get root access, without exploits
- Reached the CVS repository with r/w access
    - Extremely difficult to audit/validate.
- Did not reach svn but svn was audited anyway

Poor communication

- Tracking down what had happened was done quickly, except for CVS history
- We were confident end users weren't affected and no tainted data was distributed.
- We wanted to give an advisory that included not using CVS
  - .. but the documentation still said to use CVS!
- Confusion over who was doing what lead to unacceptable delays on the advisory

Going forward

- Organizational
  - Resolve confusion over responsibility.
  - Responsive administration vital.
    - avoid people resorting to subverting the system.
- Killed all tainted and weak ssh keys
- Significant cleanup of portsbuild
- Rebuilt and modernizing infrastructure
- Multiple authentication factors

Q/A

Spare slides

Current clusteradm

9 current members:

- Peter Wemm - peter@
- Ken Smith - kensmith@
- Simon L. B. Nielsen - simon@
- Bjoern A. Zeeb - bz@
- Brad Davis - brd@
- Sean Bruno - sbruno@
- Glen Barber - gjb@
- Ryan Steinmetz - zi@
- Ben Haga - bhaga@

Current postmaster@

- David Wolfskill - dhw@ (Mr postmaster)
- Florian Smeets - flo@
- Sahil Tandon - sahil@
- Brad Davis - brd@
- Jonathan M. Bresler - jmb@ (previous Mr postmaster)

Netapp Filer 2006-06

Note: Floppy

4x500GB storage!