The State of Traffic Engineering - an ISP's Perspective

NANOG 73

Andrew Gray – Principal Engineer – Charter Communications Advanced Engineering
Andrew.Gray@charter.com

DISCLAIMER

This presentation carries a lot of information, but also a lot of interpretation of the data as presented, viewed through a specific lens, targeting specific needs.

It provides one service provider's current view as to the state of technologies, what questions needed answering, and what was valued, for one use case.

It is not intended to be the end-all be-all definitive answer to Traffic Engineering "everyone should be doing this", but to provide food for thought for other people to consider when evaluating their own situation.

What does an ISP need from Traffic Engineering

Need: A solution to sub-optimal traffic routing

A tale of two cities

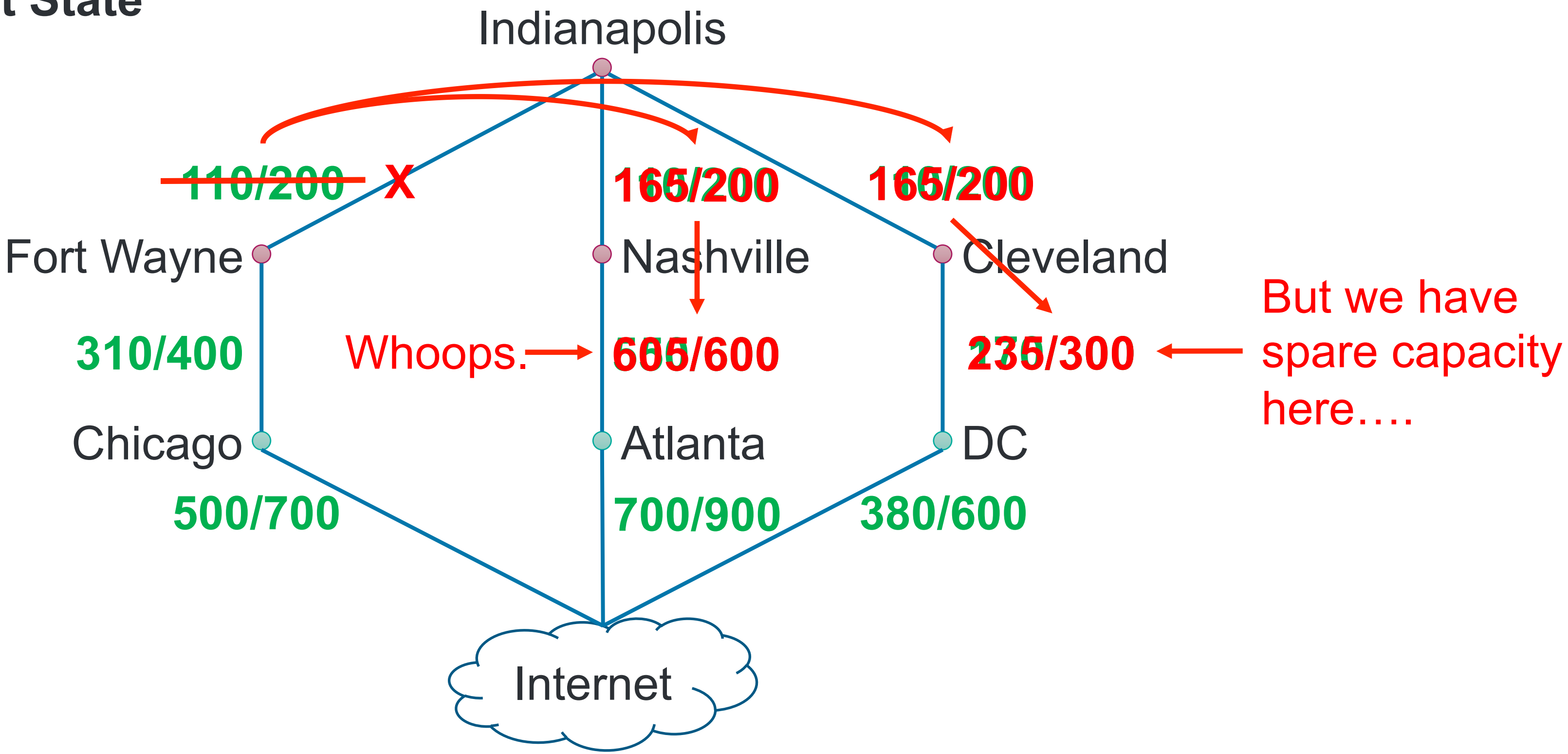
Path	Fiber (km)	RTT (ms)
Via DC	1,884	18.46
Indianapolis via Chicago, Primary	3,024	29.62
Indianapolis via Chicago, Backup	3,953	38.73

- City being serviced
- POP
- Fiber transit

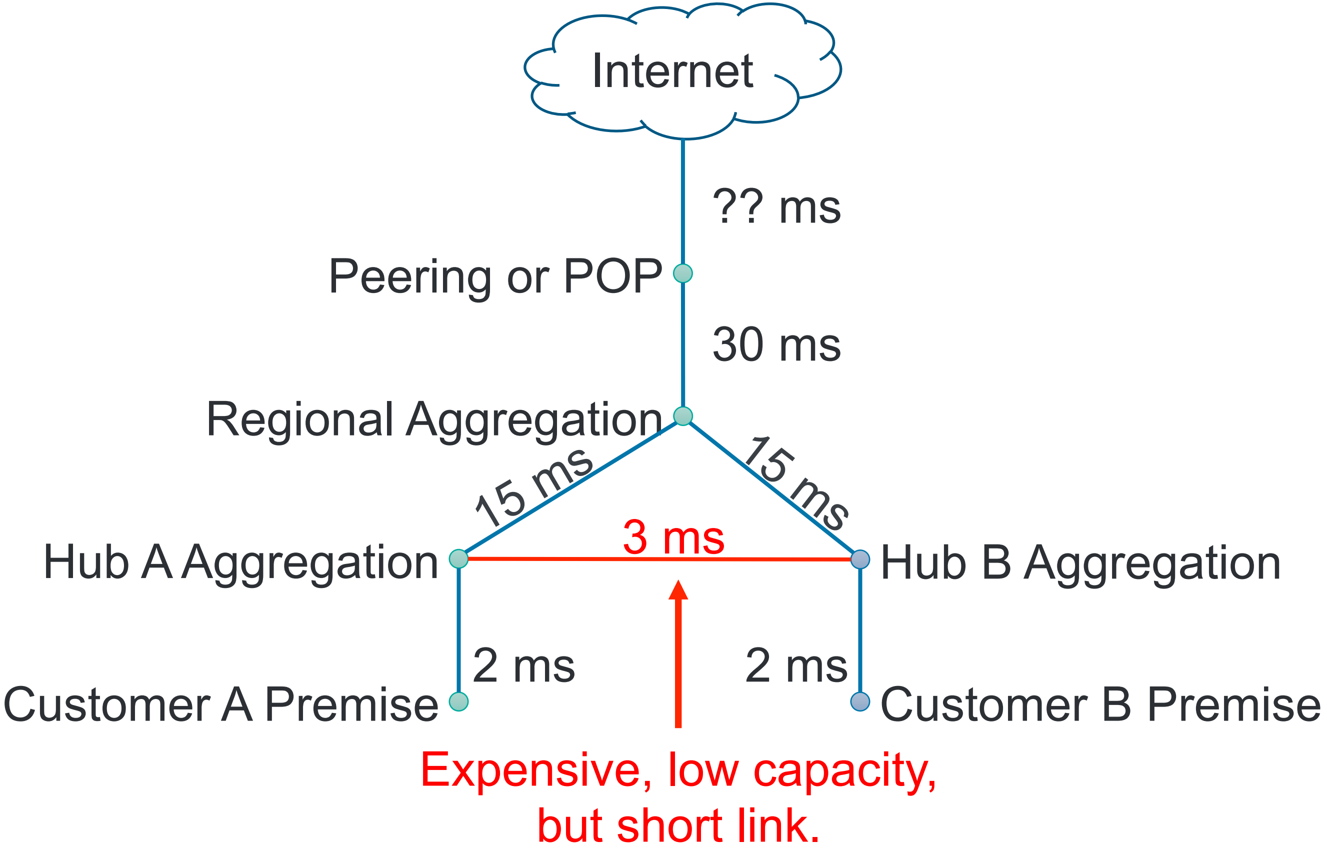
* Not actual fiber paths.

Need: Bandwidth management during outages

Start State



Need: Handle exceptional traffic differently than what routing says

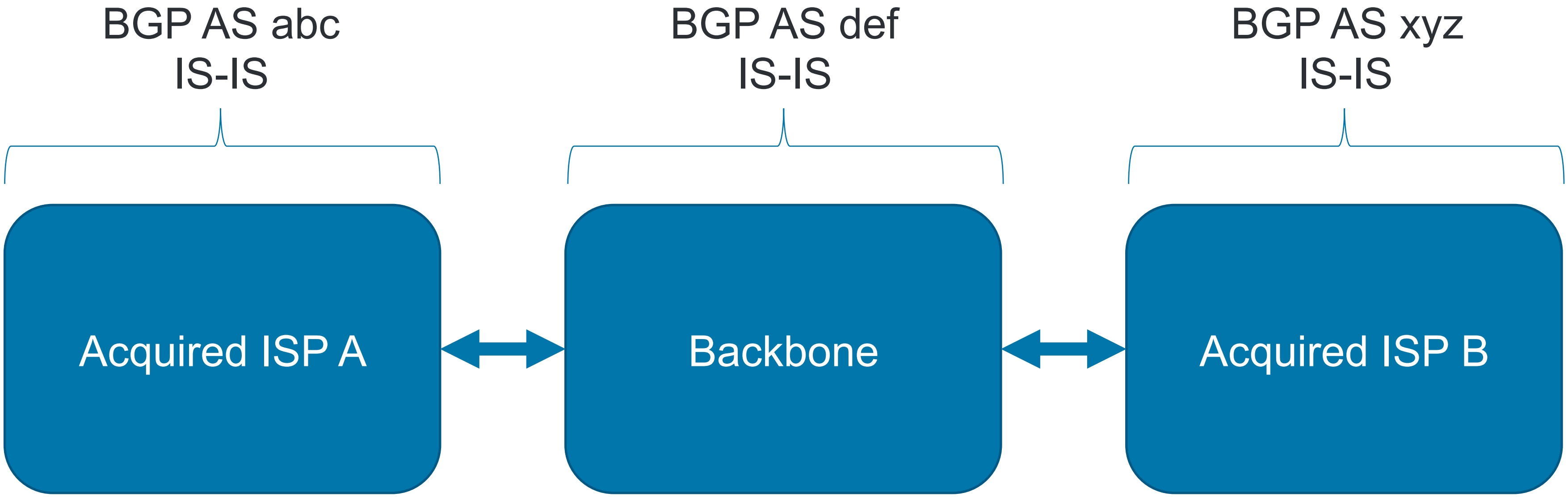


Need: Fast re-route

- Current goal is 50 millisecond restoration time.
- Technologies on the horizon (5G, for example) will require much faster (potentially as low as 10 milliseconds).

The complications an ISP faces

Complication: Multiple ASes, IGP domains, etc.

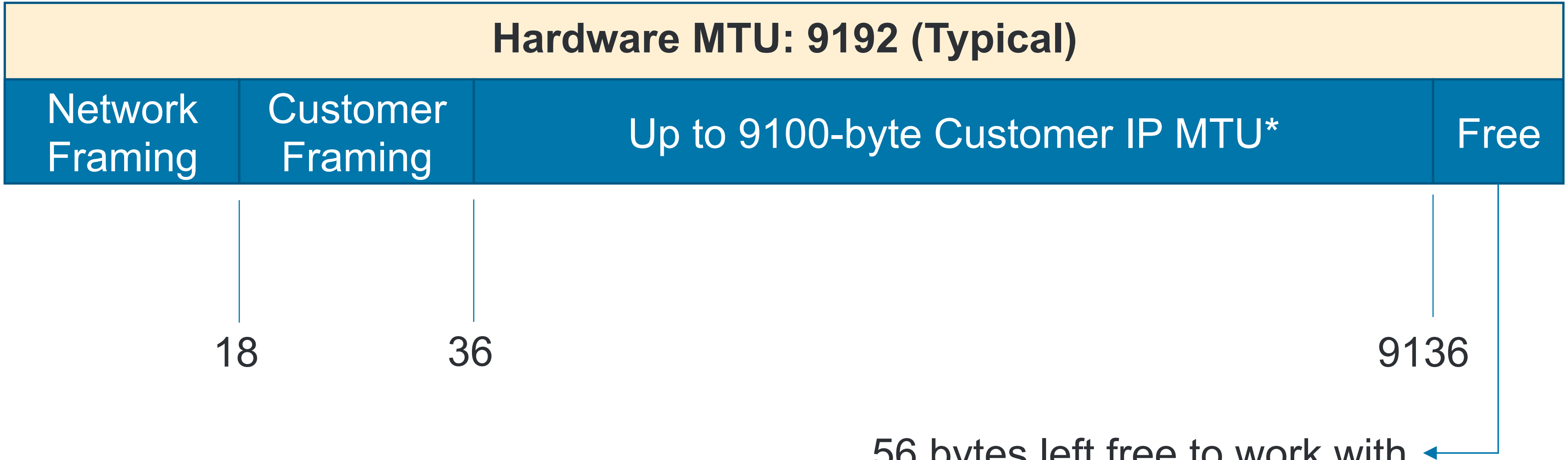


How do you get Traffic Engineering data through? (BGP-LS is at least one option…)

Complication: Multicast

- Multicast is a particularly vexing problem for traffic engineering deployments, due to the point-to-multipoint nature.
- MSOs use multicast fairly extensively for linear video distribution, and typically have multicast sessions that are single source, but potentially hundreds (or more) of receivers.

Complication: MTU



56 bytes left free to work with
Enough room for 14 MPLS labels
Enough room for SRv6 SRH plus 5 SIDs

* This is admittedly a very worst case – we normally sell circuits with a 9000-byte Ethernet MTU size, but we have legacy circuits that may be using this.

What an ISP generally does not need

Not needed (right now): Inline Network Programmability

- Referring to some of the concepts using SRv6 SIDs for special in-line network operations.
  - NOT referring to Netconf/OpenConfig type technologies – we very much need those! If only they were ready today…
- At least, not immediately. Potentially in the future, but figuring out actual, real world use cases for our backbone goes way off into the theoretical.
  - There are potentially some corner use cases in non-backbone packet handling.
- We are a bunch of network engineers who are expanding into programmers, as opposed to the other way around. This presents a different perspective as to the usability and necessity of this.

Not needed: Very fancy controller capabilities

- There is still a large amount of work being done on controllers.
  - Especially in terms of multi-vendor support, which is virtually non-existent.
  - What little does exist, vendors are not terribly fond of supporting.
- We haven’t yet found an actual use case, on our network, that requires esoteric functionality.
- Having a piece of software controlling the network makes those of us who have been around for a while very, very nervous….
  - We still get very abhorrent behavior with core router OSes, and they’ve been around for decades, with large support and testing organizations behind them.

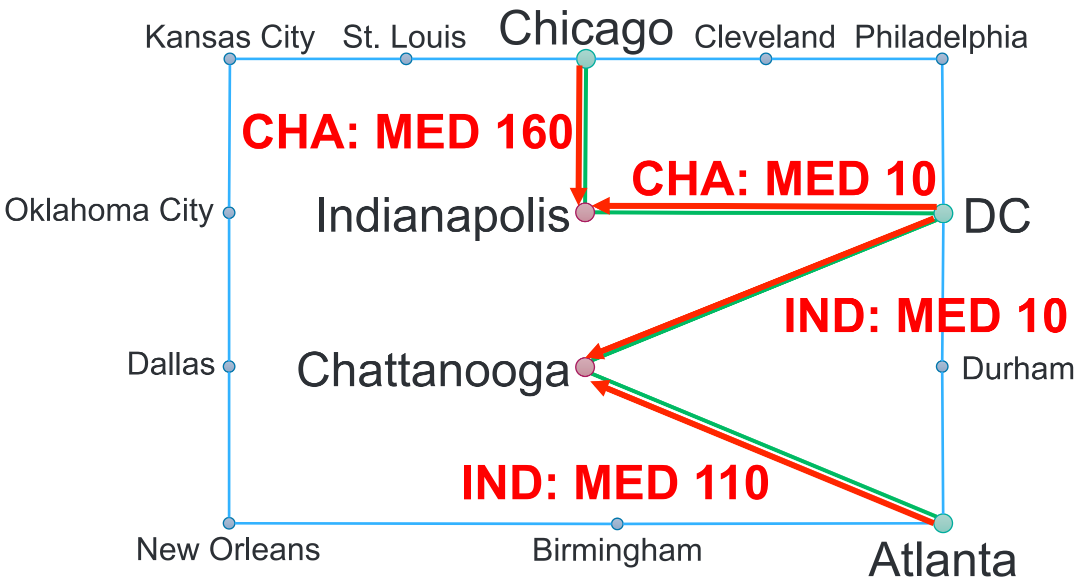
Possible solutions to review

Static Routing

- Not “Traffic Engineering” in the normal sense.
- Can work in a couple of use cases
- Is generally how a lot people solve these problems today
- Not granular or dynamic enough for our use.

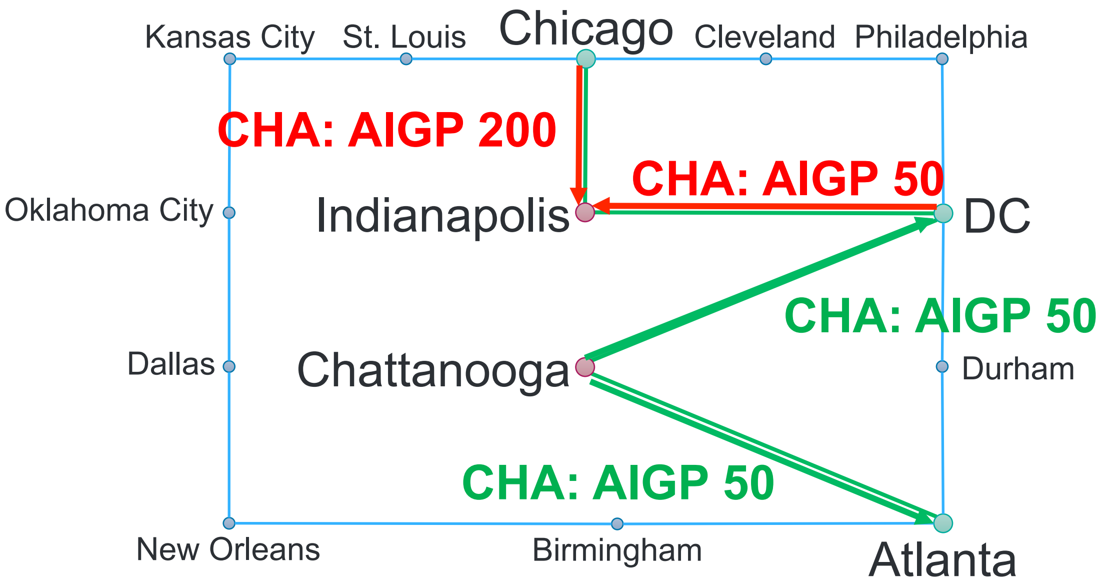
IGP as MED

- Not “Traffic Engineering” in the normal sense.
- This solves some of our sub-optimal traffic routing problems.



AIGP

- Not "Traffic Engineering" in the normal sense.
- Somewhat similar to MED, but we had a concern about it.



BGP Path Selection (Multi-vendor)

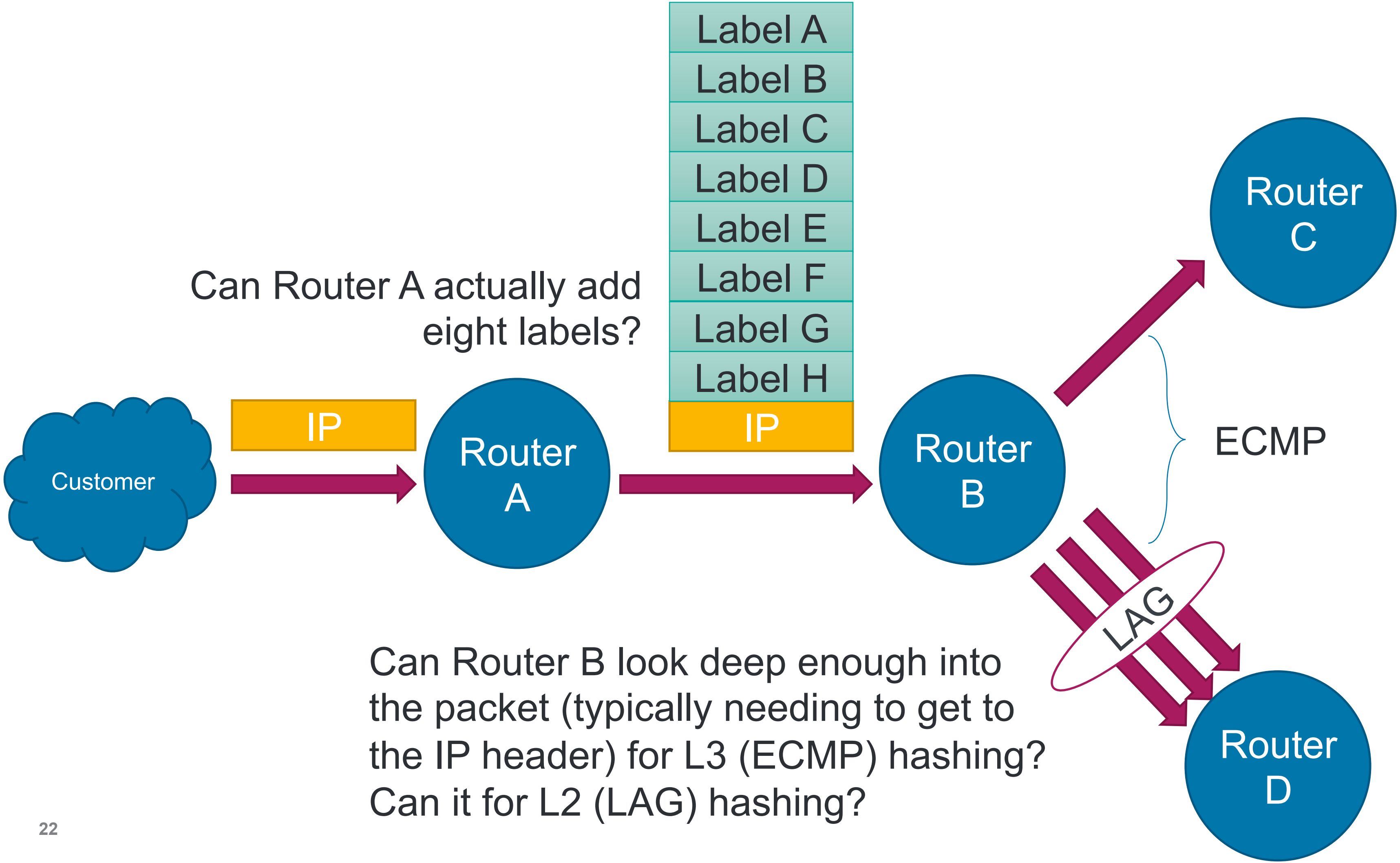
<table>
<tr><th>Vendor A</th><th>Vendor B</th><th>Vendor C</th></tr>
<tr><td colspan="3">Verify next hop is reachable</td></tr>
<tr><td colspan="2">Prefer route with lowest origin validation state</td><td></td></tr>
<tr><td></td><td colspan="2">Prefer lowest route-table preference</td></tr>
<tr><td>Prefer highest weight</td><td></td><td></td></tr>
<tr><td colspan="3">Prefer highest local preference</td></tr>
<tr><td>Prefer locally originated/ redistributed</td><td></td><td></td></tr>
<tr><td colspan="3">Prefer routes with an AIGP metric over those that don’t.</td></tr>
<tr><td colspan="3">Prefer routes with a lower AIGP metric.</td></tr>
<tr><td colspan="3">Prefer shortest AS path.</td></tr>
<tr><td colspan="3">Prefer lower route origin (IGP < EGP < Incomplete)</td></tr>
<tr><td colspan="3">Prefer lower MED (routes with no MED are MED=0).</td></tr>
<tr><td></td><td></td><td>Prefer strictly internal paths.</td></tr>
<tr><td></td><td>Prefer routes with lowest owner type (BGP < BGP-LABEL < BGP-VPN)</td><td></td></tr>
<tr><td colspan="3">Prefer strictly EBGP paths over external paths learned via IBGP</td></tr>
<tr><td>Prefer path that was received first</td><td></td><td></td></tr>
</table>

RSVP-TE

- RSVP itself was standardized in RFC 2205 in September, 1997 (20 years ago)
- RSVP has extensions for point-to-multipoint LSPs standardized in RFC 4875 in May, 2007.
- Lots of work has been done to extend and standardize it, with lots of multi-vendor interop testing.
- RSVP requires one additional protocol to be stood up to support it (RSVP itself).
- RSVP optionally can use extended traffic engineering metrics, link coloring, etc. that can be configured.

SR-TE

- Segment Routing is in RFC draft status (https://datatracker.ietf.org/doc/draft-ietf-spring-segment-routing/), first version June 2013, current version (-15) January 25, 2018.
- Segment Routing almost requires the use of a Path Computation Engine (PCE) for large scale network operation.
- Thus, two new protocols must be stood up: BGP-LS (which requires a BGP session flap) and PCEP.
- Different platforms, and even different line cards on those platforms, support different label stack depths, causing additional complications:


Label A
Label B
Label C
Label D
Label E
Label F
Label G
Label H
IP
Can Router A actually add
eight labels?
Customer
IP
Router
A
Router
B
Router
C
ECMP
LAG
Router
D
Can Router B look deep enough into
the packet (typically needing to get to
the IP header) for L3 (ECMP) hashing?
Can it for L2 (LAG) hashing?

SRv6-TE

- SRv6 is in RFC draft (https://datatracker.ietf.org/doc/draft-ietf-6man-segment-routing-header/), first version March 2014, current version (-08) January 20, 2018.
- I imagine everyone here is sick of hearing about SRv6, Segment Routing Headers (SRH), etc. – I won’t repeat the many, many, many presentations about it.
- The size of the SRH and associated SIDs limits it slightly on our network (we don’t have the overhead to spare)
- We already have MPLS deployed and enabled on our network, so SR is a less impactful change.
- Vendor support is very basic at this point, with limited hardware acceleration.
- Older hardware will most likely never support it due to hardware restrictions.

PCEs

- Path Computation Engines are useful for all three TE technologies (RSVP, SR, SRv6).
- We are keeping a very close eye on this space, but the key requirement for us is multi-vendor interop, which is very weak on the commercial providers.
- We are taking looks at the various open source options (OpenDaylight, Ryu, etc.) as well.
- We view these as almost a necessity before we do large scale deployments (although at least one other MSO has deployed a substantial RSVP topology without).

Conclusions

Summarizing 20+ years of technology in one slide

Requirement	IGP as MED	AIGP	RSVP	SR	SRv6

Summary: Current and Future

- **Every network is different.**
- We are planning on a couple different approaches for the present day.
- We are keeping an eye on SR and PCEs for the future.
  - SR + TI-LFA + PCE has some very interesting potential.
- Right now, SRv6 is off our radar.

Q & A

Feel free to reach out: Andrew.Gray@charter.com