F-Root Updates

Daniel Mahoney

DNS-OARC 32

https://www.isc.org

Intro

- Who am I?
- Brian Reid retired from ISC October 2019
- Ray Bellis is now the Director of DNS Operations
- Elmar Bins (DENIC, NIC.AT) is now working with us as well.
- This is a high level “things we’ve learned” talk.

Aging F-Root

- 8-14 U of Kit
- 2 Transit Providers
- Originating a /24 and /48 for management
- 2 enterprise servers, most from well-known manufacturers, but a frustrating mix of remote access technologies — serial KVMs, Java, some which required ActiveX controls.
- A recent push to pave over and mass upgrade to latest BSD was successful, but that's another presentation.
- Power Hungry
- Reaching the end of useful life

Aging Issues

- Even with maxed out ram, the size of a full RIB from multiple peers and providers approaches useful limits.
- Reinstalls become necessary as OSes get heavier.
- Remote hands is often the only reinstall option — across time zones and with no OS familiarity. The FreeBSD installer can be scary versus mouse-driven Ubuntu installs, for example.
- Many exchanges are moving up to 10G speeds, which existing routers cannot do.
- Replacement of parts limited by local availability.
- Layer9 problems. Often, a sponsor would graciously arrange for hardware to be donated, and move on.

Partnering with a CDN

- Partnering with a CDN, we are able to complement each others' growth strategies
- Increased network reach and coverage
- Allows ISC to focus on coverage versus complete overlap.
- Smaller deploys, embedded sites, newer and growing IXes
- ISC is still a global network operator

New Hardware development

- This slide also known as “my love letter to Dell”
- We’ve seen a lot of funky hardware, but Dell’s machines are solid.
- Smarter BMC than many other vendors, where IPMI feels “stapled on”.
- Large install base means attention to vulnerabilities and issues. Regular updates which are easily installable.
- Never again having to see Java warnings

More iDRAC praise

- iDRAC reachable via SSH, https, SNMPv3.
- Can remote mount an ISO via SMB, using only a few ssh commands.
- Java-free remote console, and media mounting via browser
- Can redirect hardware serial console into an ssh session, thus you can drive the OS installer totally via a screen/tmux session.
- With an SDcard installed, we can upload that ISO and boot natively, way faster than a CD would ever be.
- Certificates deployable signed with an internal root cert. (Let's Encrypt rolls too frequently).
- Can reset most bios settings via SSH (without driving it via a screen share).

New Software

- We could not have done this ten years ago.
- Current software routing stack landscape is much more solid — including alternatives. (BIRD, FRR)
- The term “Devops” wasn’t a thing when F was first deployed. The only tools for mass deployment were in-house at large ISPs, or things like “rsync and a version control”. This landscape has changed.
- OSes with multi-fib support.

Multi-fib support, why?

- Often asked question.
- If you have both a static default route, as well as lots of learned peers, a process will prefer the interface address closest to the traffic.
- Example: packets sent back to the mothership, in an exchange where HE.NET is seen as a peer via a route-server. HE provides transit for ISC.
- Exchange addresses are not globally routable, nor are they RFC1918.
- Adding a second fib for the named process to use (but still adding the system's default route to the new fib) allows named to use the complex routing table, but management processes to be a simple static default.

Deployment Strategies with sponsors

- VAT, customs, and other weird logistics issues mean we ask local sponsors to acquire hardware for us (some even already have a channel to do so). The power savings alone over previous design make this common sense.
- Dell part numbers vary country-to-country, so we ship a descriptive buy-list, rather than part numbers. This also means when Dell bumps revisions, it still makes sense.
- We always recommend dual 10g fiber ports, which can also do 1G with correct optics. Many DC's don't do copper anymore.
- Solid warranty required. We recommend at least a 5 year.
- Dual PSU, iDRAC Enterprise with vFlash, 8G RAM, hardware raid.
- This single-box solution is NOT a drop-in replacement, however…

Replacing the old with the new

- A good portion of the replacement process is re-provisioning.
- Where previously a provider only gave us a /30 that we spoke BGP over, we now require multiple drops and IP's configured for iDRAC and OS separately.
- Since we no longer originate a local /24, the host must make arrangements, either manually or via BGP to allow DNS responses (5-10 percent of traffic, typically) to leave via the management connection.
- MAC-Address filter updates, and often custom BIRD configs to match exchange-specific communities apply.

Software Platform

- Currently using FreeBSD (stable, but non-zero release) and BIND stable, but the goal is to allow for install of same toolchain on Linux.
- Since rebooting into a new OS installer is a few commands to the iDRAC via ssh, this is an easy way to add resiliency.
- We maintain our own FreeBSD package building architecture so we can deploy custom versions of BIND or other packages with non-default options as needed.
- Goal of F-Root control plane being completely different from ISC corporate networks or software engineering systems.

Routing Platform

- In the past, nodes used Quagga solely as a signaling protocol, and announced the F-Root prefixes to local routers, which were then advertised as behind AS3557.
- Current systems run BIRD2 and speak to any peers (either direct peers or route-servers) directly, update a FIB on the box.
- Scripts were written to translate all old Cisco-style configs to BIRD configs.
- Multi-fib.
- We continue to use a unique AS per-site as this is one of the only ways to detect routing leaks (BGP looking glasses don't know about hostname.bind or NSID)
- A number of health-checks operate on BIND and will withdraw the route if there's an issue — this is done by removing the F-root address from the loopback, rather than trying to script the routing daemon.

Puppet

- We use Puppet (open source) as a deployment tool, but currently use it more like most people use Ansible.
- In an ideal world, the only post-install task would be ‘pkg install puppet’
- All work done via puppet manifests, but with abilities built in to allow sane overrides during times of issue without puppet “helping” and reverting.
- Puppet being used as a source of truth for other systems (monitoring).
- CI being used for puppet (puppet-lint, puppet parse tests, testing environments, use of testing VMs with rollback).

Monitoring

- Previously: Cacti, Netdot, Nagios
- Now: Nagios, Puppet as backend/source of truth, basic traffic graphing via SNMP tools, probing all hardware that the iDRAC knows how to enumerate.
- Third-party monitoring tools being added and evaluated regularly.
- All sites sync pcaps back to the mothership on regular basis for analysis.
- RSSAC Stats being provided to the general public.
- Custom network tools that make use of the ATLAS probe network to detect routing leaks.

Future Work

- We maintain a large network of peering partners, and aim to make it more scriptable, monitor-able, but within reason.
- This wasn't possible on classic Cisco routers (Cisco didn't make ipv6 peers easily enumeratable via SNMP), and screen-scraping was annoyingly buggy.
- With network-as-code, this can be easily deployed — could auto-peer via PeeringDB, or bring up sessions in passive mode and alert on new peers, warn after flaps for N period of time, or N number of flaps.
- Smarter filtering detection, and detection of routing leaks, perhaps with auto-withdrawal.
- Perhaps sharing snapshots of our views of the global routing table with the various route-collector projects.
- We've started rollout of RPKI on some of our corporate networks, will enable it for the F prefix at some point.

Wrap Up

- ISC is a public-benefit not-for-profit dedicated to serving the Internet. We love what we do.
- F-Root relies on cooperation from local operators of exchanges and networks.
- We want to hear from you!