Multi-layer Traffic Engineering

Beatty Lane-Davis
blanedavis@infinera.com

infinera®
what THE NETWORK will be

Agenda

- Traffic Engineering – historical context
- MPLS-TE – the old stuff
- BGP-LS, PCEP & Segment Routing – the new stuff
- Multilayer TE
- Macro Trends

infinera®

Traffic Engineering

The historical context that led us to MPLS-TE

Traffic Engineering – A historical context

Nothing new here…

- How many resource units are required to handle this many demands placed on a system with a given level of service.
  - Erlang's original work started with how many operators were required to physically patch a given load of calls.
- What is the least amount of capacity I can get away with and keep my customers happy & how can I utilize every bit of it?
- From TDM to the latest flavours of stat muxing TE hasn't gone away as it represents a fundamental tenant of our jobs.

RIP Agner Erlang
1878-1929

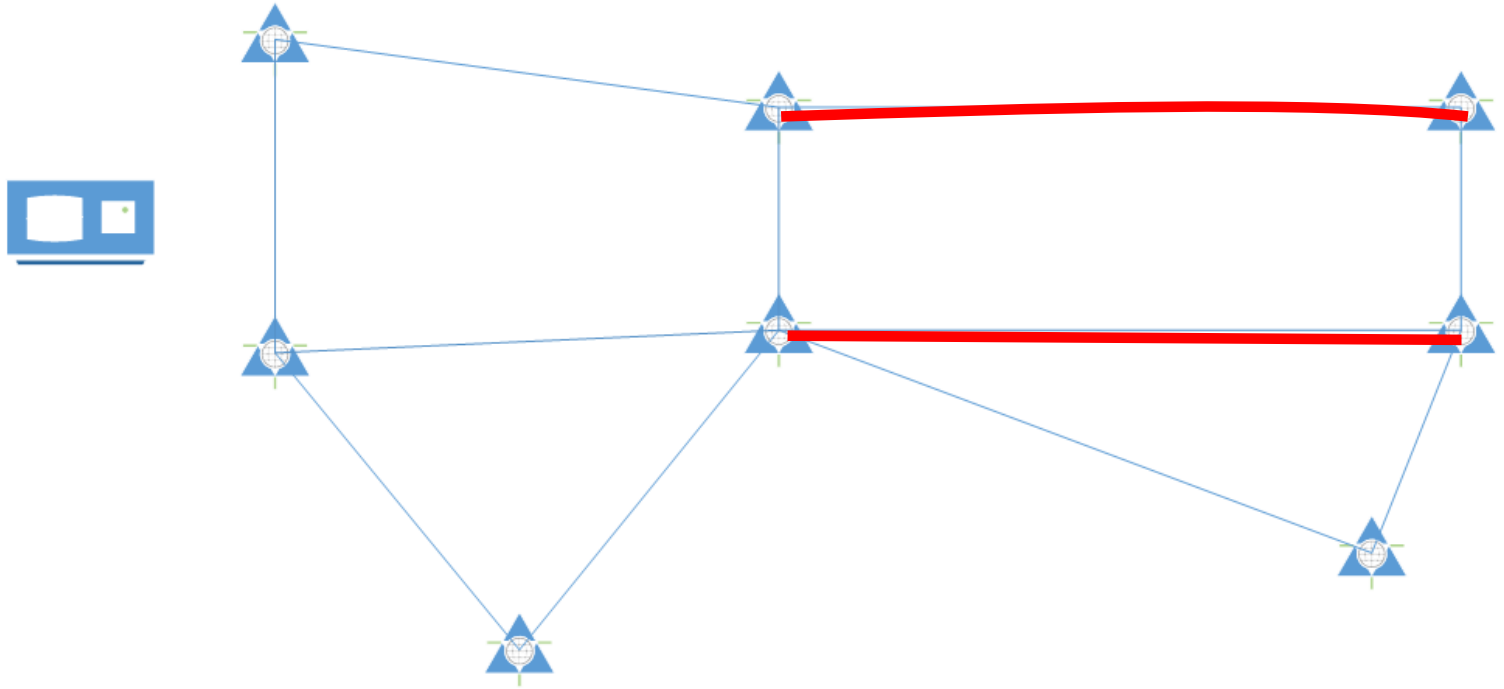
infinera®

What does TE do for us?

What does TE do for us?

What does TE do for us?

TE is hard with IP



In particular, it is generally desirable to ensure that subsets of network resources do not become over utilized and congested while other subsets along alternate feasible paths remain underutilized. – RFC 2702

infinera®

Traffic Engineering – A historical context

- TE is pretty well stuffed within the connectionless construct.
- Traditional IP traffic engineering was hard enough to be considered unworkable at even moderate scale.
  - Tweaking metrics more often than not resulted in pain relocation rather than eradication.
  - There are big nerds who would argue this point.
- Enter MPLS-TE at the latter end of the 20th century.

infinera®

MPLS-TE

Magic Problem Solving Labour-substitute – Traffic Engineering

The old stuff…

- MPLS was a visceral response to the IP-world's growing disdain for ATM.
  - Expensive boxes
  - 35%+ Cell tax
  - Net vs Bell
  - 10G SAR limitation
  - Insert your other favourite reasons to hate on ATM here…
- MPLS-TE brought a connection-oriented model to IP – a scalable approach for source-routing.
  - Chuck smart software at one box so you can yank a layer out of your POP & consolidate your infrastructure.

infinera®

The distributed problem…

- But MPLS-TE didn't solve world hunger – TE is still hard.
  - The extensions to the IGP's & RSVP gave us more information in the database & policy knobs to solve problems - but we weren't done.
- The distributed problem
  - The router driven approach - the distributed nature of the decision making led to inherently non-deterministic solutions.
  - Tunnel placement is entirely driven by who went first and when the last reboot was.
    - Inefficient stacking, stranded resources, etc.
    - Need to keep touching boxes as bandwidth requirements change.

infinera®

Distributed problem continued…

- More recently vendors have implemented ‘auto-bw’ to try to make distributed more effective.
  - Automated way to resize tunnels based on actual tunnel utilization.
  - Makes the network more dynamic to actual loads – without requiring expensive tools & integration.
- Solves some problems and like any good feature creates some new ones as well:
  - Can improve efficiency some of the time.
  - Can cause pathological meltdowns – really broke, like needs people to come kick things to recover from.
  - Boxes working harder & network in a constant state of flux.

infinera®

The centralized problems…

- Networks tended to fall into two camps:
  - Have's (huge capacity issues & the money to pay for fancy software)
  - Have Not's (the large bulk by #'s)

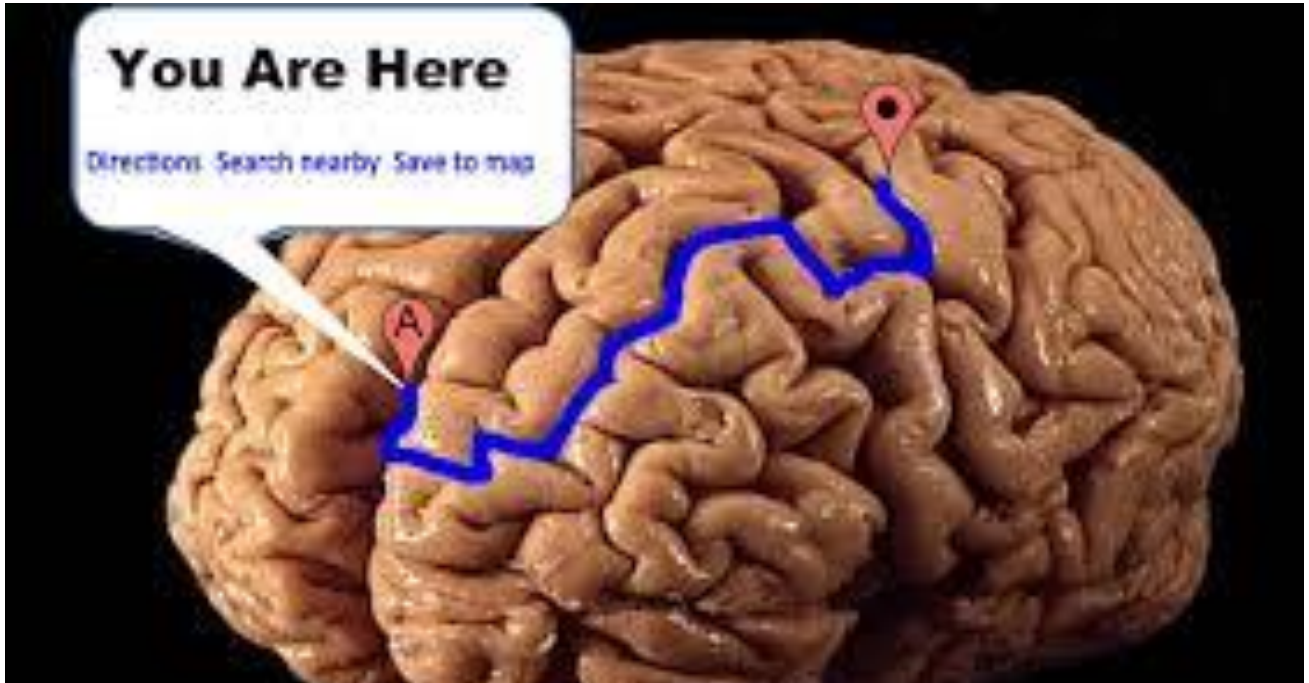
- The 'Have Not's' tended to run MPLS-TE to get FRR & do a bit of fate-sharing avoidance and tactical avoidance of the occasional crunch point.
  - But from a capacity perspective, many run largely overprovisioned.

infinera®

Centralized TE

- Traffic engineering is unquestionably more effective when a global view is taken to holistically map demands onto available resources.
  - Deterministic, optimal use of resources, planning for failure modes, etc.

- One big one is better than a bunch of little ones…



Feeding & Watering a centralized tool…

- Getting the topology graph together was a likely-familiar transform problem:
  - Export topology info from this system, transform into this format, email this guy over here, push into tool – keep current going forward.
- Getting the traffic demand matrix wasn't easy:
  - Flow stats, counters & approximation.
  - More code to map nexthop information against the graph information to determine exits.
  - Extrapolation?
- Numbers get crunched – TE tool passes ERO's to a provisioning system – provisioning system pushes config.
  - Rinse, repeat based on frequency which will provide optimal benefit.

infinera®

It got the job done, but stayed fairly niche…

- TE is a super math-heavy problem space, so the code is complex.
  - Statistics, regressions, MANY other mathy things I'm allergic to.
  - Tomogravity, min-max fairness, other unpleasant-reading documents.
- The smart tools of yesteryear were working with the boxes & tools of the vintage. Integration was labour-intensive.
- While the whole system was a bit dynamic, but not very & not easily or cheaply.

infinera®

BGP-LS, PCE-P & Segment Routing

infinera®
what THE NETWORK will be

The New Stuff

- 15 years later, we're finally getting around to realizing the benefits of a truly automated traffic-engineered network:
  - BGP-LS – takes the stuff from your TE-enabled IGP, sticks it in BGP & feeds it northbound – that's all…
  - PCE-P – provides a programmatic means of letting a centralized controller tell the boxes where their traffic should go.
  - Segment Routing – a newish, more scalable approach to source routing that goes hand in hand with centralized control like chocolate n peanut butta.

infinera®

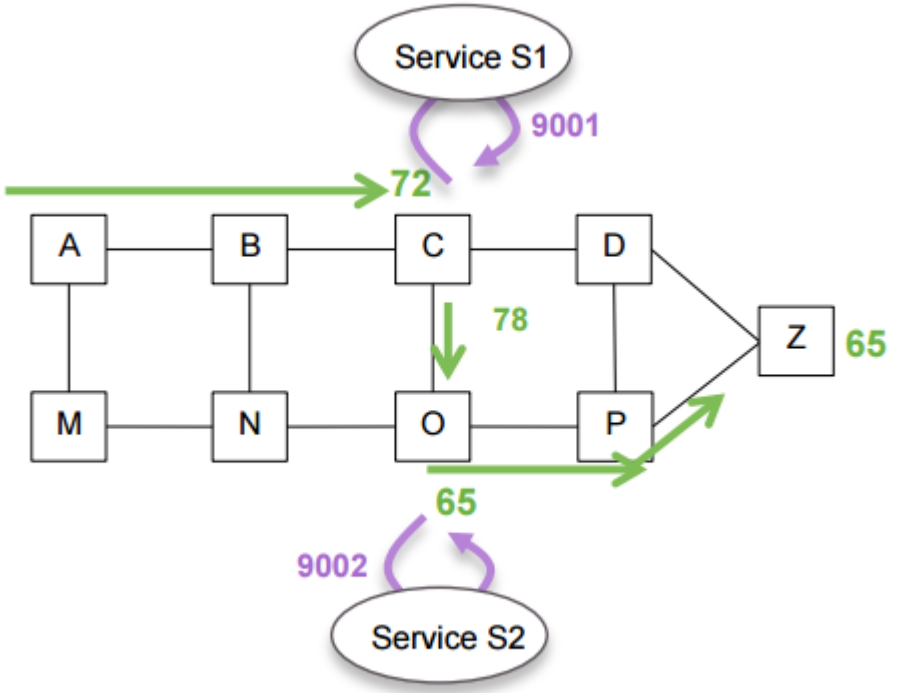
The New Stuff

- BGP-LS – really, it's taking the IGP TE info and feeds it northbound using BGP – that's all
  - Automated graph discovery.
- PCE-P – Allows centralized smarts without radical architectural changes – two options:
  - Stateless – boxes ask for a path calculation, signal based on answer. Ask again at re-optimize time.
  - Stateful – boxes delegate control of tunnel placement to central PCE which can tweak at will.
  - Both approaches provide the benefits of centralization – without needing to interact with a provisioning system, touch a config file or send the box into a commit crunching exercise.

infinera®

The New Stuff

- Segment Routing



- Think MPLS, without the signalling protocols.
- IGP's advertise 'segments' in addition to addresses.
  - Forwarding based on either MPLS labels or IPv6 addresses.
- Push source-route forwarding semantics to head-end box who pushes routing information for a given packet ONTO the front of said packet.
- But… doesn't that mean?
  - No need to push 25 labels onto each packet at each head end.
  - No end-to-end state in the middle – only re-route info.

infinera®

Segment Routing – what can't it do?

- Traffic matrix generation
  - Boxes can stream per-segment counters to a traffic matrix collector.
  - No longer do you need to marry up flow information with nexthops and correlate against interface counters:
    - The boxes count route recursion to let you know how much traffic is heading to a given 'exit segment' in realish time.
- Let the IGP do it's thing, push 'redirection' info to nodes where you need something extra done.
  - As with affinities, you really only need to give the box a few hints (labels, segments) to force things down the path you've selected
  - Oh and you can use affinities too…

infinera®

The New Stuff - Summary

- The ‘New Stuff’ gives us a nice balance between centralized and distributed.
  - The extreme edges of the pendulum lead to difficult problems which may not be nicely solveable:
    - Distributed TE
    - Completely centralized control

- The right balance allows us to leverage the best of both and focus on things that we can either save or make money with.

infinera®

Multi-layer TE

Multi-layer TE

- Global perspective of layer 0, 1, 2 & 3.
  - Optimizing path of layer 2 & 3 services over existing transport.
  - Reoptimizing transport based on layer 2 & 3 demands.
  - Physical fate sharing – SRLG's no longer require emailing spreadsheets around & manual config.
  - Latency optimization.
  - Multi-layer mesh protection – 50ms repair provided by upper layers, bandwidth presented to upper layers constrained to n failures.
    - Dynamic optical layer creates new paths for restoration bandwidth (slower.)

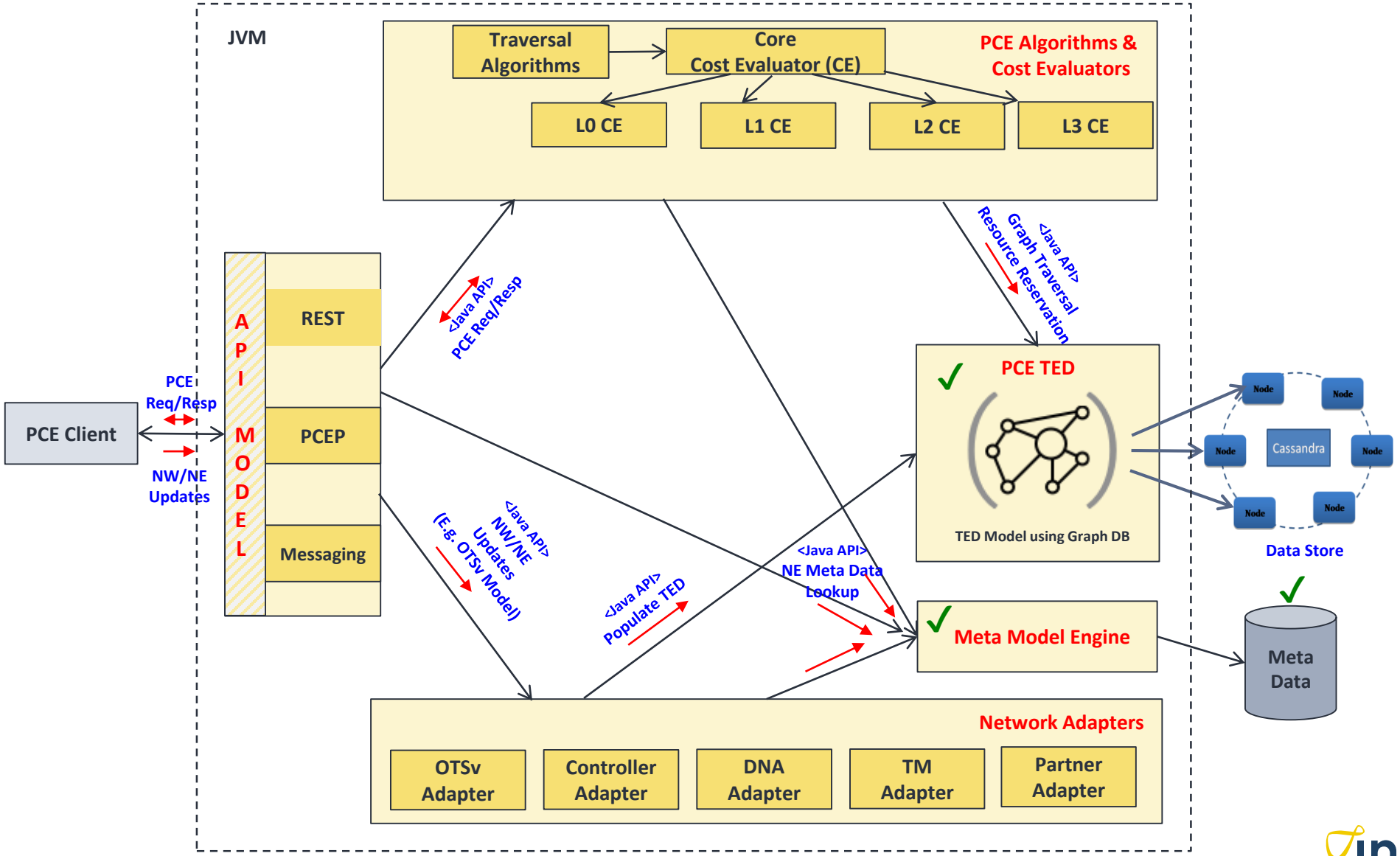
infinera®

Multi-layer TE

- Inject real-time SLA information via active monitoring and scheduler counters & the snake starts munching on it's own tail.
  - Getting the right hooks into the PCE for capacity planning and traffic forecasting takes these ideas to the next level.
  - Getting closer to proactive.
- Being able to apply inter-layer policy logic is new and interesting and ultimately necessary.
- Having multi-layer visibility and control gives us an automated way of making sure we're switching at the right level for maximum bang/buck.
  - Once you've filled up a port you want to switch as far down the stack as you can to avoid wasting $'s and gates.
- Agility…
  - Tight coupling of the layers that move the packets.
  - Turn-up doesn't need to take months.

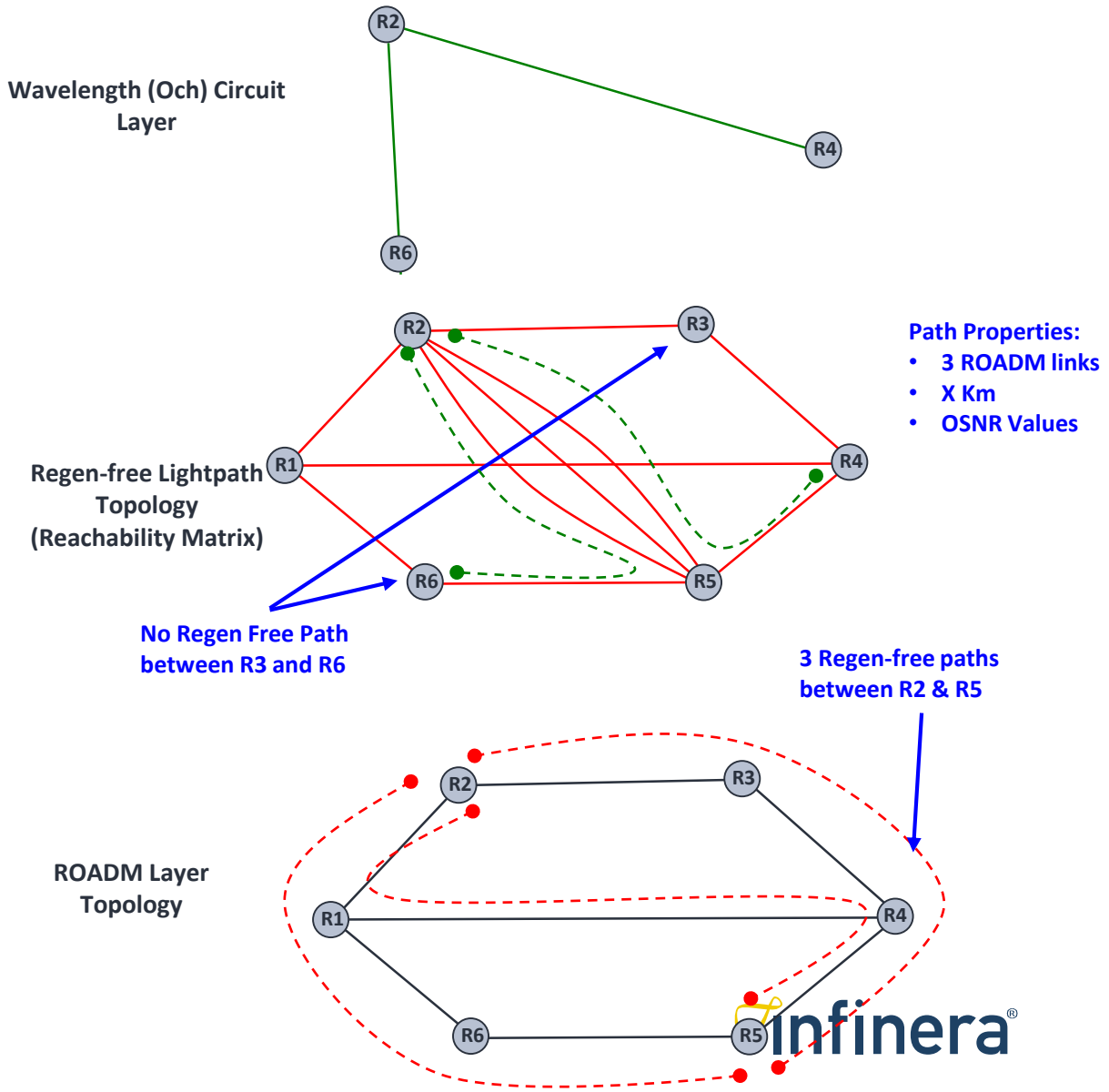
infinera®

PCE Component Architecture



L0 Restoration Planning

- Layer-0 planning is non-trivial & not real-time:
  - Lots and lots of math to be crunched to figure out where you can get optically without regeneration.
  - At scale simulations can take quite a while to complete.
    - How many places can I get to with 16-QAM without a regen, 8-QAM, QPSK, etc.
  - Results can be fed (via API) to the PCE by creating a reachability matrix

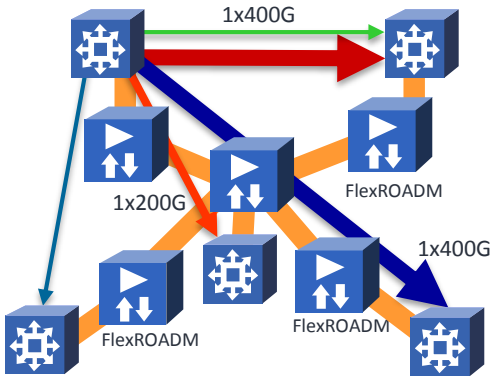


Macro Trends

infinera®
what THE NETWORK will be

Macro Trends

- Folks have been banging on about the virtues of Packet/Optical integration for quite some time now.
- The shape of VPN's is changing the shape of router roles
  - SD-WAN overlays relegate PE's to being just routers
- The economics that underpin the conceptual end-game are just now starting to come into their own:
  - Cost-effective CDC ROADM's
  - Ability to deploy huge chunks of capacity to light and pay for as required.
  - 100G coherent hitting volume
  - High-capacity feature-rich merchant packet processors designed for the carrier market



Macro Trends

- The architectural pendulum is getting closer to a sweet spot.
- Merchant silicon continues to shake things up.
  - The gap between proprietary & merchant continues to close.
- Optical networks are getting smarter, more capable/flexible & more open.
- The next few years are looking very interesting.

infinera®

what THE NETWORK will be