ESCC IPv6 Deployment Status & Issues

ESnet

Michael Sinatra

Energy Sciences Network

ESCC – January 2012

Overview

Updates and items of interest

Key addressing concepts

Survey results

Site discussions

- Addressing
- 2012 deliverables: What's working, what's not, lessons learned

More discussion

What's happening in v6-land?

- World IPv6 Launch
  - June 6, 2012
  - We will not merely be turning on IPv6 on some websites for 24 hours, then turning it off.
  - We will not merely be working with *websites* or *content.*

What’s happening in v6-land?

- World IPv6 launch
  - We’re going to turn it on and leave it on. Content providers involved:
    - Facebook (www.facebook.com)
    - Google (www.google.com)
    - Microsoft Bing (www.bing.com)
    - Yahoo! (www.yahoo.com)
    - and probably more…
  - But we’re not just doing content…

What’s happening in v6-land?

- World IPv6 launch
  - Eyeball networks have pledged to enable IPv6 for at least 1% of regular customers by June 6, 2012:
    - AT&T
    - Comcast
    - Free Telecom
    - Internode
    - KDDI
    - Time Warner Cable
    - XS4ALL
  - But you can’t get to the IPv6 broadband Internet unless you have…

What’s happening in v6-land?

- World IPv6 launch
  - A BRAND NEW CPE ROUTER!!...
    - Cisco/Linksys
    - D-Link
  - These brand new IPv6-enabled home routers can be YOURS if…
- Anyway, World IPv6 Launch *should* increase overall IPv6 traffic, although Happy Eyeballs might diminish some of this.
- I have some ideas as to how ESnet can help with this effort (connectivity testers, looking glass, etc.) and will be working with Greg and the Tools Team in the next few months to see what we can get implemented.
- What would you like to see?

What’s happening in v6-land?

- Stuff we’ve been doing:
- Briefed a government agency on IPv6.
  - Discussed addressing and the end-to-end nature of the protocol.
  - Talked about large-scale NAT on the horizon.
  - Was very well-received
- Not just tooting my horn here: This shows how government agencies can communicate and collaborate on this stuff.

What’s happening in v6-land?

- Federal IPv6 Task Force: Phil and I are involved.
- A lot of stuff being discussed.
  - New tool to document obstacles.
  - Addressing, addressing, addressing.
    - Ron Broersma gave a talk on addressing (slides available)
    - DoE is interested in addressing; particularly how much collaboration should we be doing? Should there be guidelines?
    - An engineering contractor even went so far as to say that we “should follow the IETF and IANA” and do only provider-assigned addressing from higher level DoE networking entities. (That presumably means DOEnet and ESnet.)
    - → I don’t agree with this, although ESnet is happy to make PA space available to sites.

What's happening in v6-land?

- More on addressing:
  - There's some desire for collaboration among DoE offices and sites regarding addressing (e.g. what works, lessons learned).
  - But there isn't much support for a true top-down addressing model.
- Task force will continue to meet monthly.
  - DoE is pretty well-represented.
  - So is DoC, VA, and a number of other agencies.
  - VA is doing very well.
  - DREN (John Baird) wants to help.
  - → There is plenty of room for collaboration among federal agencies (did I just say that??).

What's happening in v6-land?

- News from JT:
  - Casey Deccio's dnsviz.net now checks IPv4 and IPv6 separately (useful for MTU/firewall/etc. issues).
  - Conference venue didn't have IPv6 on the wireless → Internet2 fixed!
  - Box contract doesn't have any IPv6 stipulation, and Internet2 (and mainly the member campuses that were driving this—including UCB) apparently didn't think this was important.
  - → Both of these issues caused a lot of fur to fly (that's good).
  - → Box issue has implications for DoE sites!
  - → Outsourced/"cloud" deals *must* take IPv6 into consideration.

Other things that happened this year

- All of ESnet's perfSONAR nodes became IPv6-capable.
- Naming question:
  - perfSONAR tools prefer IPv6 over IPv4. This is important to remember!
  - But they are arguably "public-facing services." How do we name them so that people know which protocol they're testing?
    - psnode.example.com → both A and AAAA records
    - psnode-v4.example.com → A record only
    - psnode-v6.example.com → AAAA record only
    - This satisfies the mandates and allows people to test each protocol separately.

Other things that happened this year

- ESnet’s IPv6 routing table became complete (pretty much)
  - Cogent de-peering had put a hole in ESnet’s IPv6 routing table, since Cogent won’t peer with HE.
  - ESnet bought transit from HE, so we weren’t getting HE IPv6 routes. Some Level3 routes also would have been missing, but we peer with Level3.
  - Solution: Get transit from Level3 as well.
  - IPv6 transit with Level3 turned up on ~12/12/11.

Other things that happened this year

- We began to better understand NDP cache exhaustion.
- Scanning an entire /64 block can kill your router before you get 0.01% through the subnet!
- The problem is that routers have limited space to store NDP entries (the IPv6 equivalent of ARP).
- Scanning a /64 with very few hosts on it (in the world of IPv6, 200 hosts on a /64 is *very few!*) can cause your router to run out of NDP entries.
- Requires software patches and some standards tweaking to fully fix.
- In the *near term* might want to reserve /64s for subnets but set the prefix to something lower.
  - /120 = 255 hosts; /127 = 2 hosts (ptp router links)

Some addressing concepts leading into survey results

IPv6 subnetting (how are we doing it?)

NAT in IPv6 (not!) (or not yet!)

ULAs for IPv6: They’re not the same as RFC 1918!

IPv6 Addressing: Subnetting

- Why reserve an entire /64 for a user LAN?
  - It makes host address management much easier and more flexible. This will be discussed in a later module.
  - It makes things easier for *you.*
  - It is what is pretty much guaranteed to be supported by vendors. (I know, “pretty much guaranteed….”)
- In general, you will want to configure /64s for user LANs.
  - You will generally *not* want to configure any prefix shorter than a /64 for a user LAN. /48s or /56s are a waste and probably won’t work right.
  - You will want to come up with an addressing plan for all of that space.

IPv6 Addressing: Point-to-point

- Originally, point-to-point links (between routers or other devices) were intended to be /64s.
  - This seems crazy. Why use a full subnet, theoretically capable of numbering every host on the planet, for just two hosts??
  - Remember, you have an enormous amount of address space to work with. It's not about conserving addresses—it's about keeping your address space organized so that you—and your routers—can easily keep them straight.

IPv6 Addressing: Point-to-point

- More recently, it has become a best practice to use /126s or /127s as point-to-point links.
  - This is not to conserve addresses! It's actually for security reasons.
  - Two issues: ping-pong routing and neighbor-discovery table exhaustion.
  - See http://tools.ietf.org/html/draft-kohno-ipv6-prefixlen-p2p-03
- → You may want to configure point-to-point links with /127s, but reserve the entire /64 for *each link!*
  - This guards against misconfigurations elsewhere in the backbone, where someone accidentally configures a connected interface with your entire point-to-point block as a /64 and black-holes it on that router!

Addressing plans: The basics

- You will need to decide how to divide up and subnet your /48 or /32.
  - An important consideration is whether to divide the address space organizationally or by router/PoP.
  - Router-based addressing makes things easier on your routers, but organizational addressing may make things easier on your brain.
  - You may also want to address geographically (e.g. by building) or by service (web servers, wireless, voip, VMs, etc.).
  - Combinations of the above are possible, given the size of the address space.

Addressing plans: The basics

- Some reminders:
  - Avoid IPv4-centric thinking. You should no longer be concerned primarily with conserving *addresses*, you should be concerned with conserving *assignments and/or allocations* and ensuring that your address space remains relatively unfragmented, whatever the basis of your plan*.*
  - Don’t think this will be your end-all and be-all addressing plan. Keep in mind that it may have to be re-done.
  - See RFC 5375 for a discussion of addressing plans.

Carving up address space

- You will want to make your address space as contiguous as possible so that it can easily be aggregated for you routers and so that it makes things more organized for you and your staff.
  - You could just address sequentially.
    - Department (or router) 1 gets 1 /62
    - Department (or router) 2 gets 1 /62
    - Department (or router) 3 gets 1 /60
    - Oops, now department 1 needs another /64. Your space is already fragmenting.
  - Fortunately, there are (simple?) ways of dealing with this, depending on how much binary math you want to do.
  - This is probably something that you once did for IPv4, but your IPv4 space may be so fragmented that it's a lost cause.

Carving up address space

- One method: Bisection, as discussed by Owen DeLong of Hurricane Electric.
    - Address space is a pie. Keep cutting it in half until you have the correct sized pieces.
    - Suppose I have a /48 and I currently have 20 departments. Think about how my organization may grow (or not), I may want to split the space up into 32 chunks. Using the bisection method, I would cut my /48 in half five times so that I now had 32 /53s.
    - → Be sure to set aside blocks for infrastructure: routers, switches, DNS servers, etc.
    - → You may also want to ensure that the divisions fall on nybble boundaries because this makes it much easier to break up blocks in your head. It's also much easier to delegate DNS, as we will see. (You may want to stick with /52s, /56, /60s, etc.)

Carving up address space

- Another method: “A flexible method for managing the assignment of bits of an IPv6 address block”: RFC 3531.
  - Assign blocks on the leftmost (bit-wise) portion of your address allocation/assignment by manipulating the leftmost bits.
  - Assign blocks on the rightmost (bit-wise) portion of your address block using the rightmost bits.
  - Assign others using the centermost bits.
  - This allows you to assign blocks of different sizes while maintaining maximum flexibility. You fill in space around bit boundaries last so that if more (less) space is needed, you can easily double (halve) a block because your bit boundaries are flexible.
  - But it’s hard to do (and keep track of) in your head!

Carving up address space

- You can combine methods. In a way, bisecting is pretty much a leftmost-bit-wise method of dividing the initial block. Bisecting and then assigning nets using centermost bits may help keep your bit boundaries flexible.
- A full implementation of RFC 3531 is probably overkill for an end site.

Addressing plans: IPAM

- Once you have an addressing plan, you need to manage your assignments. IP Address Managers do this, and some do it quite well.
- Flat files and vi may have worked for IPv4, but they probably won't work for IPv6.
- Of course, your IPAM may not support IPv6 at the same level as IPv4 (if at all), so flat files and vi may be all you have.
- Some commercial IPAMs that support IPv6 (or at least claim to—YMMV):
  - BT Diamond IP
  - Nixu NameSurfer
  - Men and Mice IPAM
  - Infoblox
  - BlueCat Proteus
  - Probably others!

Addressing plans: IPAM

- Some open-source IPAMs that support IPv6 (or at least claim to—YMMV):
  - IPplan (6.00 beta only)
  - NetDB (Stanford)
  - NETDOT (U. of Oregon)
- → IPv6 support in IPAMs is improving rapidly.

What about private addresses and NAT?

- Remember, we now have tons of address space. We don’t need NAT to conserve addresses.
- Therefore, there is no NAT functionality in IPv6.
- What about the security provided by NAT?
  - Whether NAT provides any security is debatable, especially considering the rise of social engineering attacks and session hijacking, and the general decline of network-based (e.g. scanning) attacks. In fact, sequential scanning is much harder in IPv6.
  - The same level of security can be provided by filtering a portion of one’s IPv6 space within the network and applying a stateful firewall to allow outbound connections. This would have the same effect as a stateful NAT box.
- But should we really be using globally-routable IPv6 addresses when they will only be used within our network?

ULA vs. globally-routable addressing

- IPv6 does have a concept of private addressing. It was originally called site-local addressing and used addresses within FC00::/7.
- For a variety of reasons (see RFC 3879) the site-local addresses were deprecated and in RFC 4193, they were replaced with *Unique Local Addressing* (ULA).
- ULAs are self-assigned out of FD00::/8 in blocks of /48, using a pseudo-random method. The idea is to keep the "U" in ULA, so that ULA addresses from different organizations will not collide.
  - Prevents issues arising from leakage of address information.
  - Makes the merging of networks with ULA space much easier.
  - Makes troubleshooting easier, too. How many times have you seen addresses in 192.168.0.0/24, 10.0.0.0/24, or 192.168.1.0/24. showing up on your networks? There's plenty of space defined by RFC 1918, but people often use the same /24s over and over again, so it's hard to see who is doing what with the same space.

IPv6 private addressing

- But ULAs are not all milk and honey.
  - Within just about every community (standards, operations, policy), there are concerns about ULAs. In many cases there are strong objections.
    - Embeds routing policy in addressing.
    - Has many similar problems as does RFC 1918.
    - No clear idea of how ULAs evolve in the future and whether they will ever be carried in the global internet.
    - Part of the initial idea of ULA was to provide PI addressing. Now that most RIRs provide PI addressing, this is less of a reason for ULA.
  - There is a violent lack of consensus on the value of ULA.

IPv6 private addressing

- Because of this lack of consensus, you should be concerned about using ULAs.
  - There is currently no guarantee that they won't someday be routed. If you are going to operate on the assumption that your ULA-addressed hosts can't possibly be exposed to the Internet, then you may be surprised to someday find out that people on the Internet can suddenly reach your hosts.
  - Or, you may be using space that you think is "yours" and the ULA policy changes and an RIR assigns or allocates it to someone else.
  - Better to just use existing space assigned to you and block it at your border, if you want it to be private.

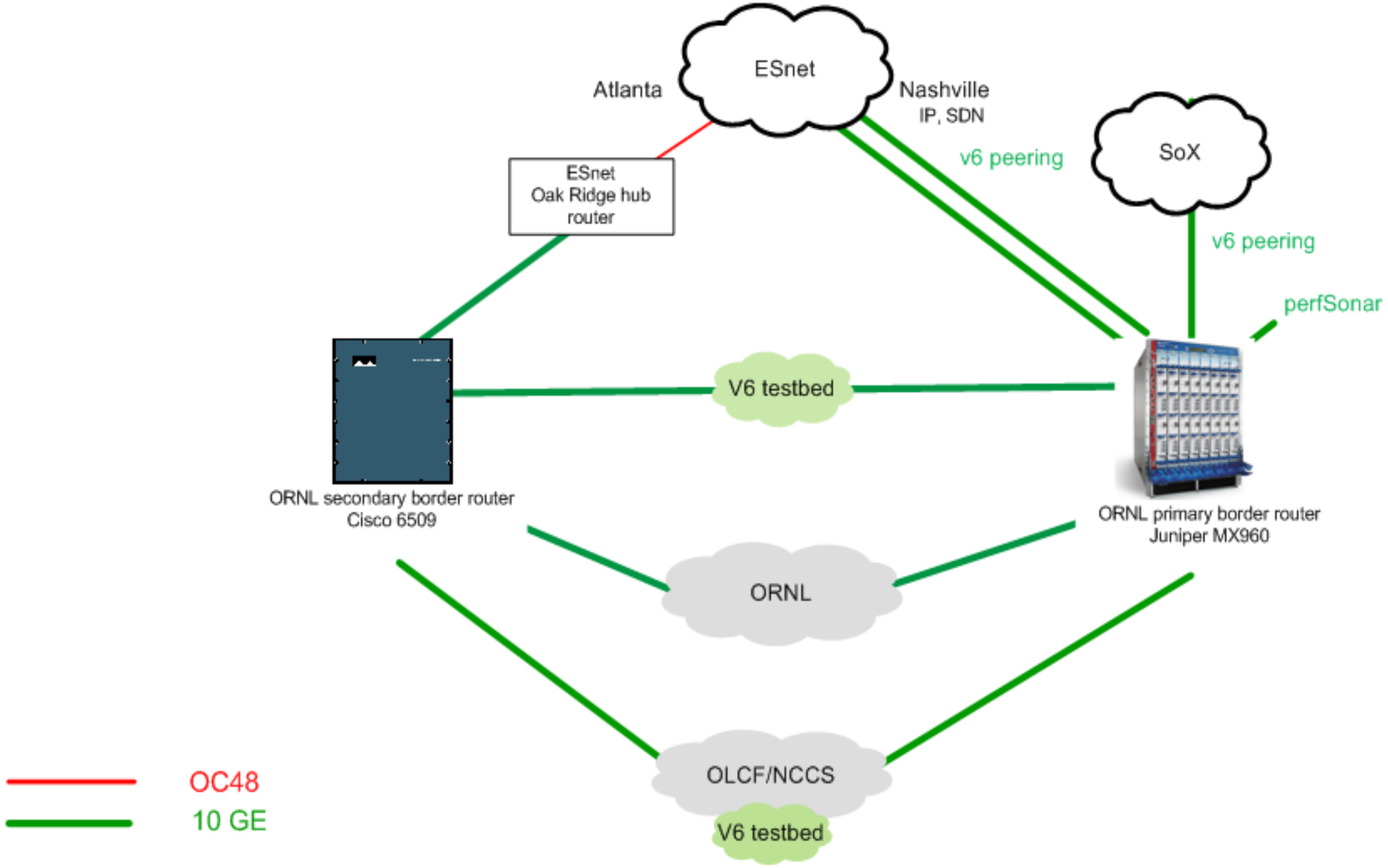
Survey results

Site discussions

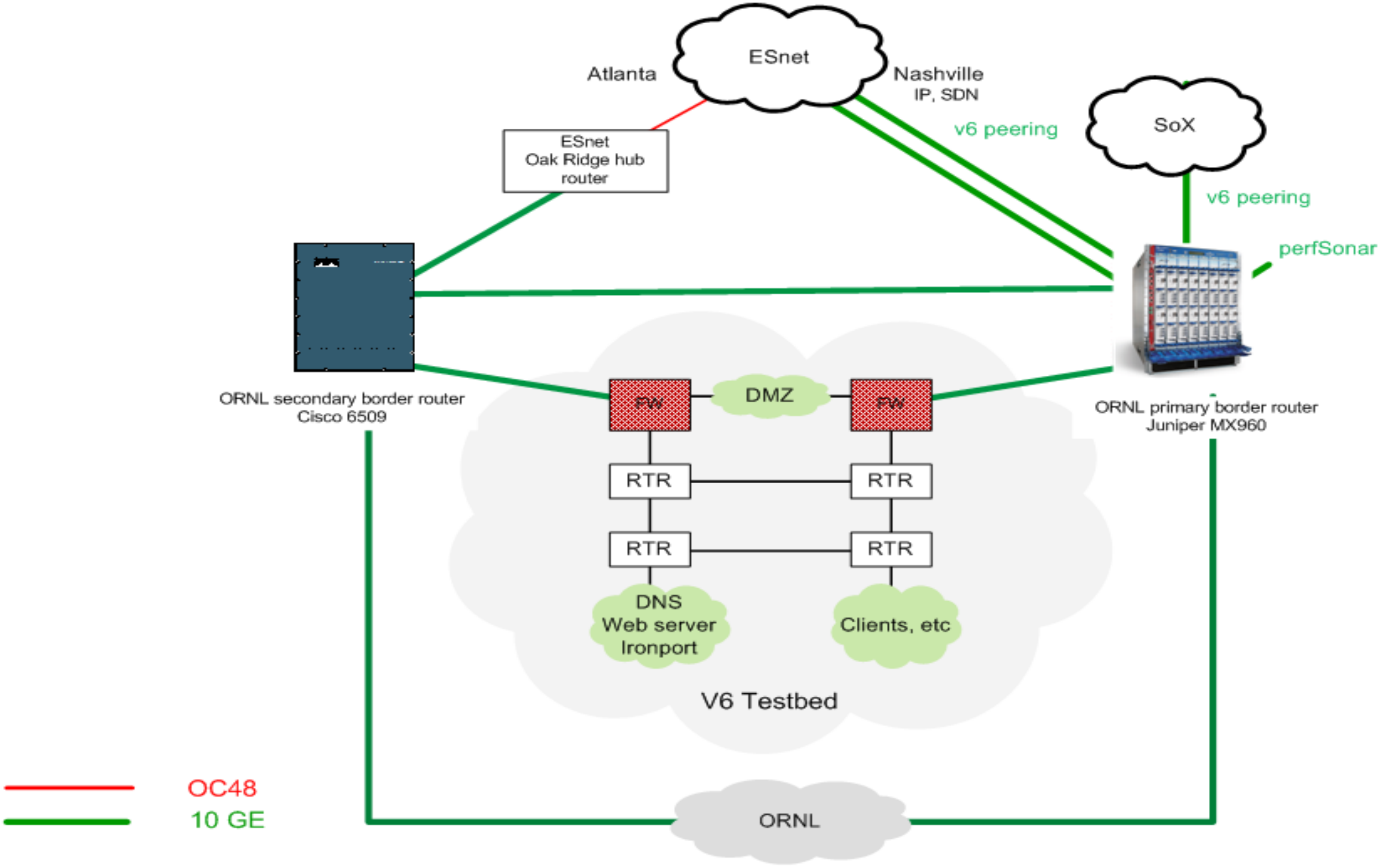
ORNL Near Term IPv6 Plans

- Currently peering with ESnet and SoX
- Just enabled v6 for border perfSonar
- Testbed
    - web server
    - DNS
    - email
    - firewalls
    - security tools
- Integrating IPv6 deployment with firewall upgrade
- Plan to implement v6 capable DMZ initially

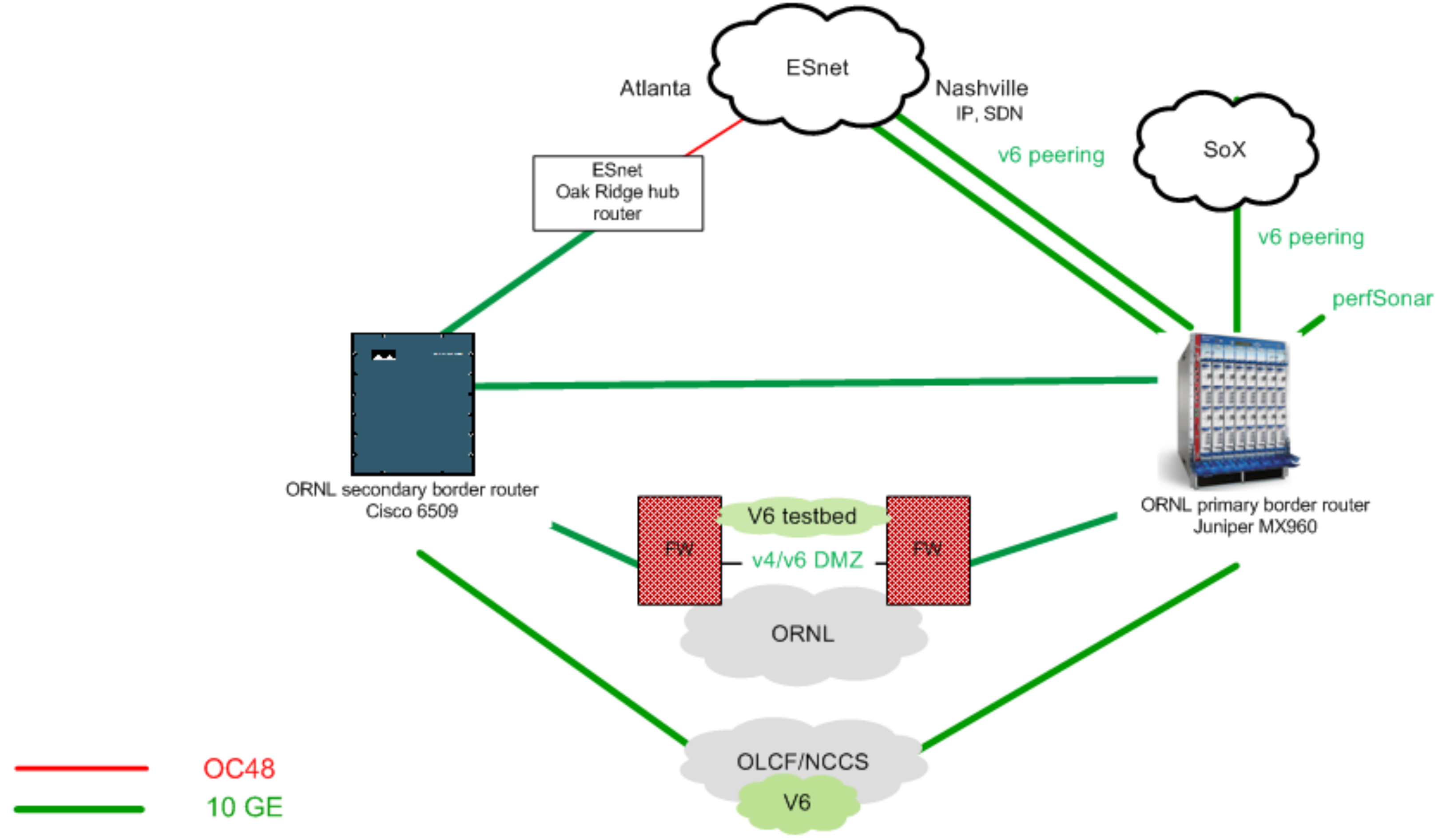
ORNL V6 Testing Overview



ORNL V6 Testbed



ORNL FY12 Target



FNAL IPv6 Status

Vyto Grigaliunas

Winter 2012 ESCC meeting

January 26, 2012

Office of Science
U.S. DEPARTMENT OF ENERGY

Fermilab

2012 IPv6 Milestones

DNS:

- IPv6 on InfoBlox vetted in test bed
- Expect production external DNS servers compliant by Mar 1

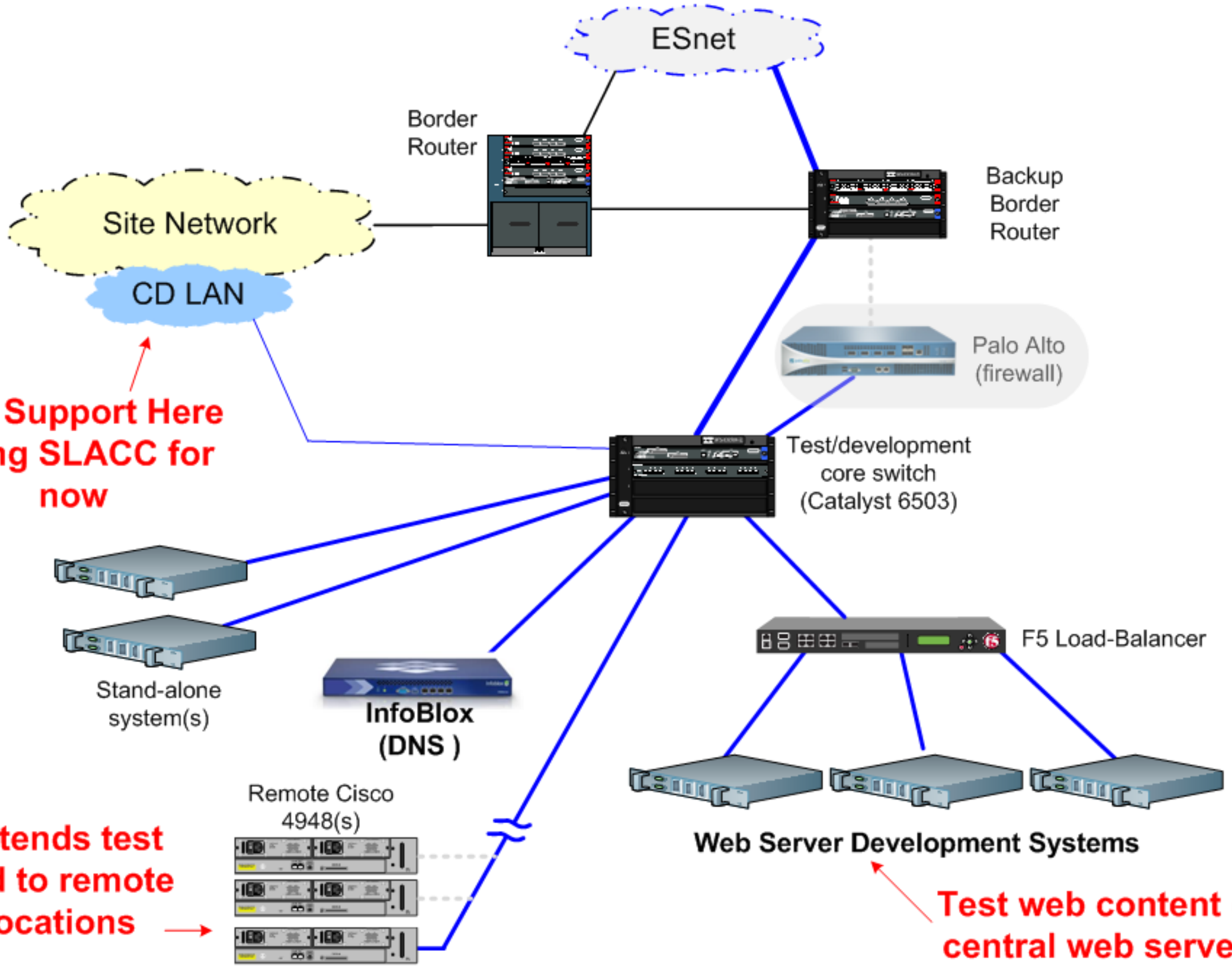
Public web pages

- Intend to migrate all central web server pages to IPv6
- Central web service development systems in test bed
  - Some resource issues with central web support effort
  - Target is production IPv6 support for fnal.gov by May 1

E-mail:

- External AV/Spam service from MessageLabs is a problem
  - No immediate schedule for MessageLabs IPv6 support
- Investigating other options to MessageLabs

IPv6 Test / Development Network Today



IPv6 Support Here Using SLACC for now

Extends test bed to remote locations

Test web content on central web servers

IPv6 Addressing Plan

We own two /48s; one PA and one PI

- PI block will be used in production network

Draft address plan completed; in review/approval process

General approach:

- 16-bit subnet field
- vLAN-based subnet allocation:
  - Believed to closely map to IPv4 subnets
- Will use readable decimal representation of vLAN number
  - Not hex equivalent of vLAN number
  - eg: vLAN 41 = IPv6 subnet [0000 0041], not IPv6 subnet [0000 0029]
- Allocates only ~6% of address space in /48

IPv6 Addressing Plan (II)

Reserving additional blocks of address space for location-based & service-based allocation schemes:

/126s to be used for pt-to-pt links

Misc:

- ULA’s not supported
- No decision yet on SLACC vs stateful auto-configuration
- Static address assignments will use decimal notation of last octet of IPv4 address (by default)
- Allocates only ~6% of address space in /48

IPv6 Support Problem Areas

OSPF authentication ‘untidy’

IPAM isn’t clean and will require additional work

Having some problems with rogue 6-to-4 “tunnels”

IPv6

- Current Test / PDN Equipment IPv6 Knowns
  - SSH, FTP, DNS, Squid (reverse cache)
  - F5 can provide 6 to 4 connectivity (more testing needed)
  - SMTP (Iron Ports If no IPv6 support)

IPv6

- Upgraded Address Assignment from /48 to /44
  - Interesting ??? Procurement
  - Legally belongs to BNL not BSA
- To Do
  - Update IPv6 Addressing Plan
  - Reconfigure peering with ES net
  - Reconfigure PDN test deployment / DNS
  - Give back /48