Turning the Network

NANOG 70

Andrew Gray
IP Engineer IV
Cox Communications (AS22773)

Who We Are

Cox was built up through a number of acquisitions, and each market was generally allowed to operate autonomously.

Why that's important

- Until approximately 2012, the general method of operation was for the central network engineering group to make general suggestions about technologies to use and how to deploy them, and allow engineers in the markets tailor custom solutions.
- There are over a dozen different markets in six different regions, all being tied together by a Cox backbone network.
- This gave a lot of flexibility, and a lot of good solutions came out of this work (quite a few of which were integrated into the final design)
- But it led to the inevitable chaos of network designs.

Taking stock of the situation

We started off by asking all the markets a few easy, straight-forward questions about what they were doing today.

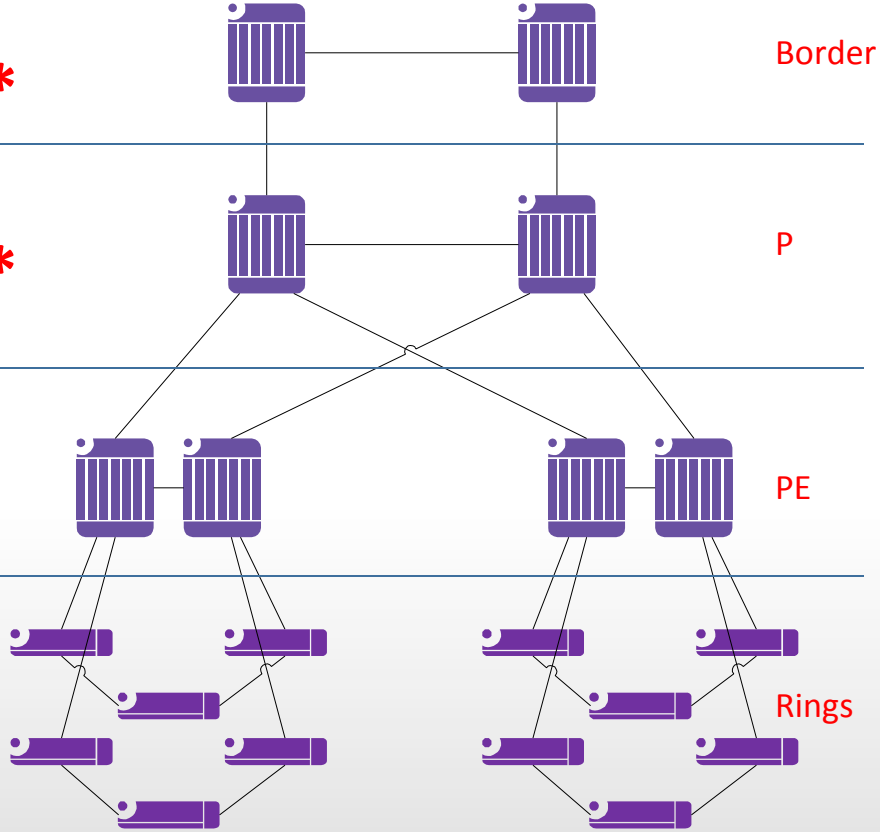
Terminology and (extremely) high level diagram

Inter-Market Border Routers (2)*

Market Core P Routers (2)*

Market Core PE Routers (2xN)*

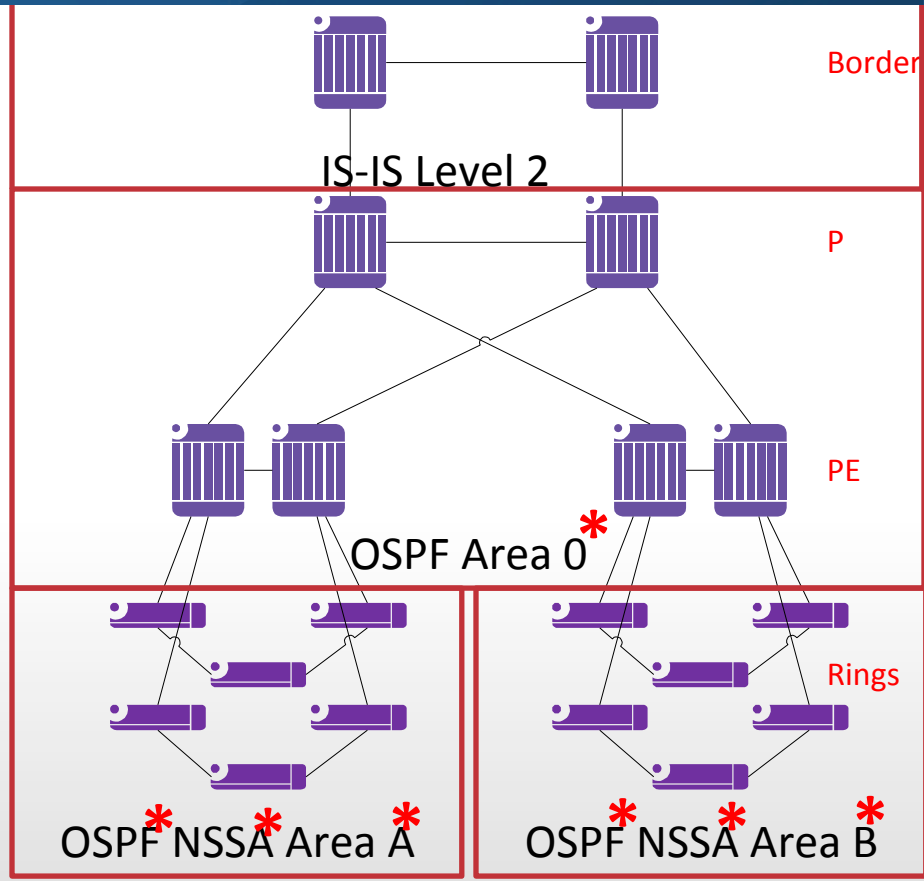
Customer Access Fiber Rings* (lots)



*** May not be accurate for a specific market**

Problem 1: Existing Network Topology

- Different area numbers
- Stub areas, NSSA areas, regular areas
- A couple places static routing instead of OSPF
- I lost a bet – we didn't find RIP anywhere



*** May not be accurate for a specific market**

Problem 2: Management Connectivity

- Almost every market had a different method of gaining management access to their network:
  - Some via pseudo-Internet connections
  - Some via private VPLS/VPRN connections
  - Some firewalled, some ACLed, some Something Else
  - Some MTU 1500, some MTU 2048, some MTU 4000, some MTU 9000, some MTU 9100, some MTU 9212
  - Some interfaces LAG'ed, some not
  - Some interfaces null-encap, some dot1q, some qinq

Problem 2: ~~Management Connectivity~~ Almost Everything

- Management Connectivity
- NTP configuration
- Syslog configuration
- Security posture
- Naming standards for everything (hostnames, interface names, interface descriptions, etc. etc. etc. etc. etc. etc.)
- Interface configuration (encapsulation type, etc.)

Problem 3: IPv6 Support

- Cox Business has supported IPv6 customer internet connections for years, however our internal management and telemetry networks were not IPv6 aware.
  - Not really a problem today due to private IPv4 space
    - …although we've run out of that, and now have multiple, overlapping RFC 1918 blocks
  - Some management tasks not IPv6 ready either
    - That list is shrinking, however – things like syslog, SSH, authentication are ready. NTP is lagging behind on some platforms. Our management platform is capable, but was not enabled.

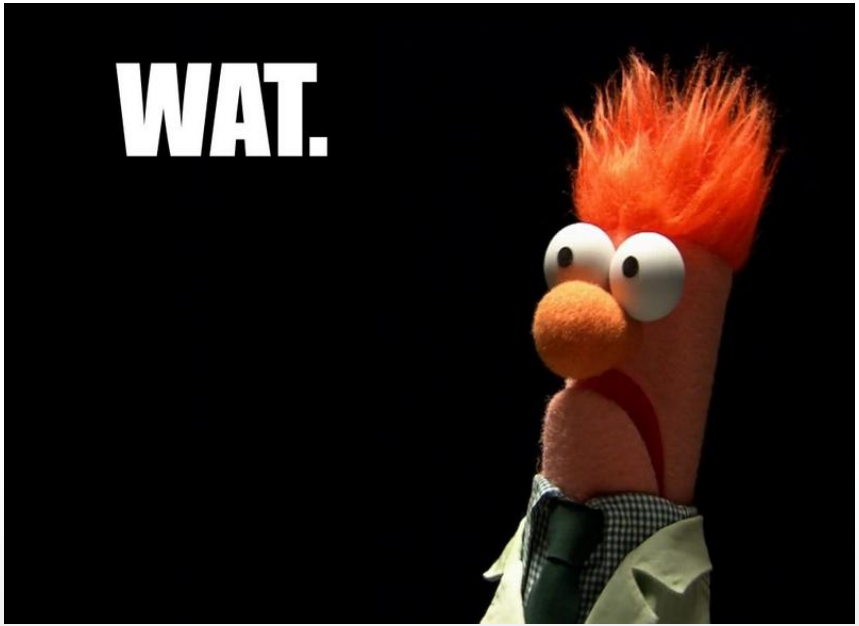
Root problem

Having so many variables was making it extremely difficult to engineer new technologies and solutions:

- BGP labeled unicast
- SDN
- IPv6 management
- Improved security posture
- IP unnumbered on rings
- Consistent failover times
- Consistent QoS policies
- Automated network configuration auditing
- New platforms

Root problem

- Testing new software loads and hardware platforms for all the configurations in use in the markets.
- Inter-op testing between all the different configurations.



Getting everyone together

OSPF to IS-IS migration

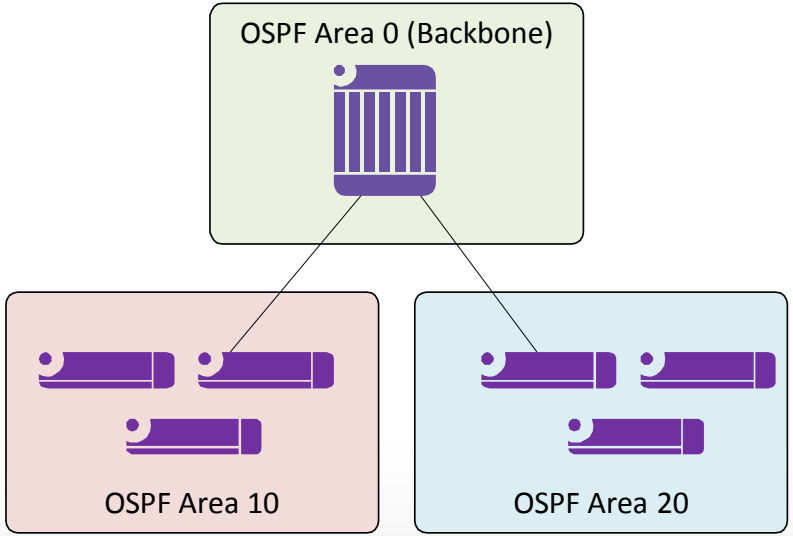
- One of the main project goals was to unify to a single IGP.
  - OSPFv2 does not support IPv6
  - Our existing OSPFv2 deployment was not conducive to unifying.
  - OSPFv3 was considered as an option.
  - IS-IS was already in use in the network core, and across other segments of Cox.

Issues with IS-IS

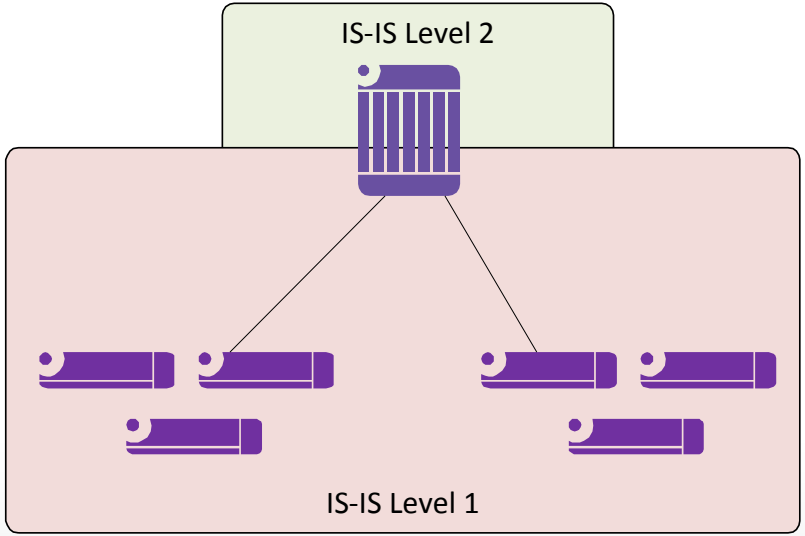
IS-IS is not a perfect fit:

- We needed some mechanism to split up the routing domain more effectively than just level 1/level 2 separation.
- Officially tested IS-IS limits were lower in some cases than their matching OSPF limits.
- We discovered we had been violating the officially tested limits of OSPF (some times by many multiples).

OSPF Areas vs IS-IS Areas/Levels



Each area sees only their own area's routes, plus summaries from backbone. Nice, small routing tables.



Everything inside IS-IS level 1 sees every other level 1 device, plus a level 2 summary.

IS-IS Areas are not OSPF Areas!

- We can’t use multiple area numbers in this instance, as we have the same router terminating the different areas.
- In this case, the router acts as all of the areas at the same time – level 1 area A is the same as level 1 area B, and are put together.
- We have some devices with very small route table capacity (especially with IPv6).

Possible solution: Multiple VRFs

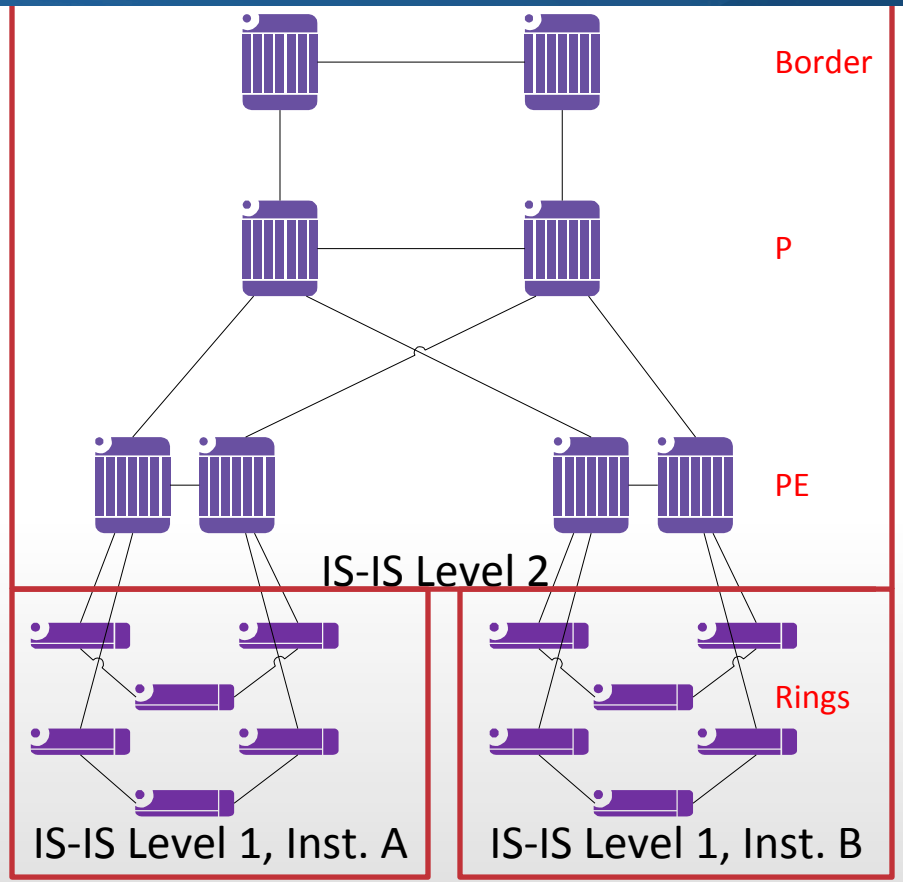
- Multiple VRFs were considered for about two emails.
- There would be a fairly high administrative overhead to manage them.
- Our then version of code didn’t doesn’t support running MPLS out an interface in a VRF anyway.

Solution: Multi-Instance IS-IS

- A type of MI-IS-IS is defined in RFC 6822, but according to that specification, multiple topologies inside multiple instances are not permitted.
- Our vendor platform does not have that restriction (fortunately) – others probably don't either.
- We standardized on 32 instance identifiers:

Instance	Use
0	Core instances
1-15	Ring instances
30	Rings with "unique" circumstances
31	Labs

Solved: Nice, consistent IGP

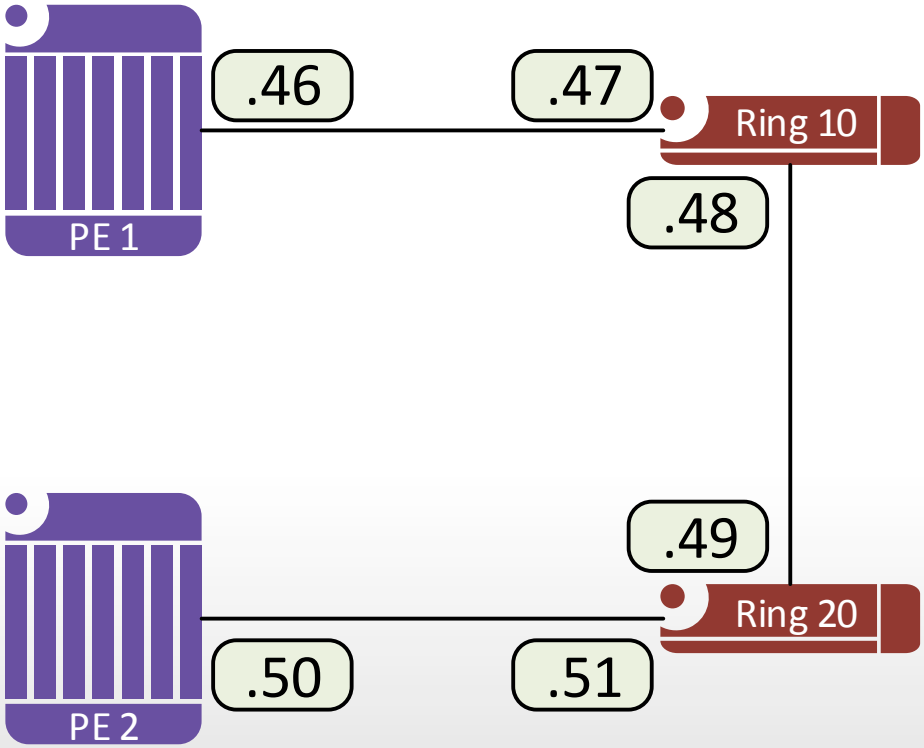


Bonus Points

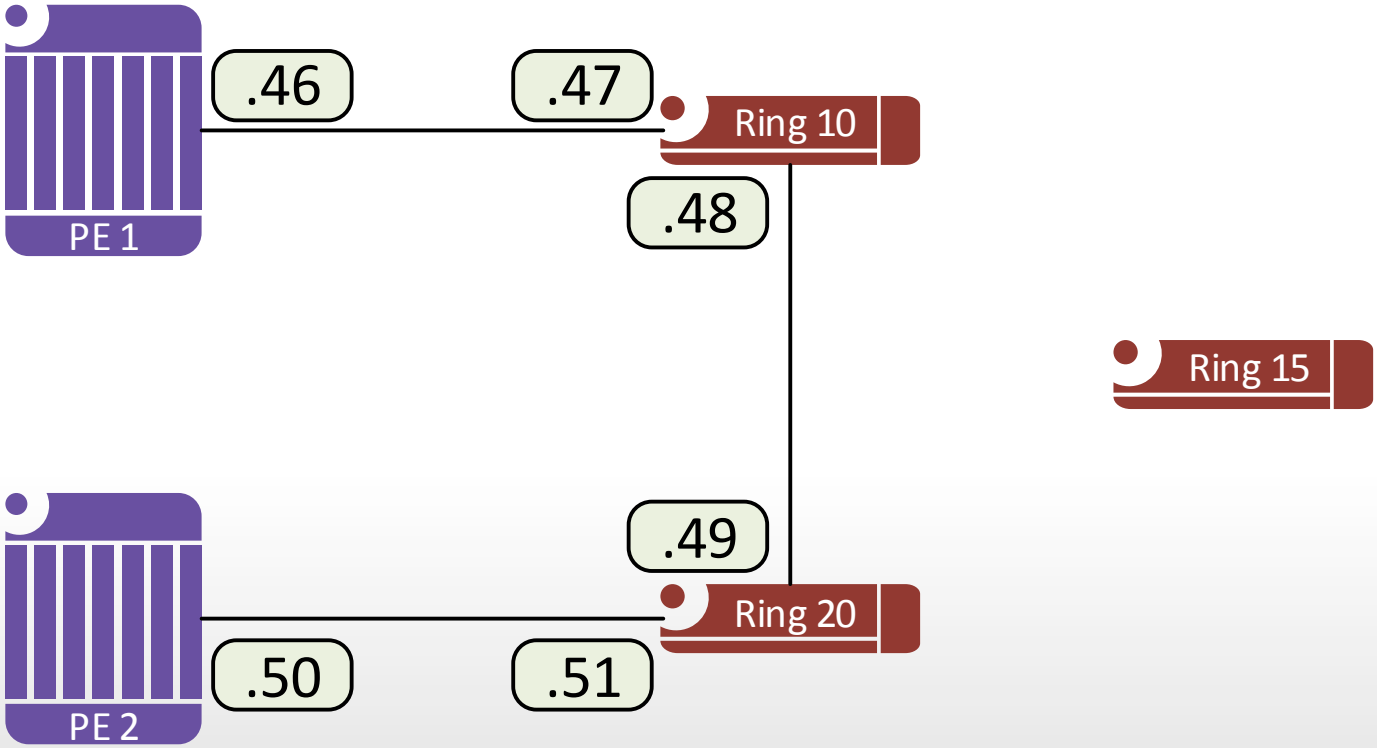
IP Unnumbered – What is it?

- Links no longer get dedicated IP addresses, they instead re-use another address (the loopback, in our case)
- IP Unnumbered can help solve our IP addressing challenges, plus give us benefit in ring maintenance.
- Most of our hardware platforms already supported it, we got our vendor to implement it on the remaining platforms with a code released in March, 2017.

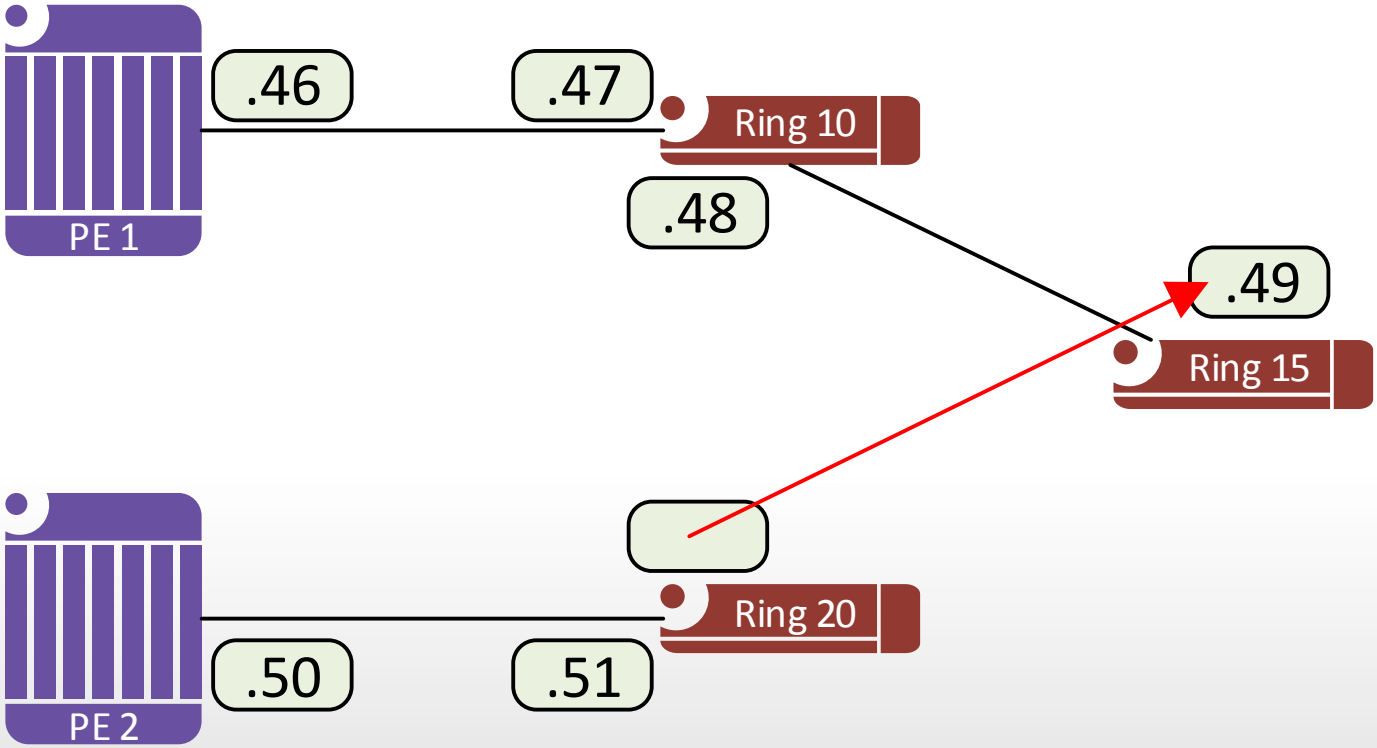
The Benefit to IP Unnumbered



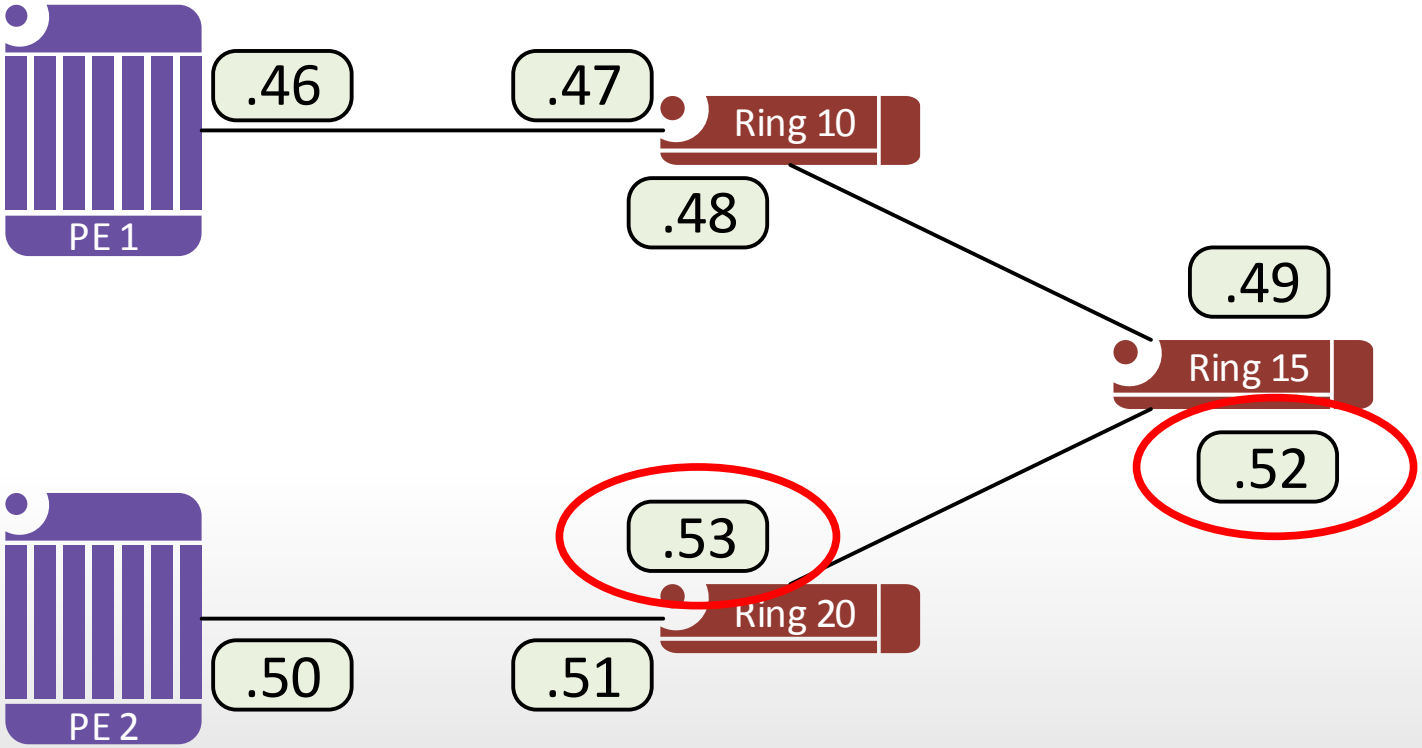
The Benefit to IP Unnumbered



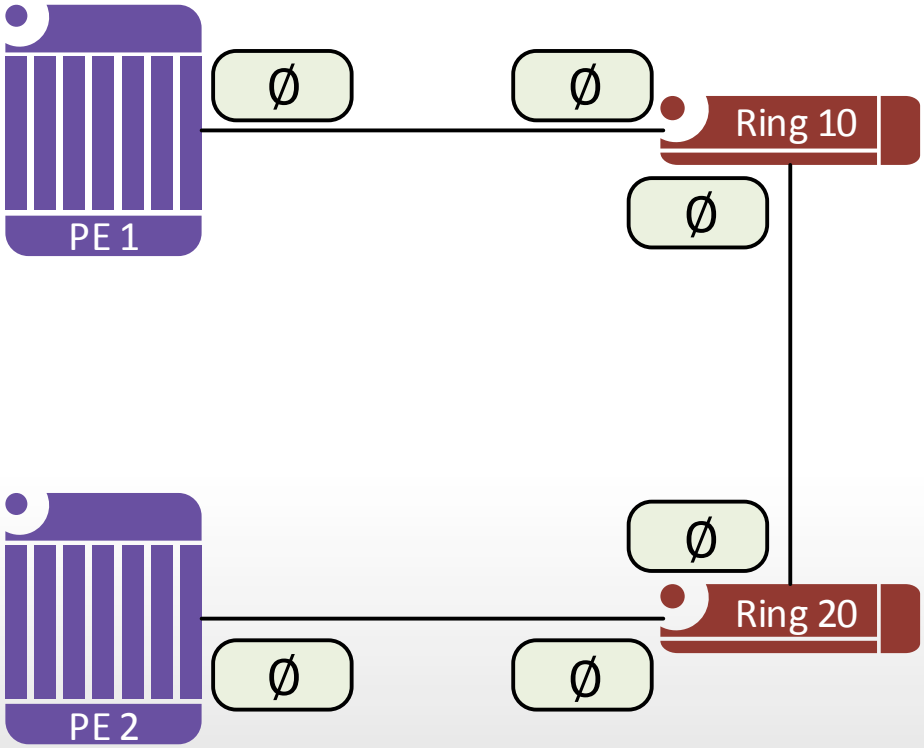
The Benefit to IP Unnumbered



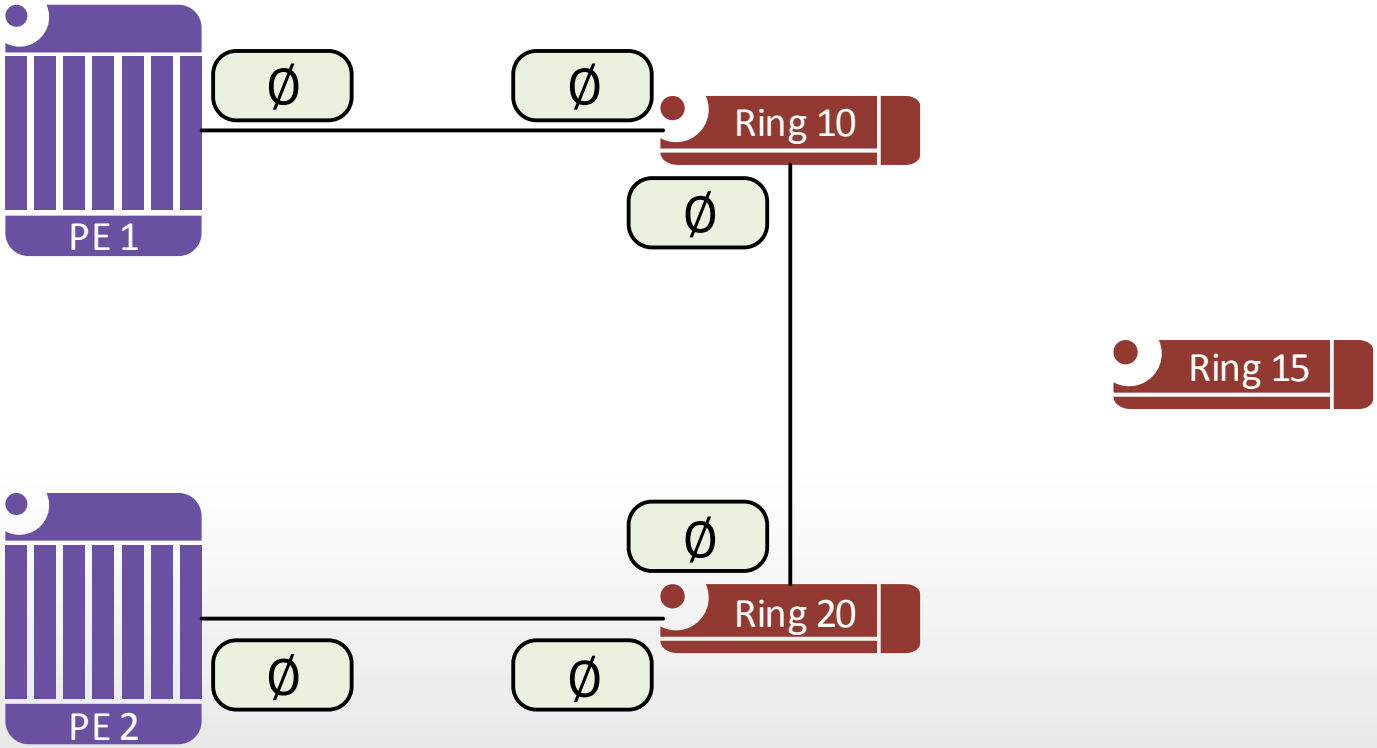
The Benefit to IP Unnumbered



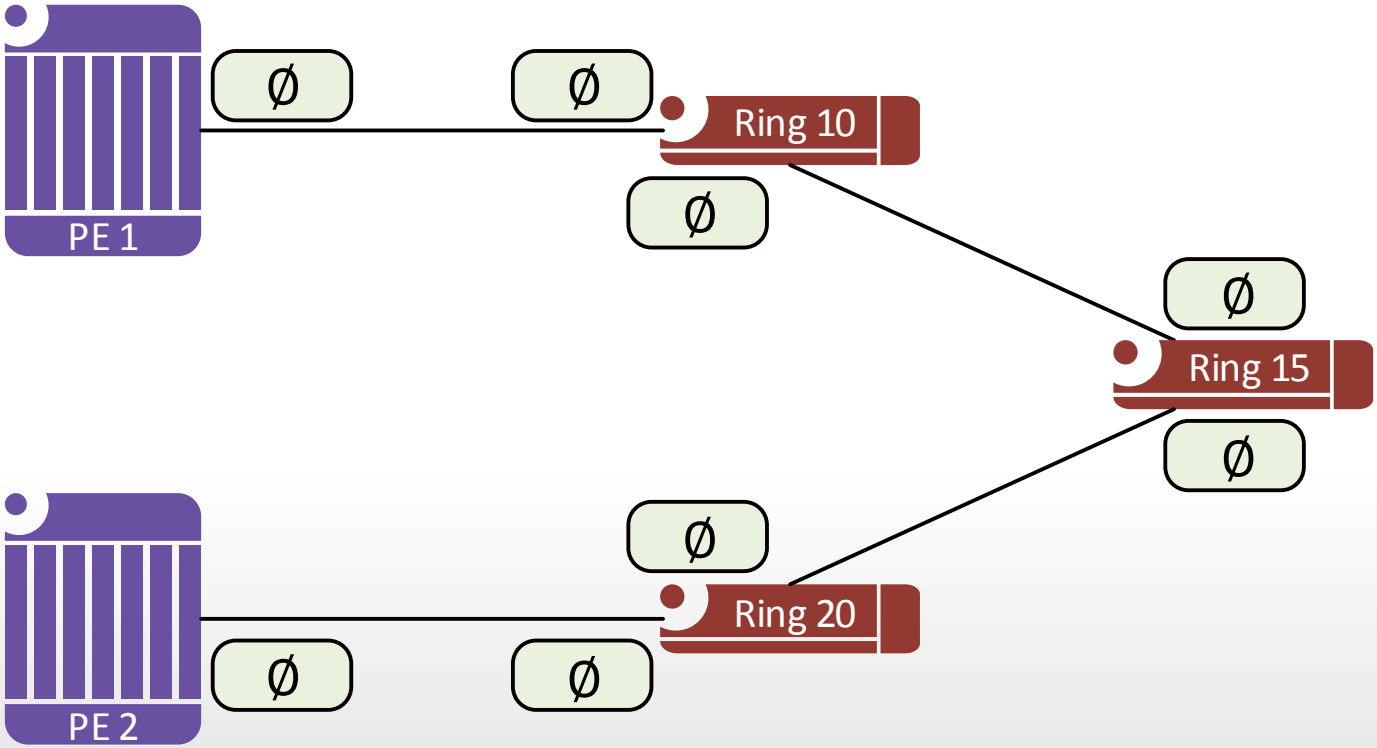
The Benefit to IP Unnumbered



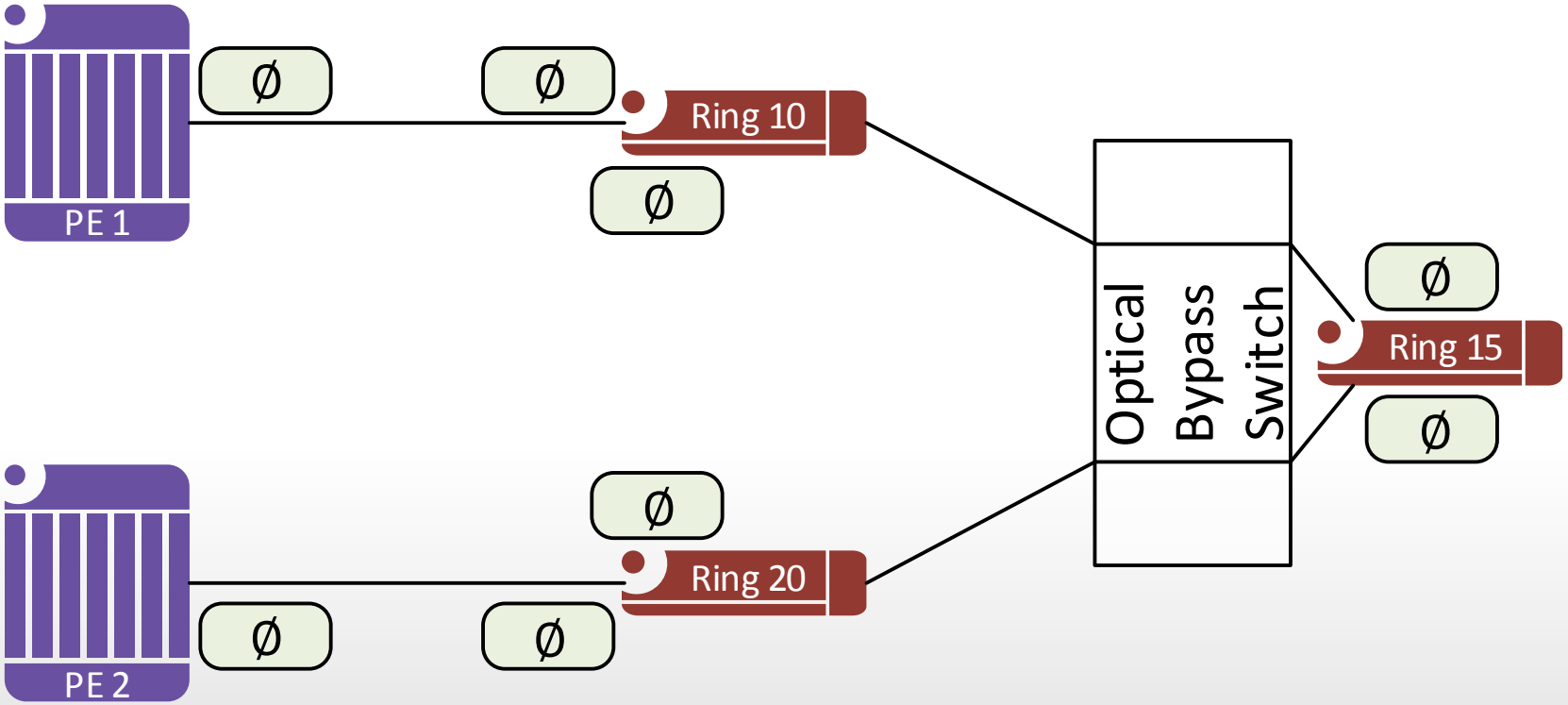
The Benefit to IP Unnumbered



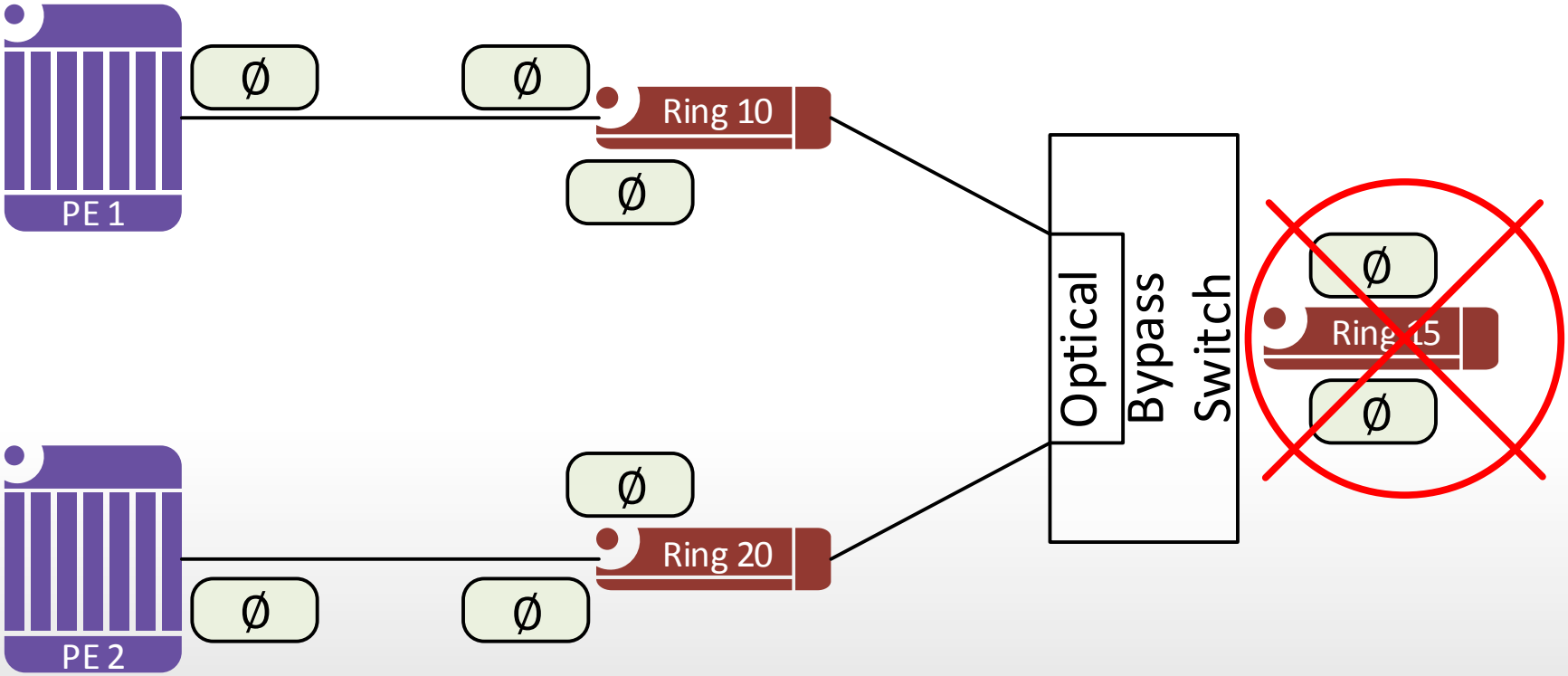
The Benefit to IP Unnumbered



The Benefit to IP Unnumbered



The Benefit to IP Unnumbered



A few notes if you look to deploy IP Unnumbered

- BFD is not supported.
  - EFM-OAM (802.3ah) is a good substitute though.
- RSVP FRR has minor issues dealing with signaling bypass tunnels.
  - Get around this by utilizing a second, dummy, only-ever-seen in RSVP address to borrow from.
  - We just took the system IP and made the first octet a 1.
  - Don’t worry APNIC, these addresses are never put into IGPs, advertised, etc.
- Be careful of potential link distances
  - Your standard 10km optic may not be the best idea any more…
  - Do you assume 1 failed node? 2? 3?

Why not just use IPv6 and link-local addressing?

- Lack of full suite of protocols supporting IPv6 – the key one for us being RSVP.
- Transitioning to Segment Routing was an option, but an evaluation at the time showed that there would be not much gain for us over the existing RSVP-TE based network, it would be a lot of work, and SR was felt to be too immature.

IP Unnumbered – The Benefits

- Conserves IP addresses
  - But that turns out to be a fairly minor benefit
- Reduces the amount of pre-provisioning we have to do on a box.
  - A box can rattle around in a field person's van for a month, and be deployed on a customer site quickly.
  - Once the ring connectivity is up, and the loopback address set, the rest of the box can be set up from remote, and our field person can go on to the next job.
- Disaster recovery
  - When hurricanes or the like take out substantial number of nodes, having this self-healing capability is useful.

IP Unnumbered – Next steps?

- We are experimenting with having the devices pull their loopback addresses via DHCP (ideally DHCPv6), to further reduce the amount of configuration the person on site needs to do.
  - The device's pre-loaded configuration would talk to our management system, and support personnel could find the IP address from the device's serial number there.
- We believe with this we can get to a zero on-site configuration needed (although we still need to pre-stage the device with a baseline configuration).

Timeline

- This project is still in progress (but getting very close!)
  - Initial internal planning discussions in January, 2015.
  - First preliminary HLD in June, 2015.
    - Roughly 40% of the design here would change before the final version.
  - June through November – MANY meetings, bringing in additional market engineers for input and sets of eyes, and many revisions.
  - December, 2015 – “98%” finished HLD… we thought.

Timeline

- January, 2016 – Re-work to bring IP unnumbered rings support in.
- January – July – LOTS of testing, banging on things in the lab, learning fun new ways to crash a network (more routing loops than we'd care to admit to…)
- May, 2016 – First phase, Area 0 and Management standardization, work started.
- July, 2016 – "Final" HLD (version 27!), transition procedures written.
- September, 2016 – First market completes IS-IS transition.
- October, 2016 – Last market completes IS-IS transition.

Timeline

- April – May, 2017 – IP Unnumbered rings tested and validated in lab.
- July, 2017 (est) – Last components needed for IP Unnumbered deployed, expecting first rings to come up.
- September, 2017 (est) – All new rings configured using IP Unnumbered. Off-platform services start being signaled via BGP-LU.
- 2019 (est) – Enough rings operating in unnumbered mode that we convert all remaining legacy rings to it.

Additional Lessons Learned

Lessons Learned

Problem: A lot of projects like this falter or fail because of trying to get everything done at once and perfect.

Solution: Breaking the project up into multiple phases helped immensely, including phases defined as “in the future… sometime.”

Lessons Learned

Problem: For projects like this, *everyone* has a few skeletons hiding in the corners of their network design, that will come out to bite you.

Solution: We kept all the engineers involved over many months, did multiple reviews, and provided targeted training for market engineers who were the local experts in their area. This training served two roles – to both familiarize the markets with the work to be done, and as a last check against market-specific skeletons to ensure adequate time can be spent to evaluate them.

Lessons Learned

Problem: The way your “standard” network in your test lab is configured doesn’t match the “standard” in production.

This one bit us a few times in different ways – our lab has cards and platforms that we had evaluated and decided not to deploy that had strange issues. Our lab deployment of our management platform fully supports jumbo frames – production has issues doing so.

Solution: We focused our FOA (first office application) in a market where we had two engineers that had been on this project nearly since the beginning to help smooth over issues that were discovered night-of.

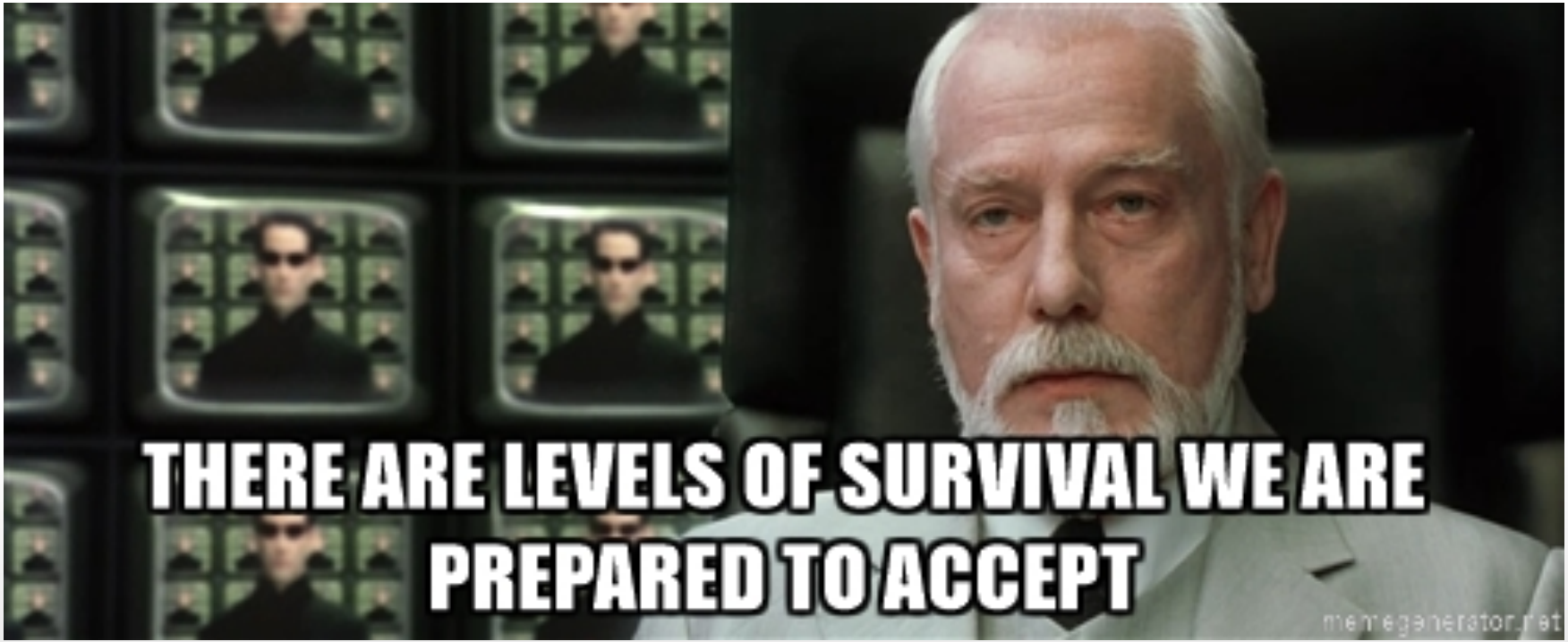
Lessons Learned

- We were not nearly as secure as we thought we were.
- Different policies applied in different places with different permission sets.

Take Away: Get as many people thinking about this as you can! Not jut your own engineering groups, either.

- Ask your vendor
- Ask your operations folks
- Ask your people doing the hands on work

Security


THERE ARE LEVELS OF SURVIVAL WE ARE PREPARED TO ACCEPT
memegenerator.net

Overall Success

- We were highly successful in implementing these changes to date.
  - 8 training sessions for 102 engineers performed
  - 580+ core nodes converted from a variety of configurations to a new standardized IGP+BGP topology, with standardized QoS, security policies, and other large chunks of configuration.
  - 57 maintenance windows required for the core changes.
  - 1 customer-impacting event, lasting only a few minutes (root cause was yet another skeleton in the networking closet lying in conflicting QoS policies)

Last Comment

Yes, THAT AS22773:

```
Global Per AS prefix count summary
----------------------------------

   ASN    No of nets  /20 equiv  MaxAgg  Description
  4538          5569       4190     74  ERX-CERNET-BKB China Education and Rese
  7545          3774        391    265  TPG-INTERNET-AP TPG Telecom Limited, AU
 22773          3692       2969    152  ASN-CXA-ALL-CCI-22773-RDC - Cox Communi
 10620          3541        546    149  Telmex Colombia S.A., CO
  6327          3377       1333     83  SHAW - Shaw Communications Inc., CA
 39891          3334        171     18  ALJAWWALSTC-AS, SA
  8551          3250        377     41  BEZEQ-INTERNATIONAL-AS Bezeqint Interne
 17974          2991        903     72  TELKOMNET-AS2-AP PT Telekomunikasi Indo
 20940          2968       1071   2109  AKAMAI-ASN1, US
  4766          2760      11150    755  KIXS-AS-KR Korea Telecom, KR
```

We know. We're working on it.

Q & A

Questions? Comments?

Feel free to contact me:

Andrew.Gray@cox.com