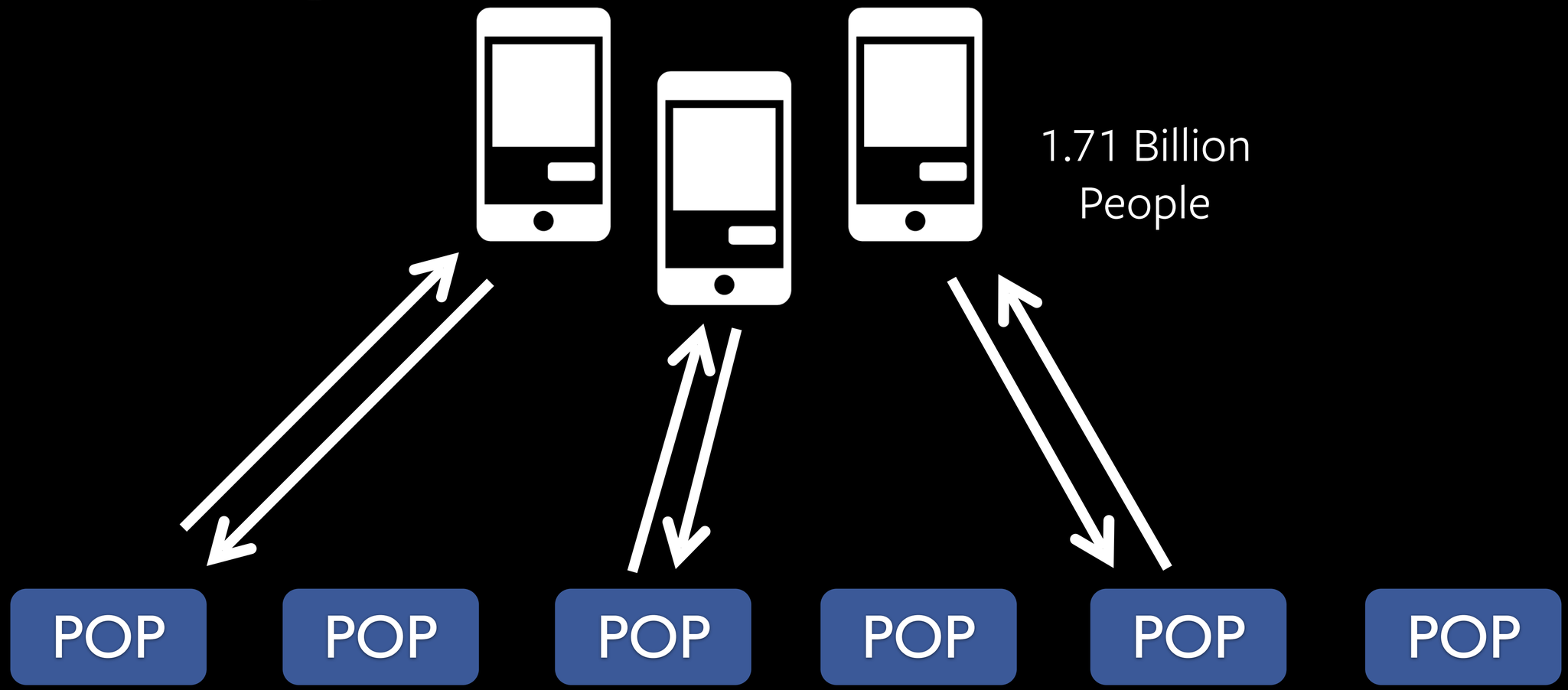
Being Open:

How Facebook Got It’s Edge

James Quinn
Network Engineer, facebook


15% of global egress
is already IPv6
1.71 Billion Users
1.1+ Billion Daily Users
2 Billion Photos daily
60 Billion Messages daily

Our edge connects to the world



What is the service we support?

- **dynamic**
  - newsfeed
  - likes
  - status updates
- **static**
  - images
  - css
  - js

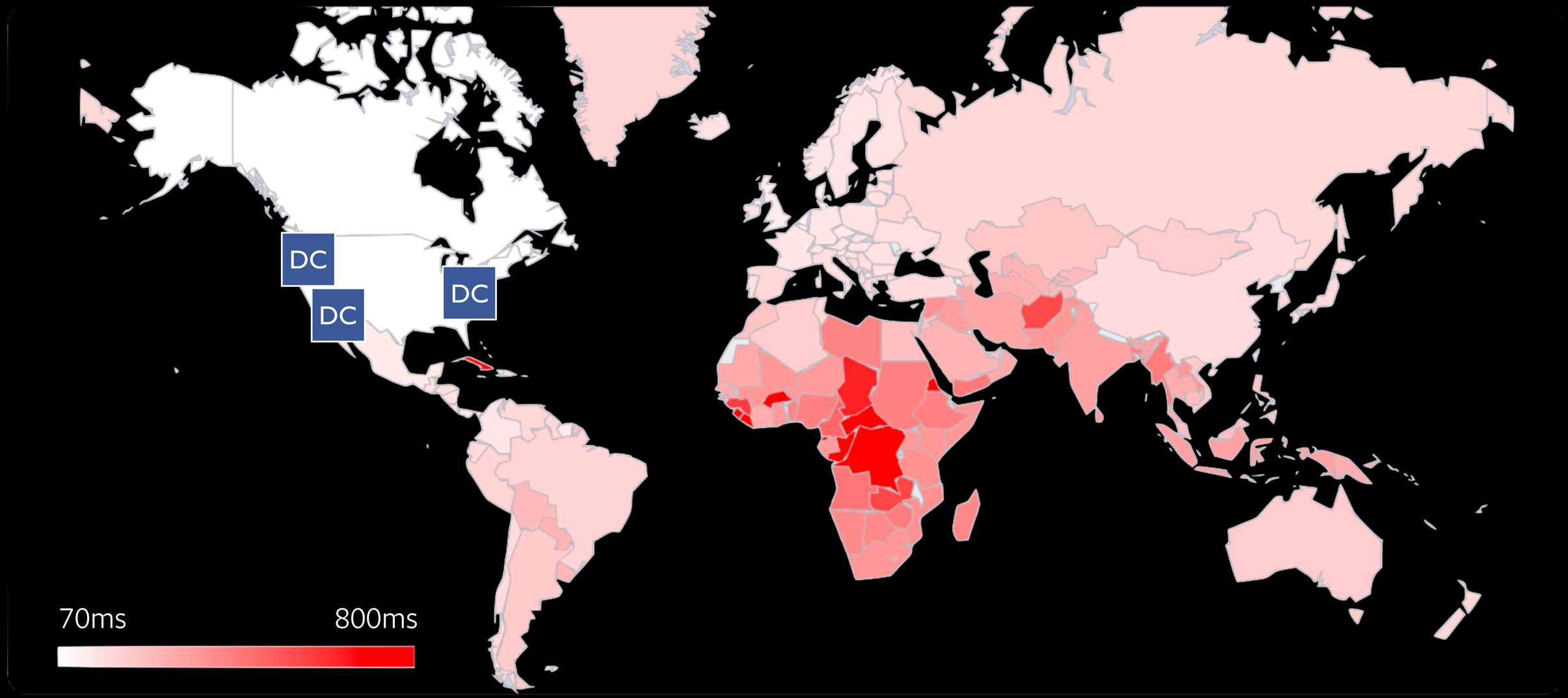




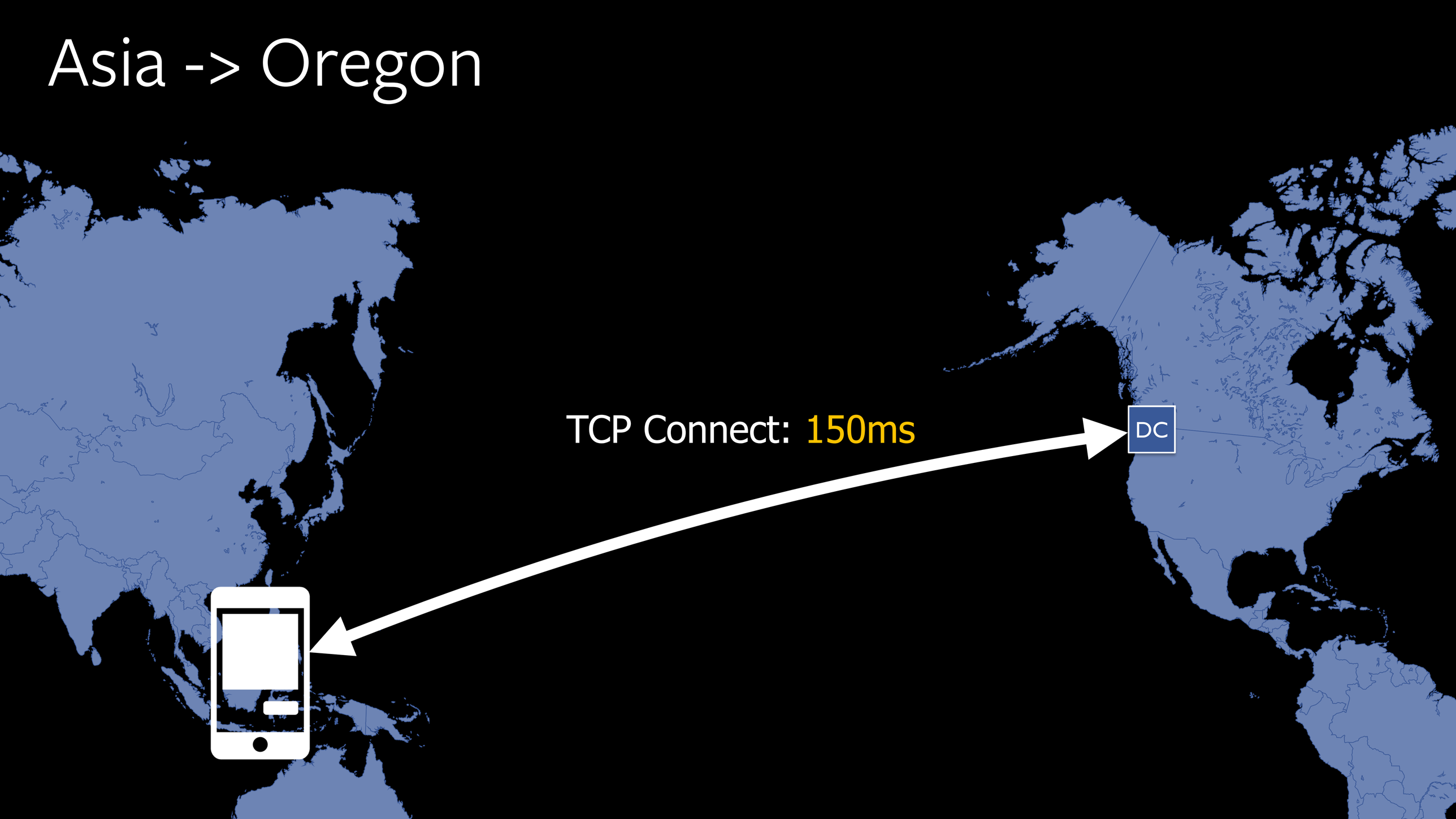
Credit: Lou Stejskal
"Justin Bieber Q & A"
https://www.flickr.com/photos/loustejskal/26577668795
Creative Commons BY 2.0 License

International RTT

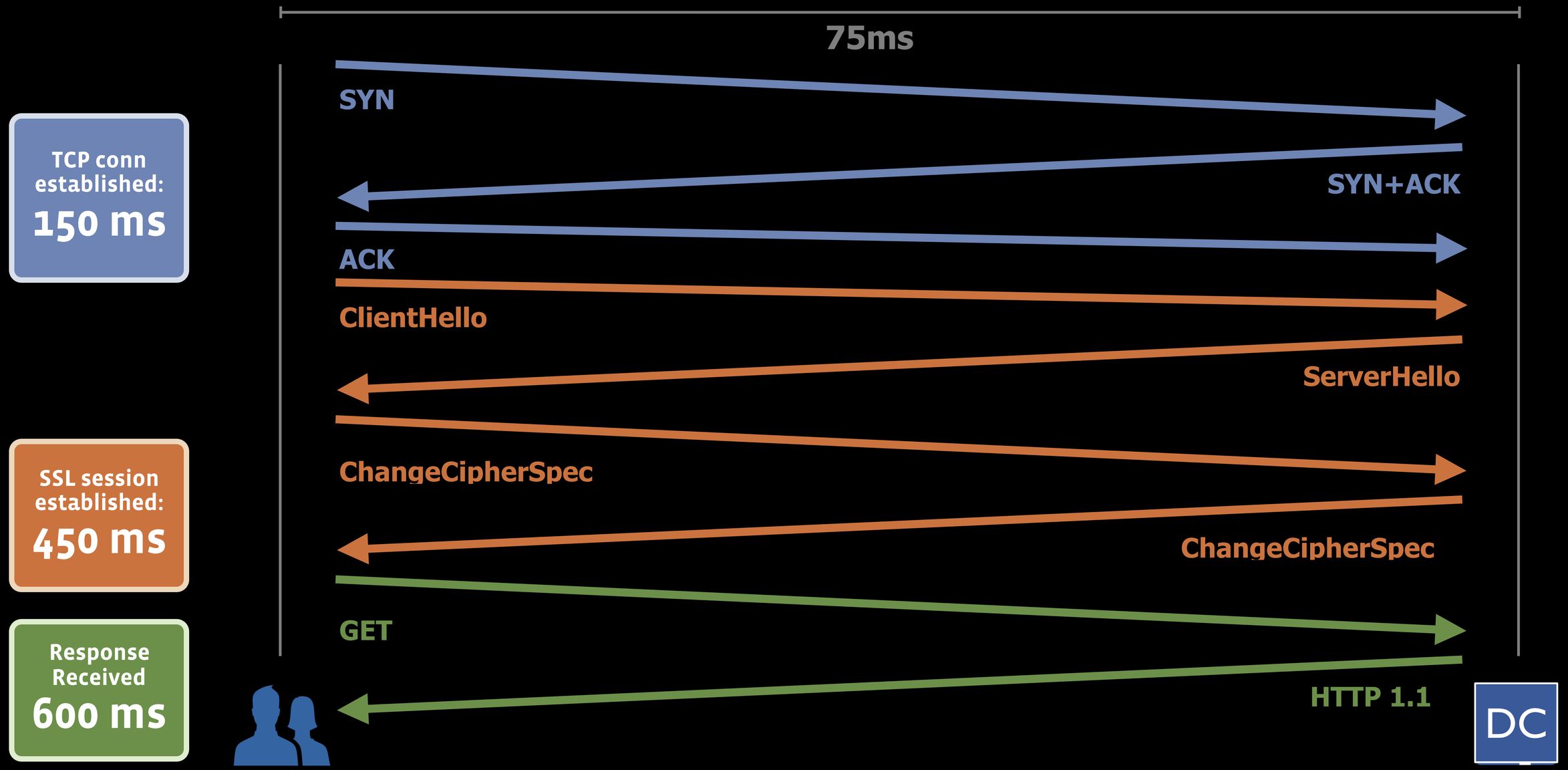
circa 11/2011



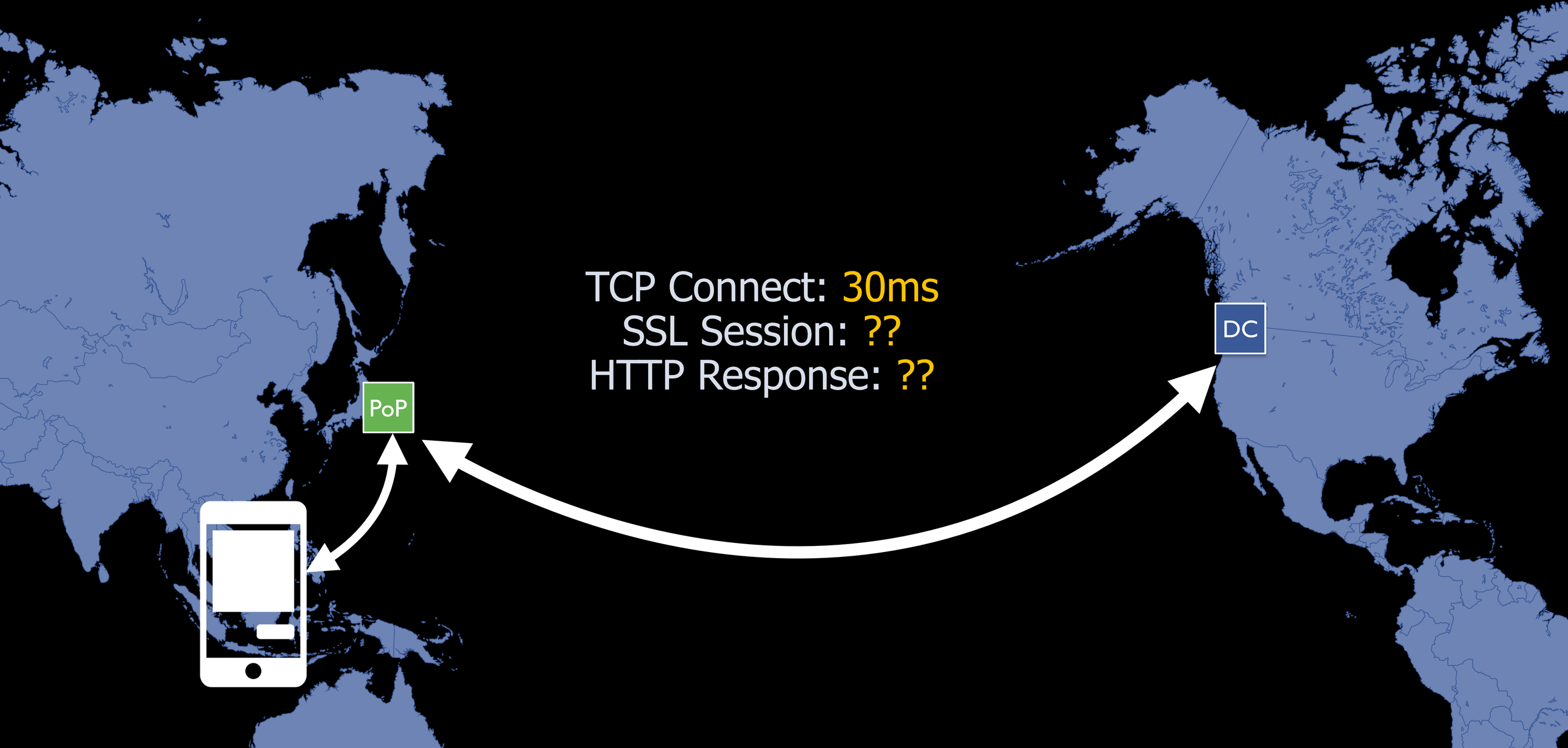
Asia -> Oregon



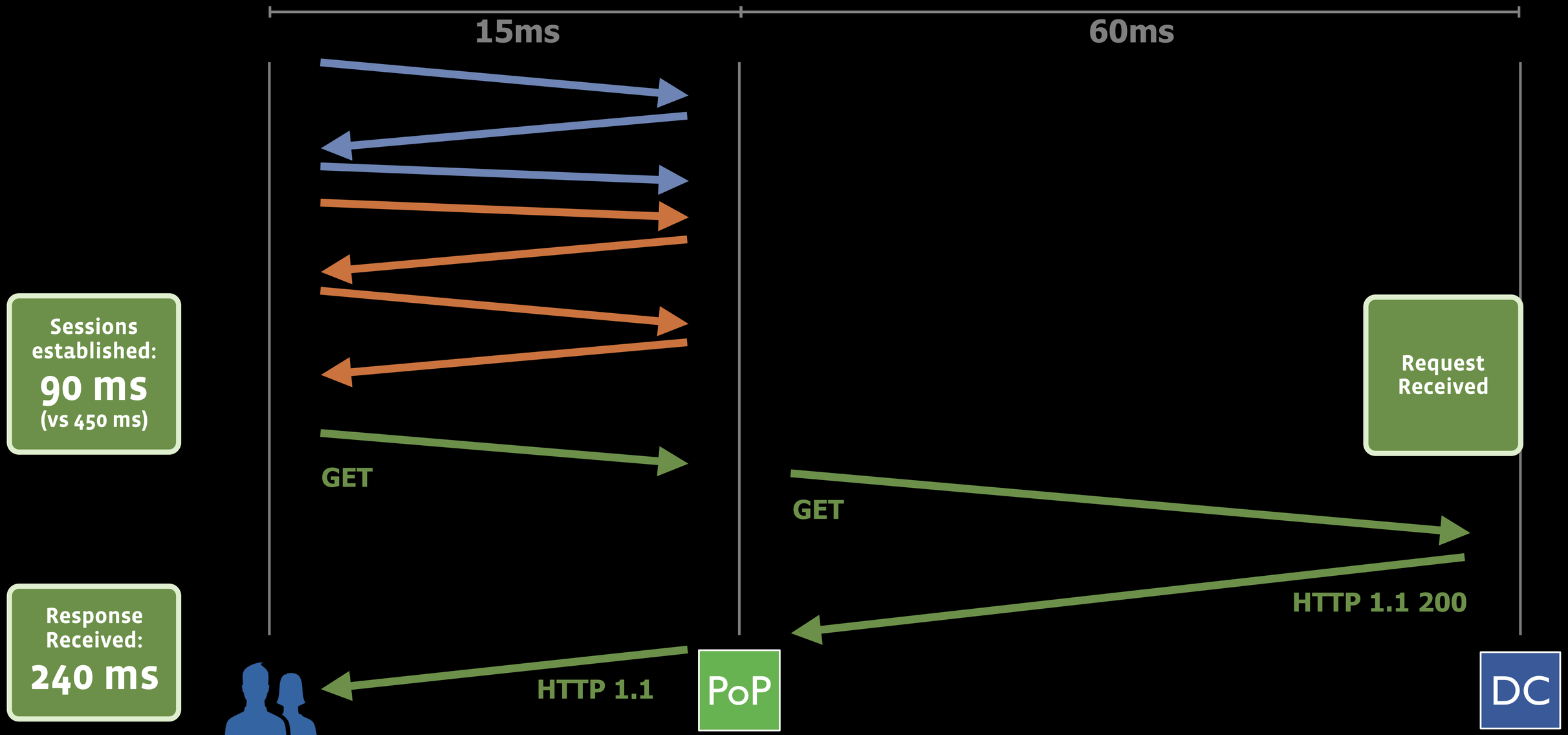
HTTPS Asia -> Oregon



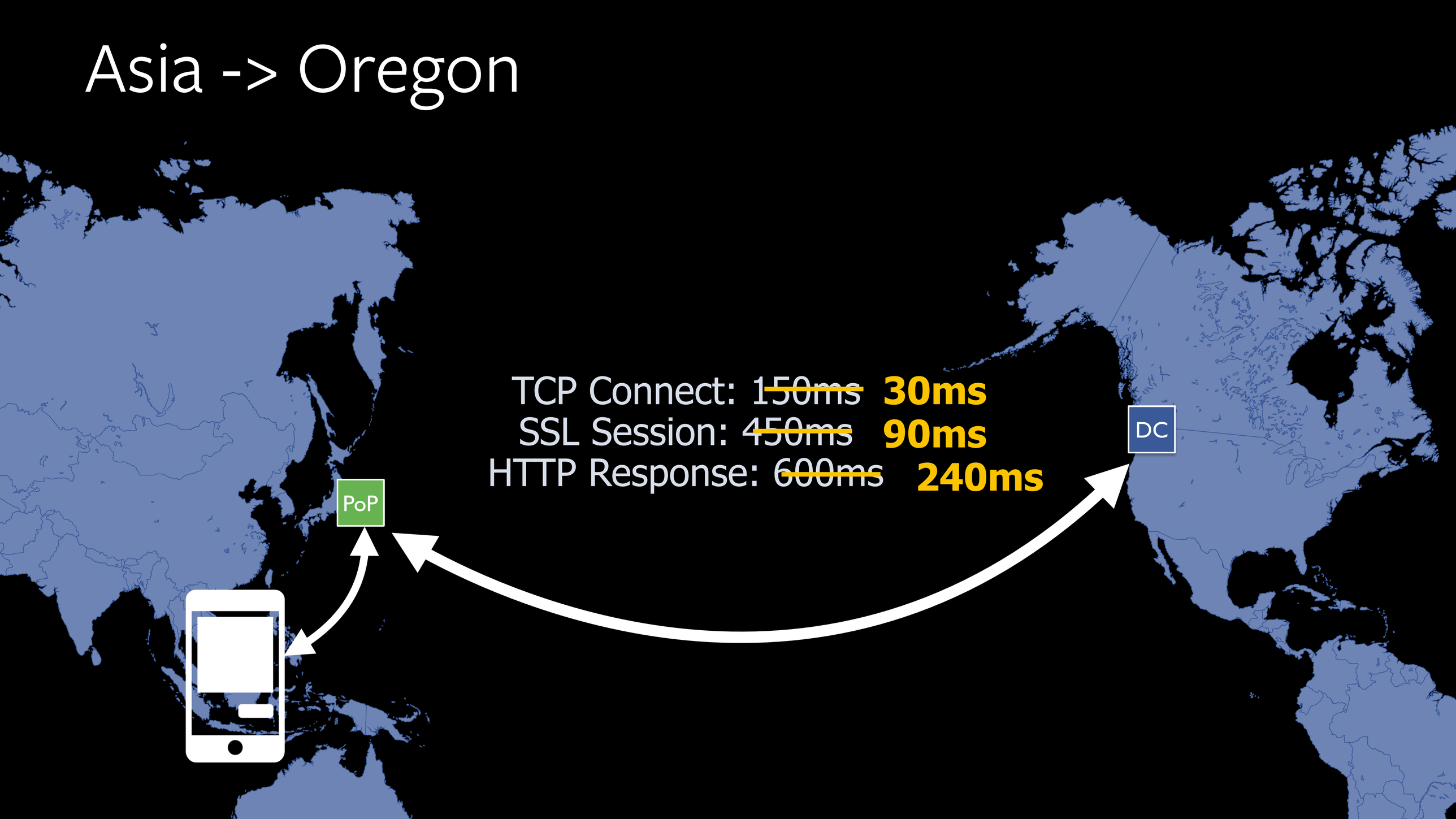
Asia -> Oregon



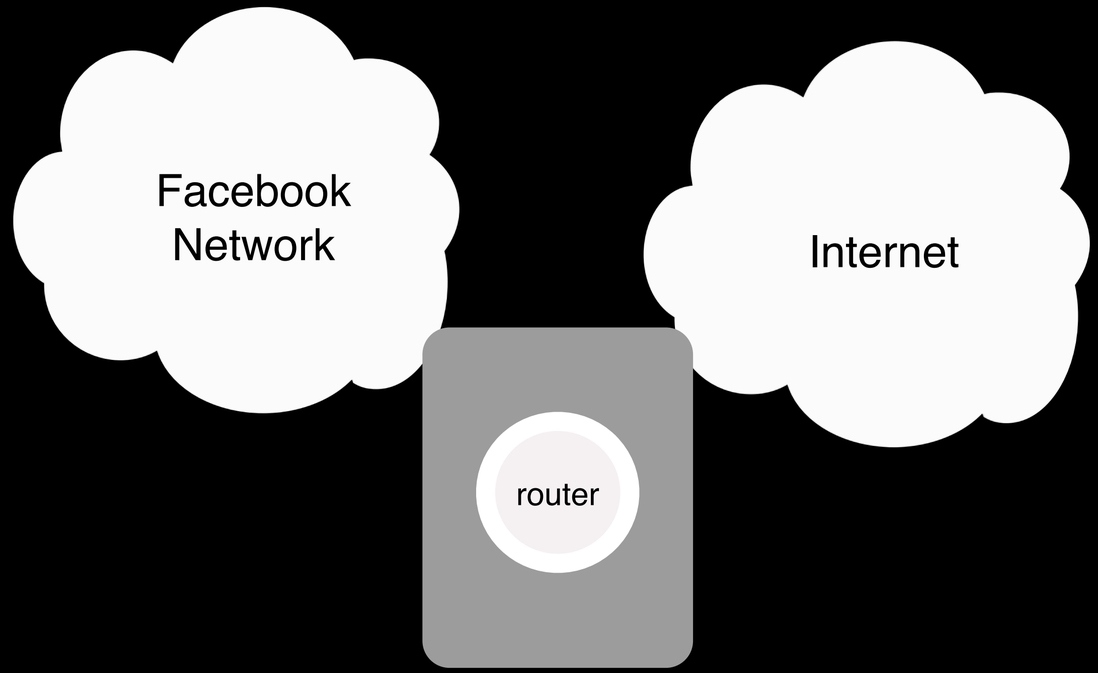
HTTPS Asia -> POP -> Oregon



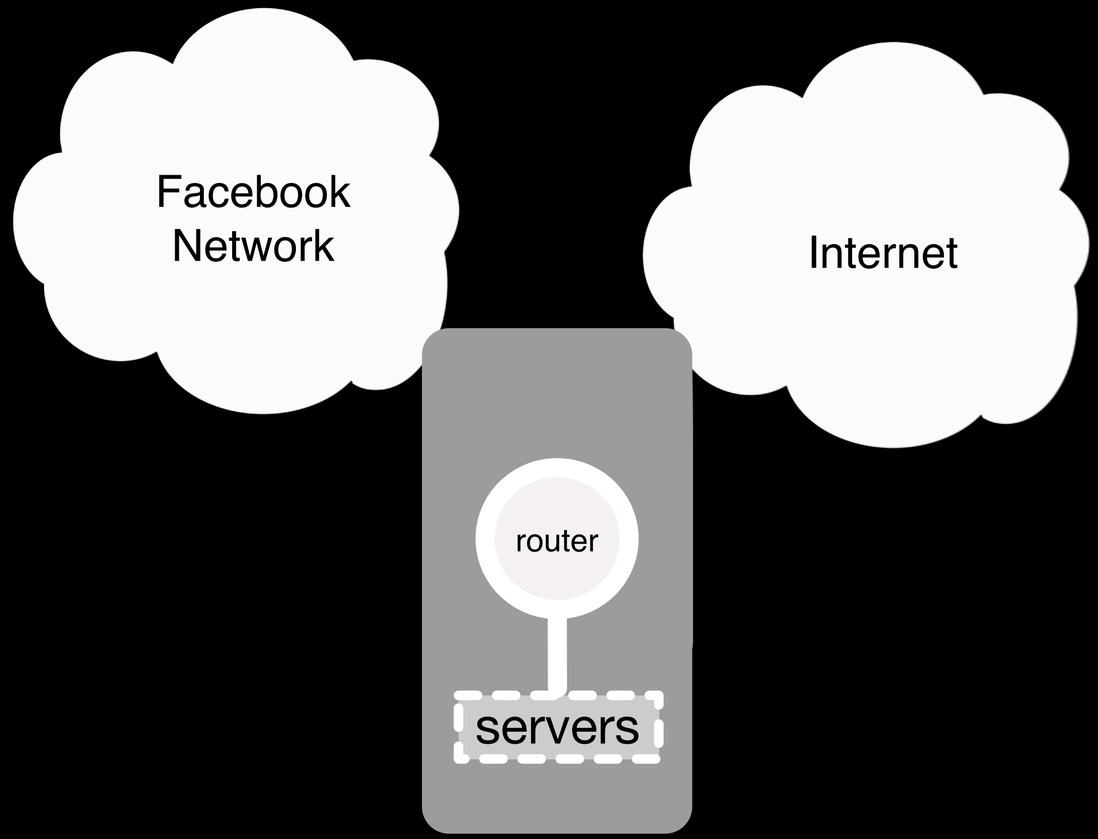
Asia -> Oregon


TCP Connect: ~~150ms~~ **30ms**
SSL Session: ~~450ms~~ **90ms**
HTTP Response: ~~600ms~~ **240ms**
PoP
DC

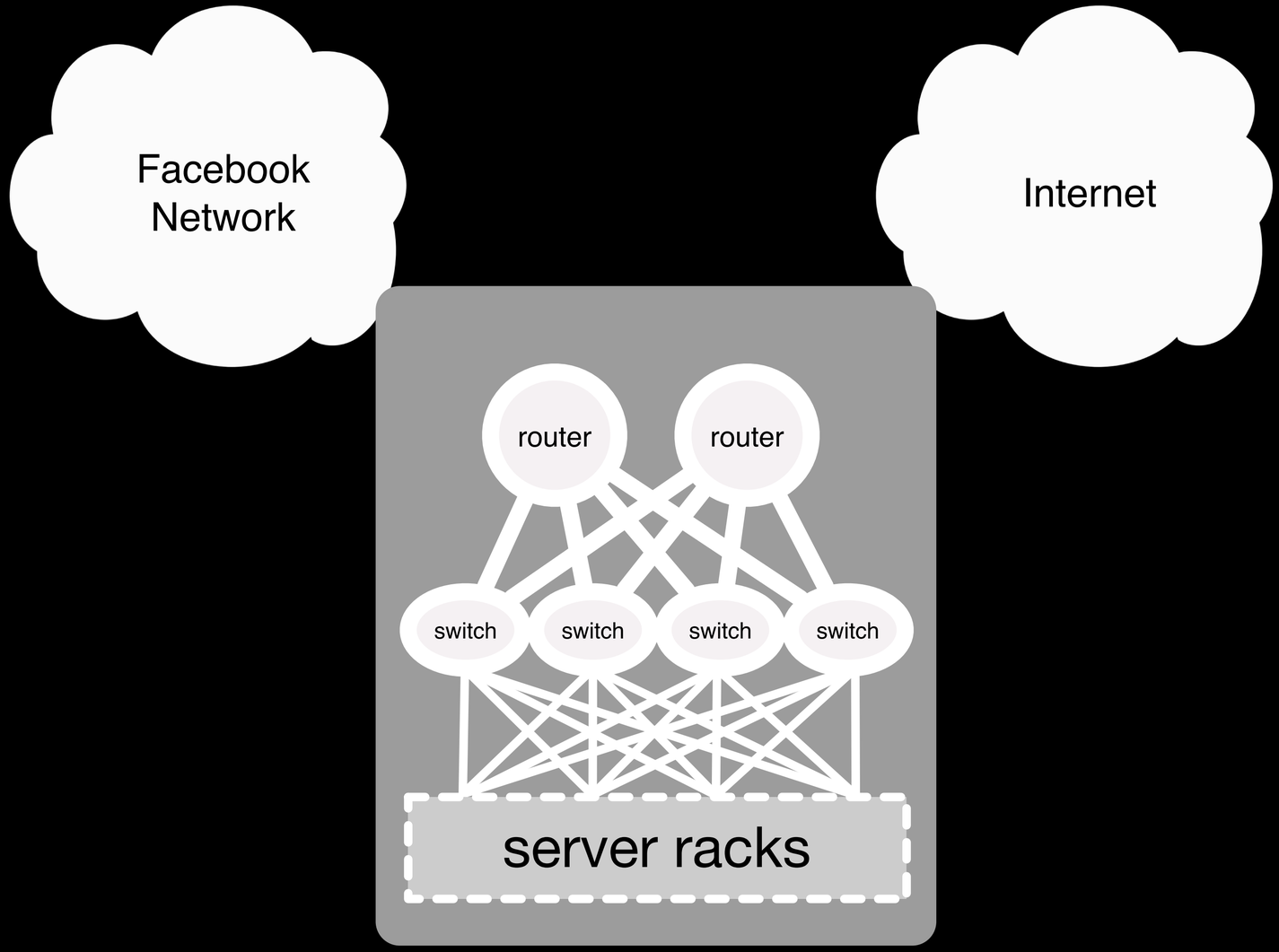
edge routers -> edge clusters



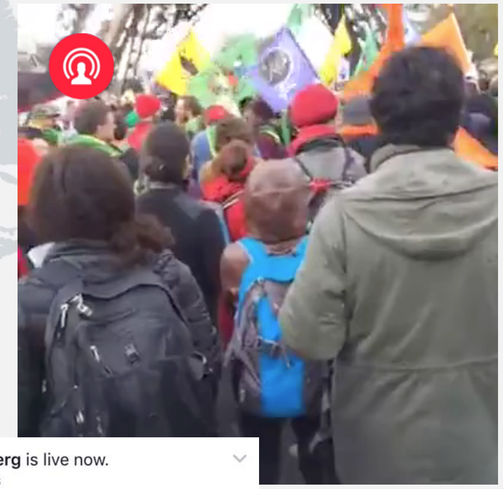
edge routers -> edge clusters



edge routers -> edge clusters



LIVE Explore to see who's live around the world





Mark Zuckerberg is live now.
Public · 2 mins

June 1st 2016



Like Comment Share

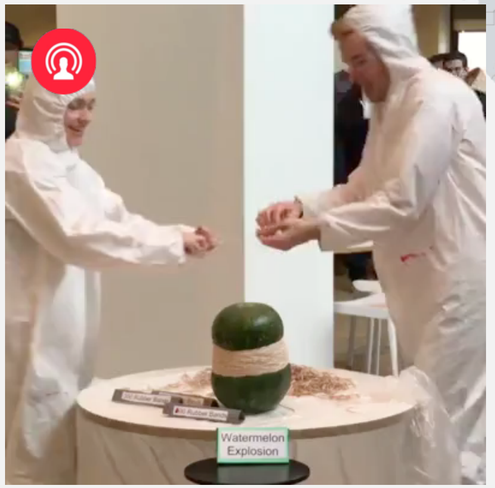
9.7K

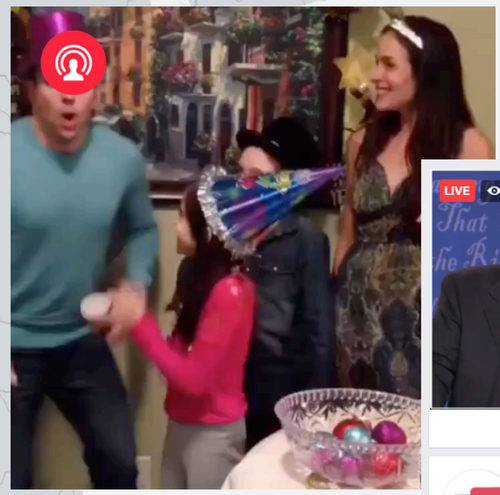
277 shares

يمنا ال اثكزي
Hey mark, congrates on the space chat initiative :)
Just now · Like · Reply

Video Is Growing









Video Discussion

Share Your Thoughts

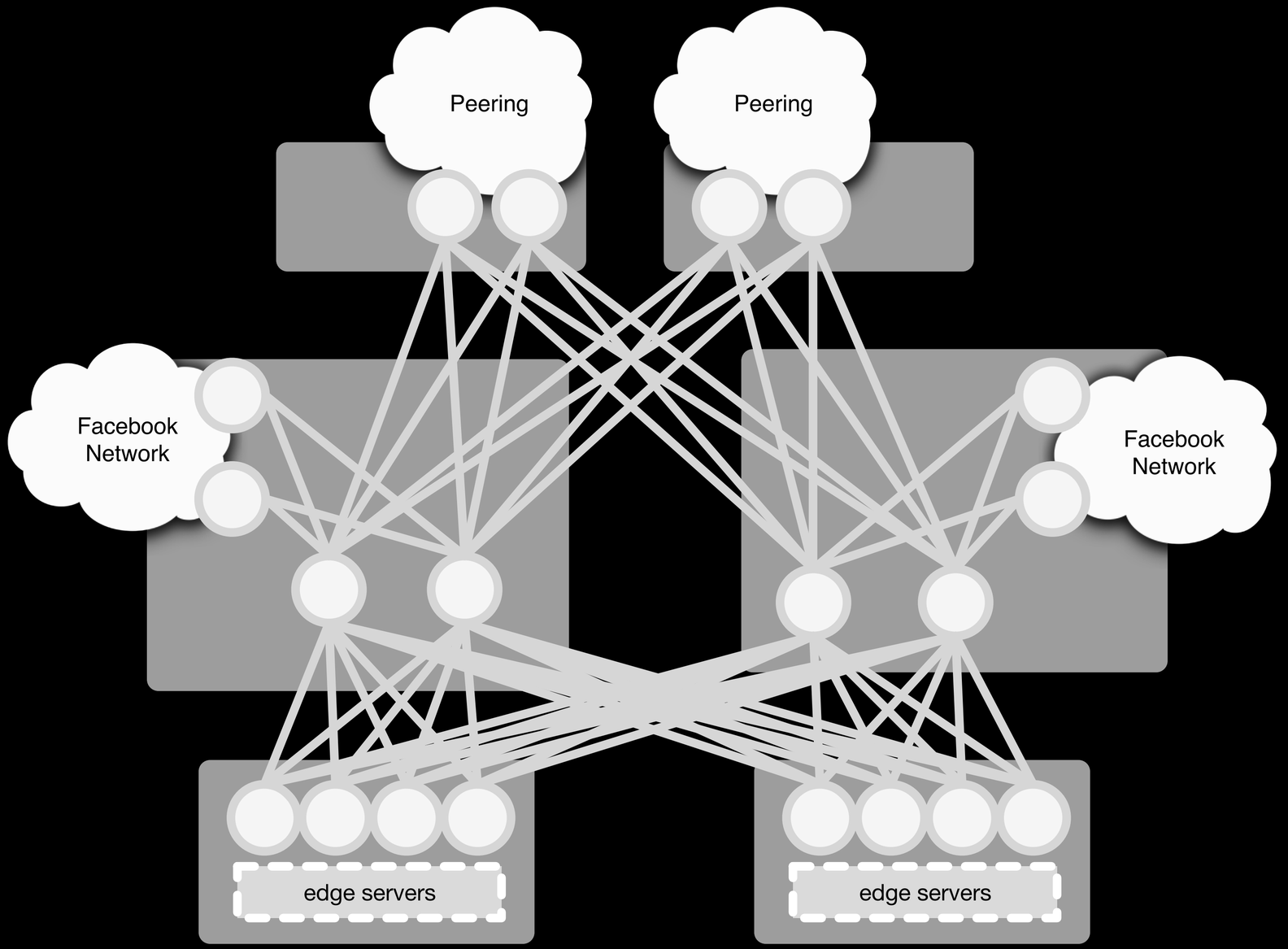
What do you think about the presidential debate? Share your thoughts with friends.

Post

Christopher Kusek
Just now · Twitter ·

My favorite part about the Sony hack investigation by the FBI was where it was definitely without a doubt North Korea. Or, not. #Debates

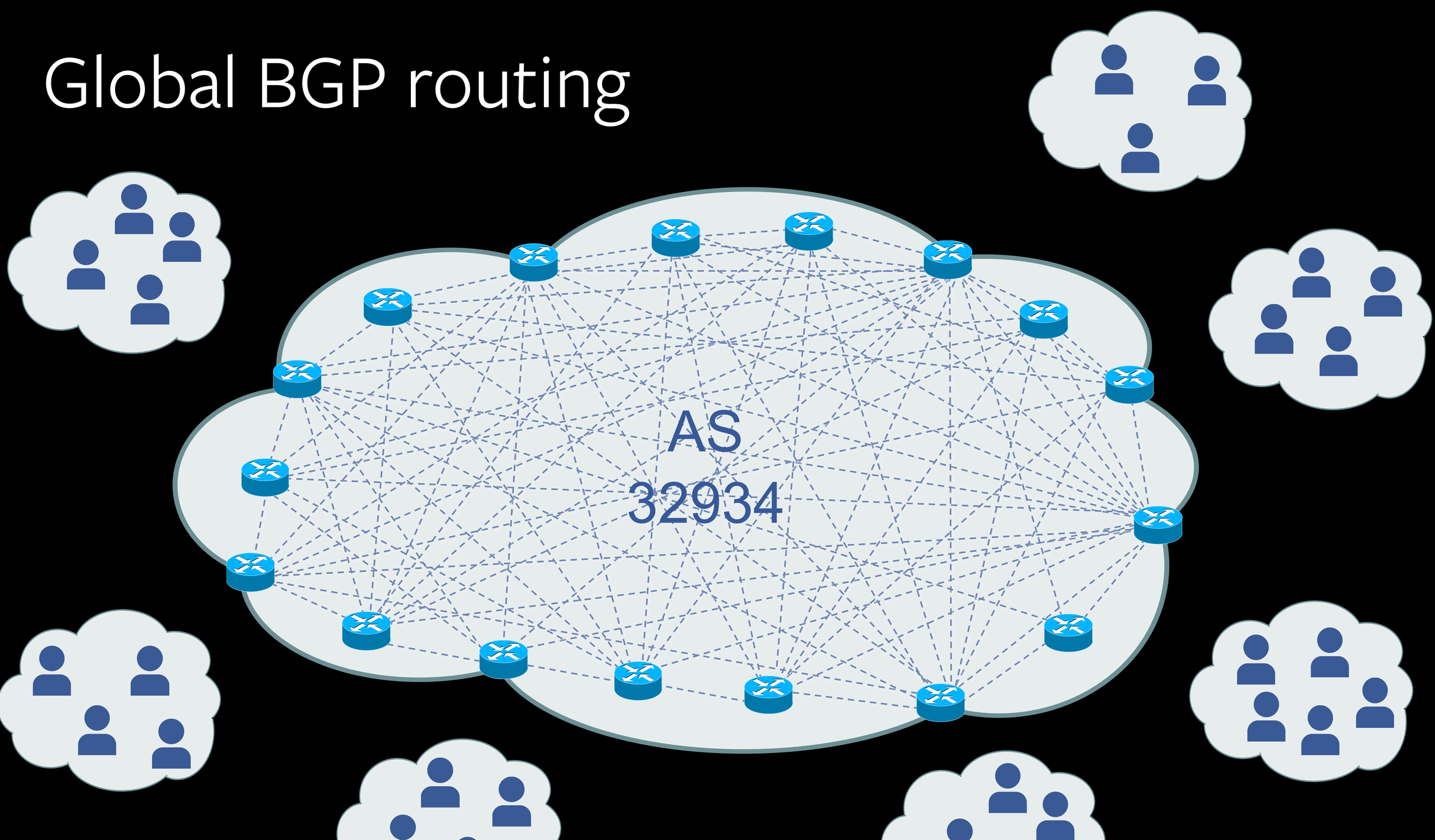
-> edge metro topology



100G
Everywhere!

Evolving beyond BGP

Global BGP routing



Which POP is best?

Considerations:

- Closest Edge to user





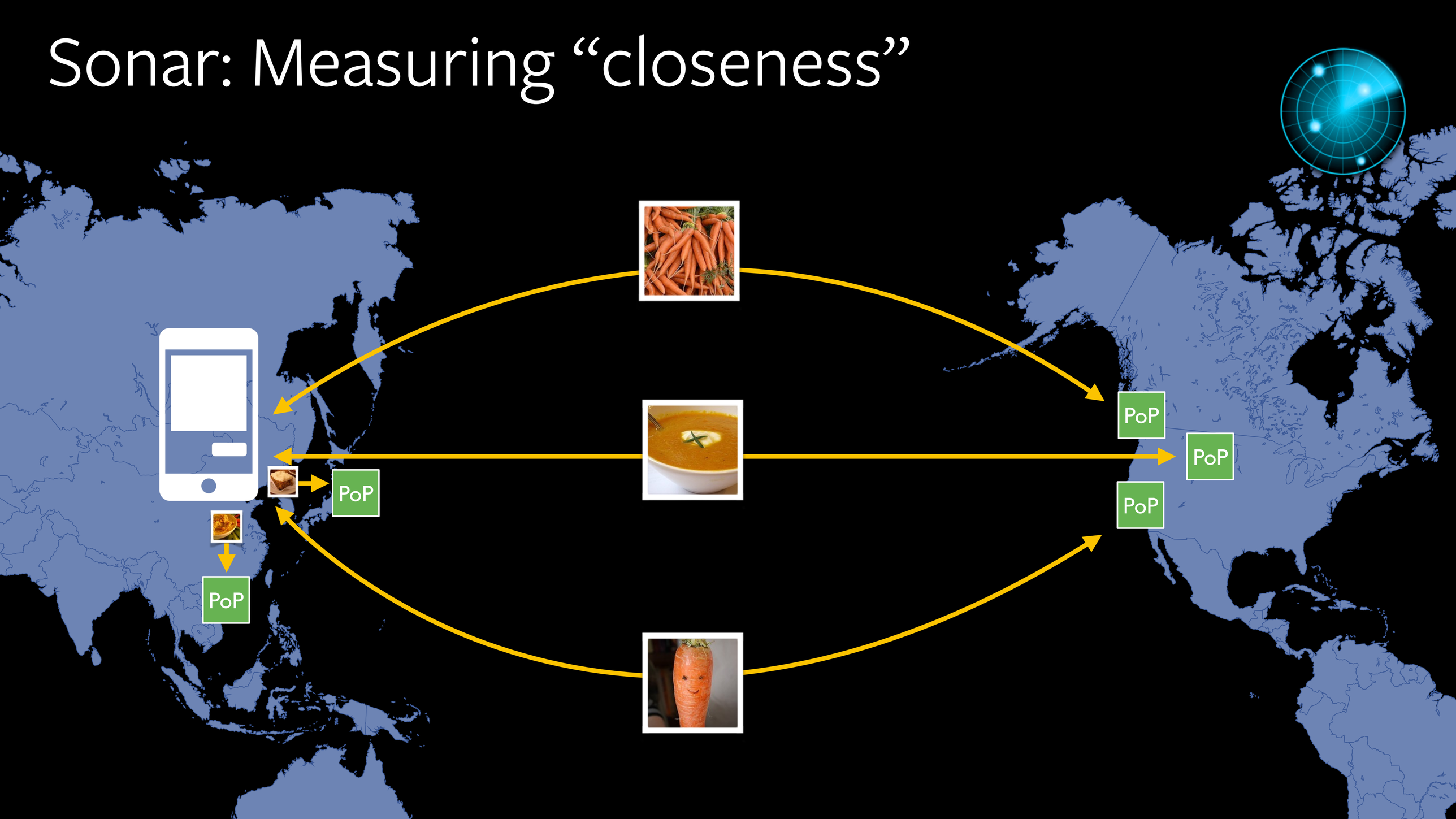




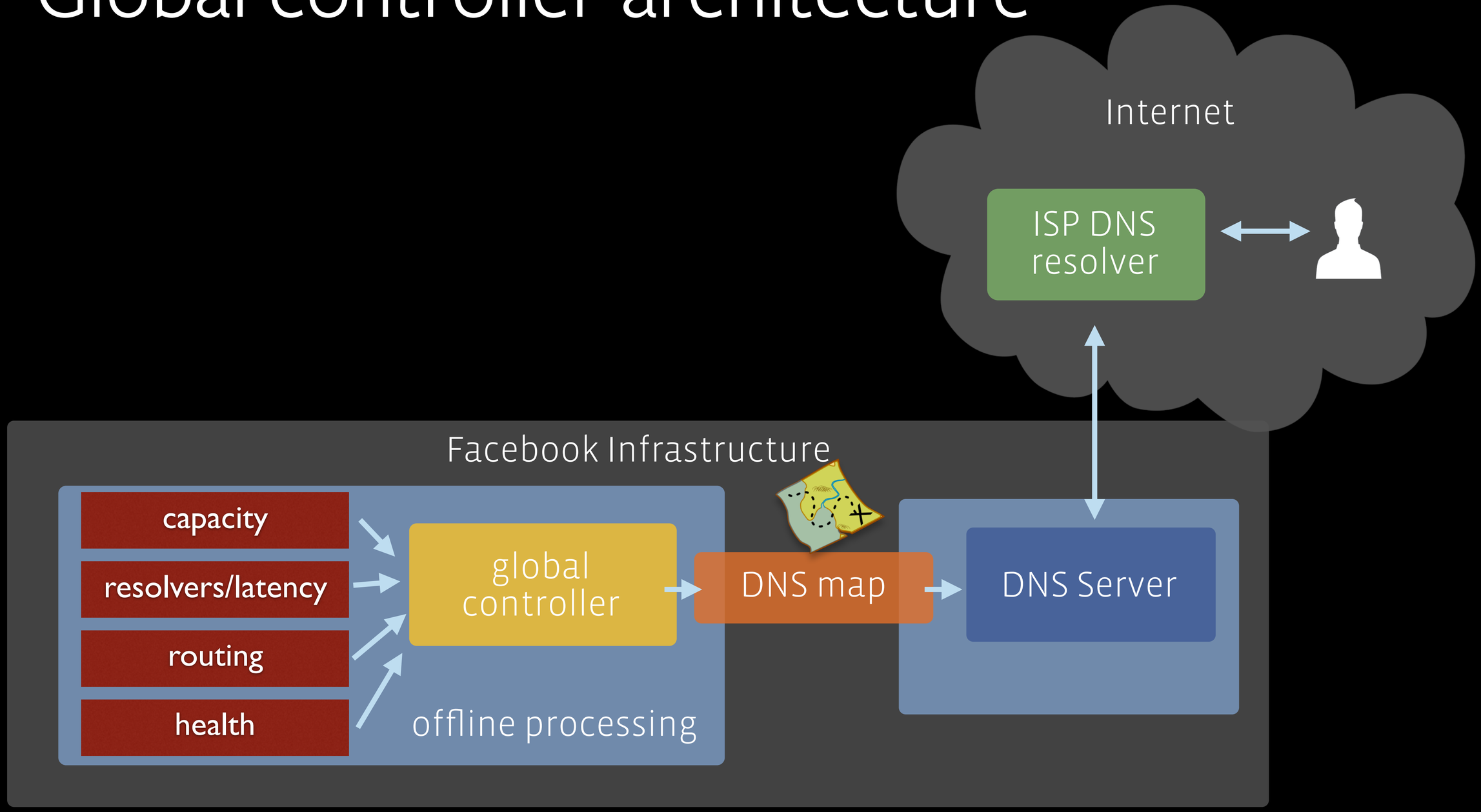
POPs

???

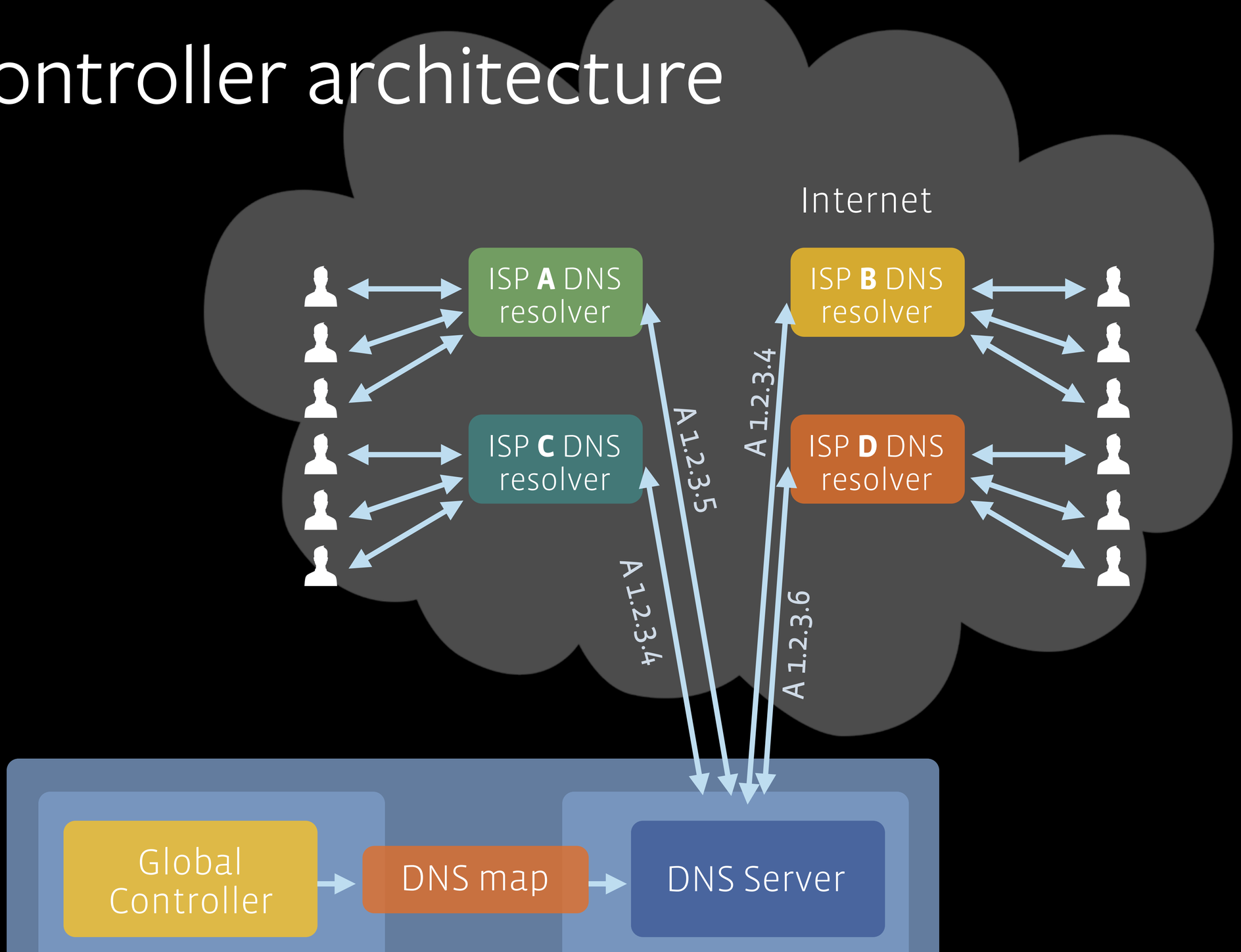
Sonar: Measuring “closeness”



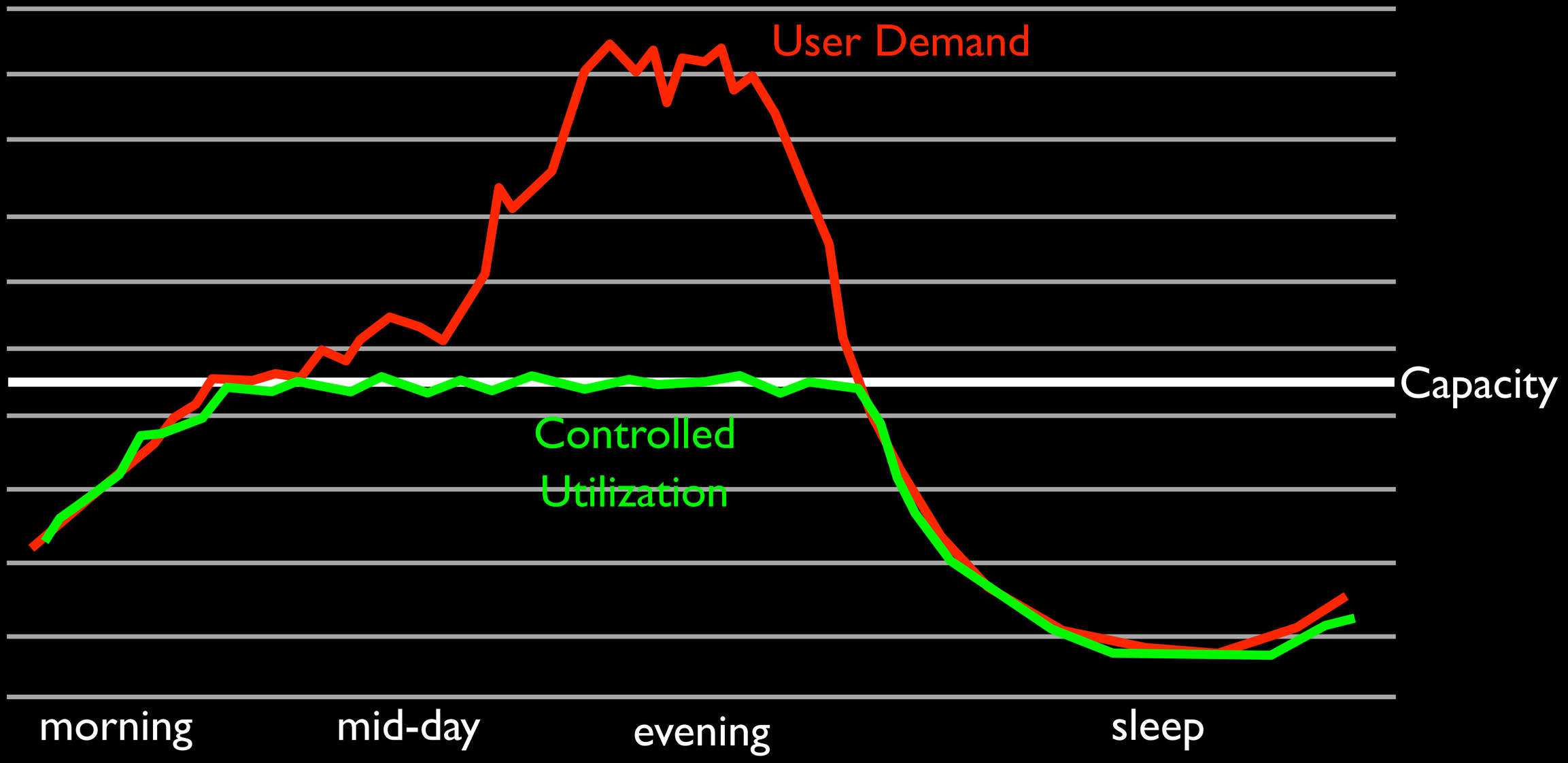
Global controller architecture



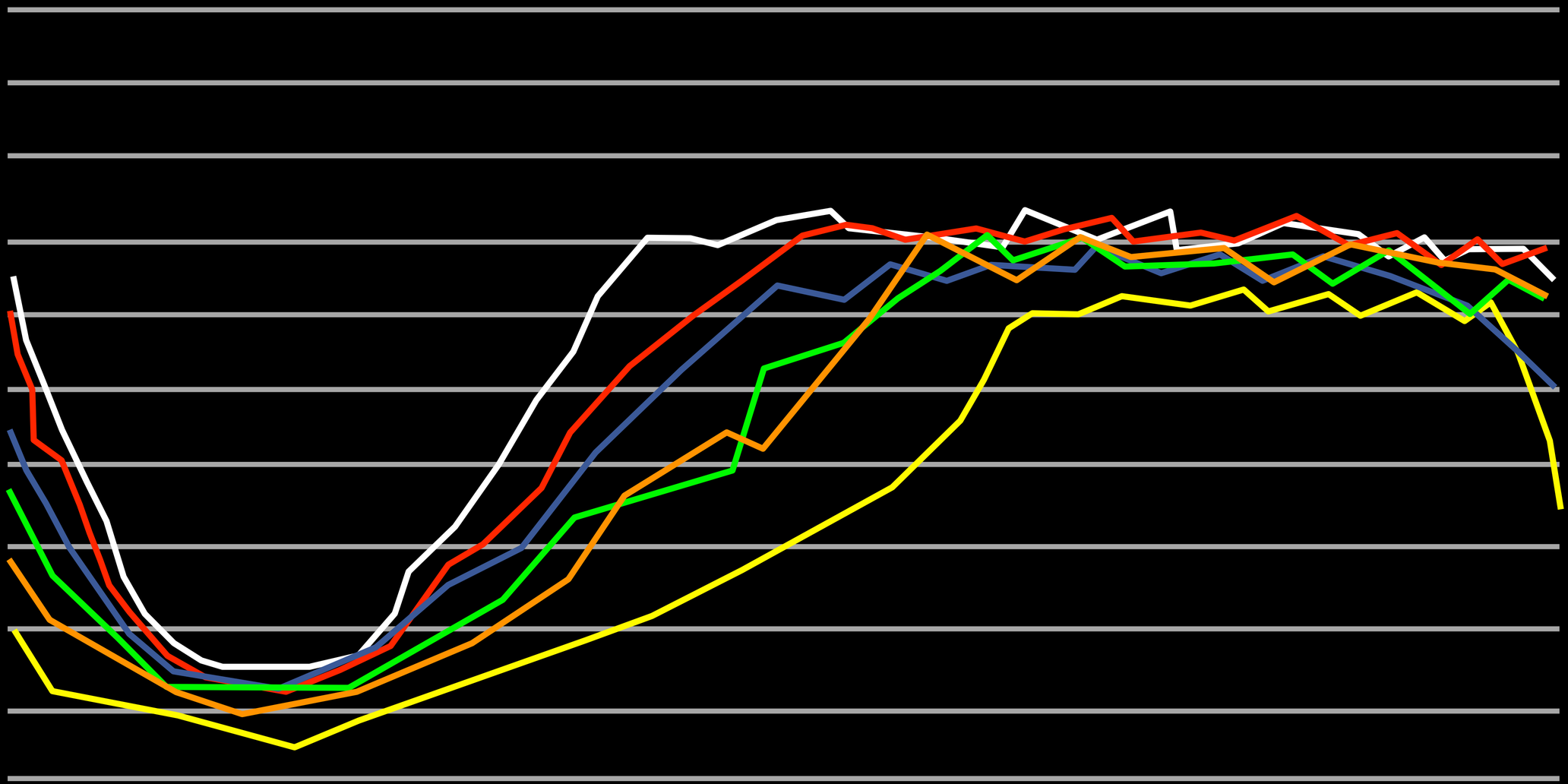
Global controller architecture



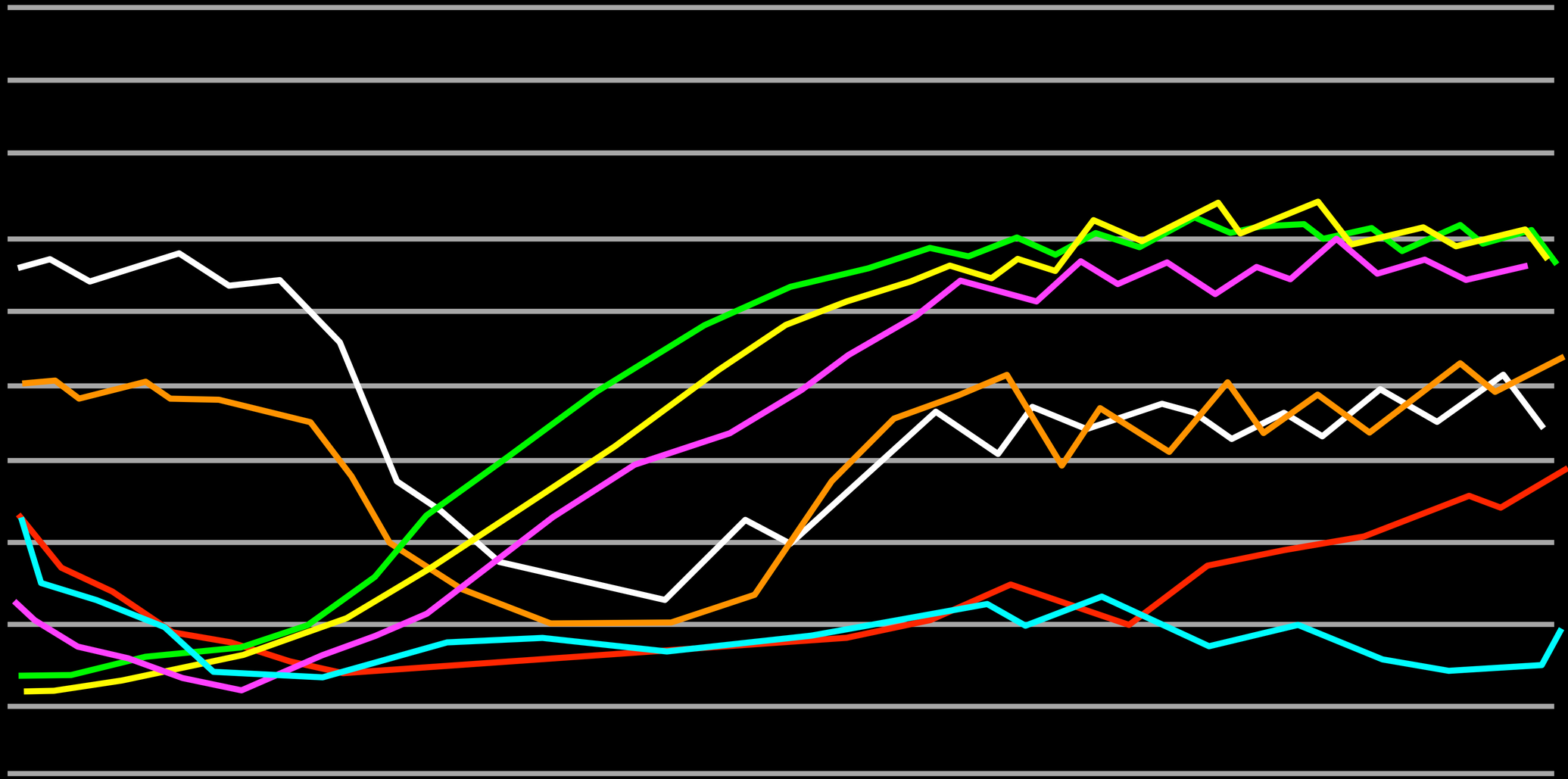
Global controller in action



Regional load shedding

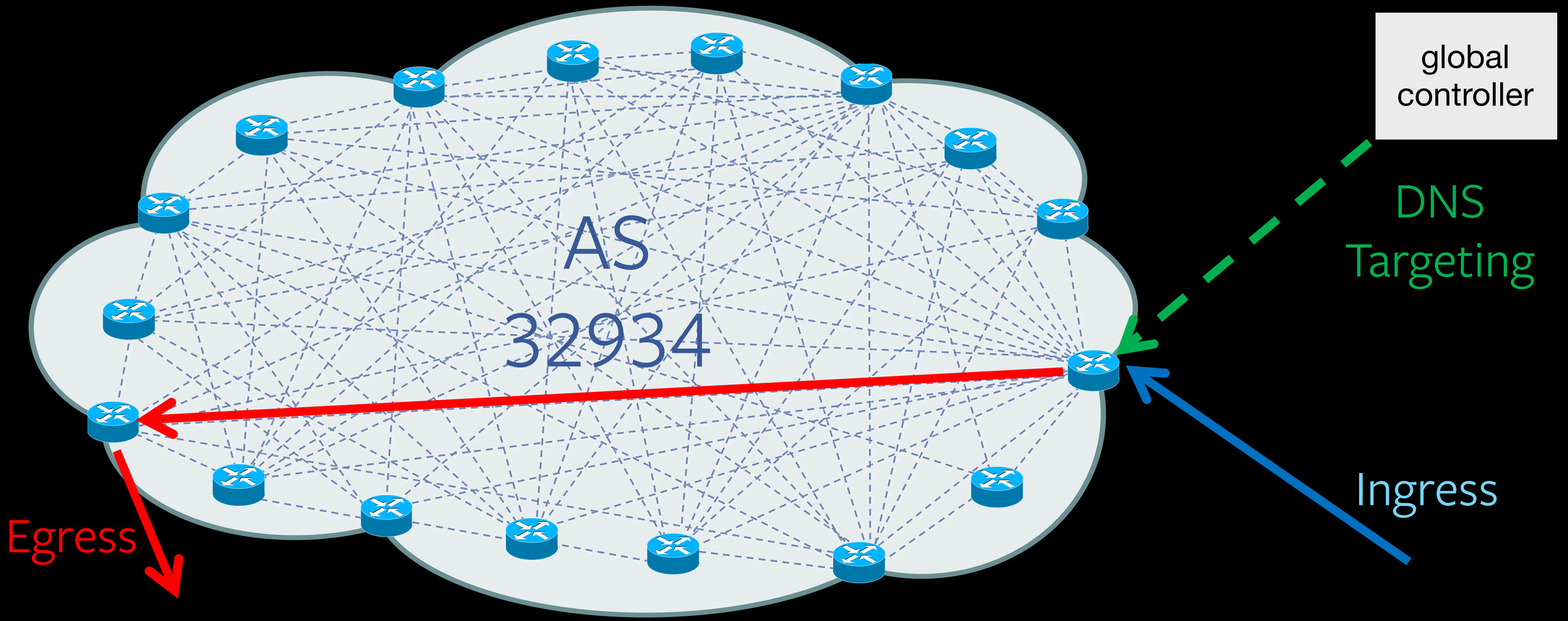


Global load shedding

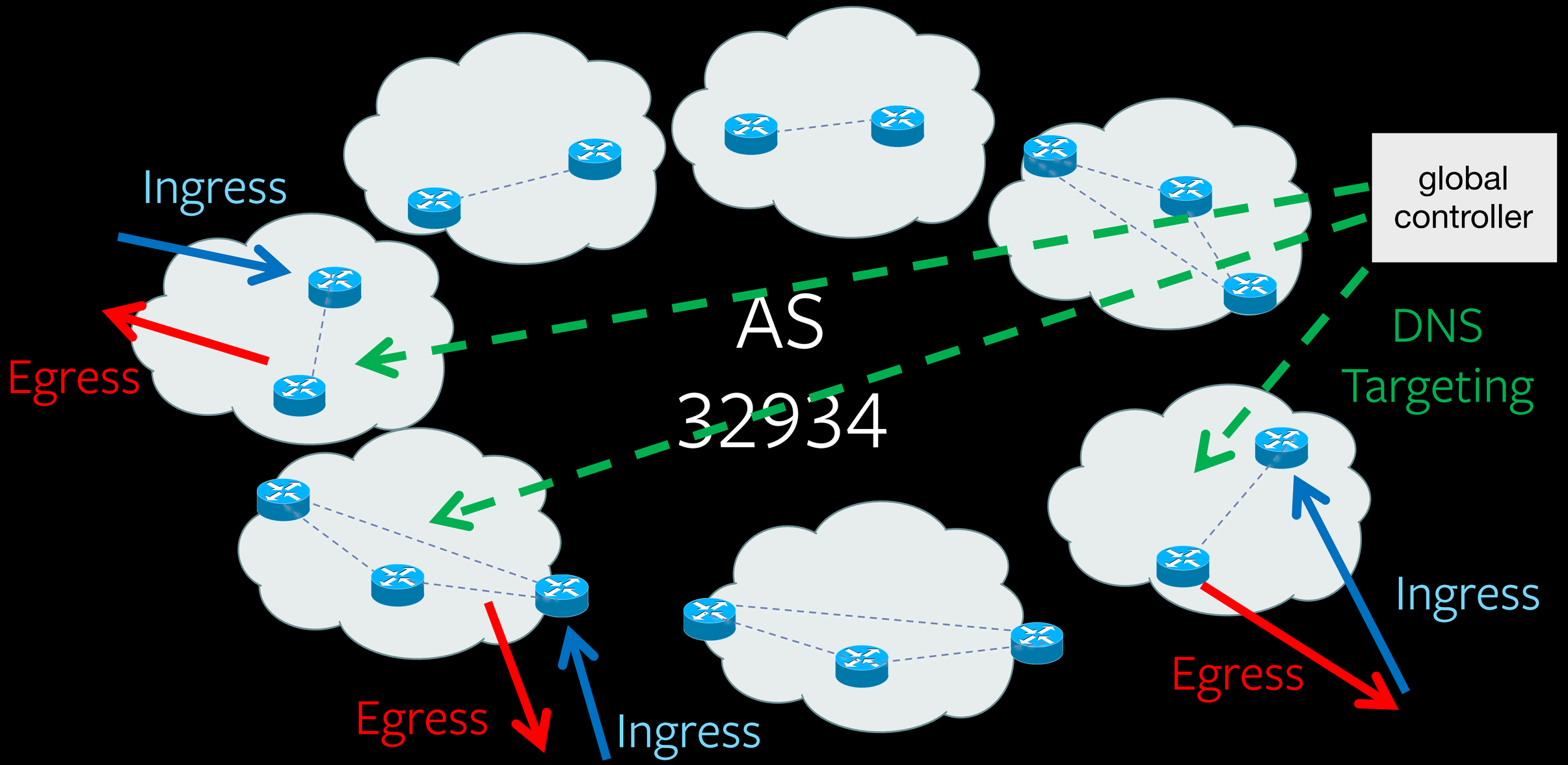




But what about BGP & targeting?



Metro-level BGP peering islands



Even on an island, BGP is limited

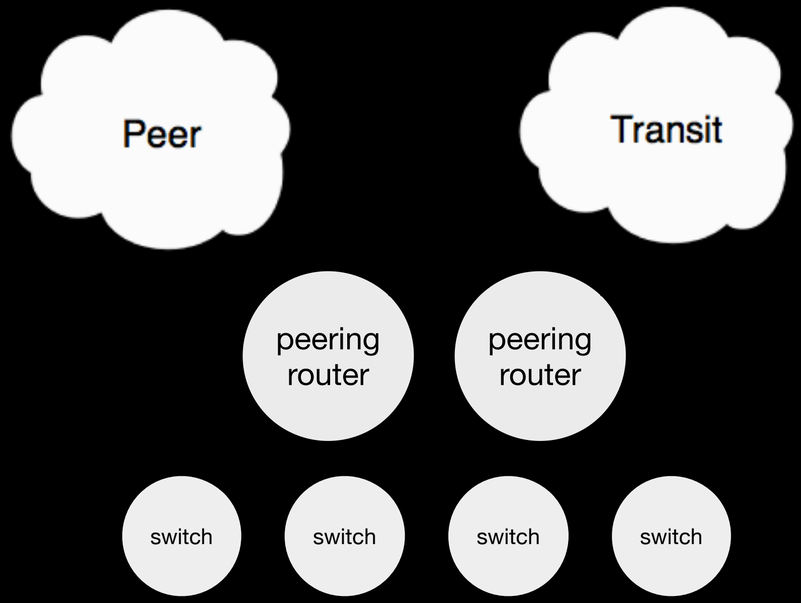
BGP cares about...	We care about...
1. longest prefix 2. local preference 3. AS path 4. MED	1. capacity 2. packet loss 3. latency 4. service performance

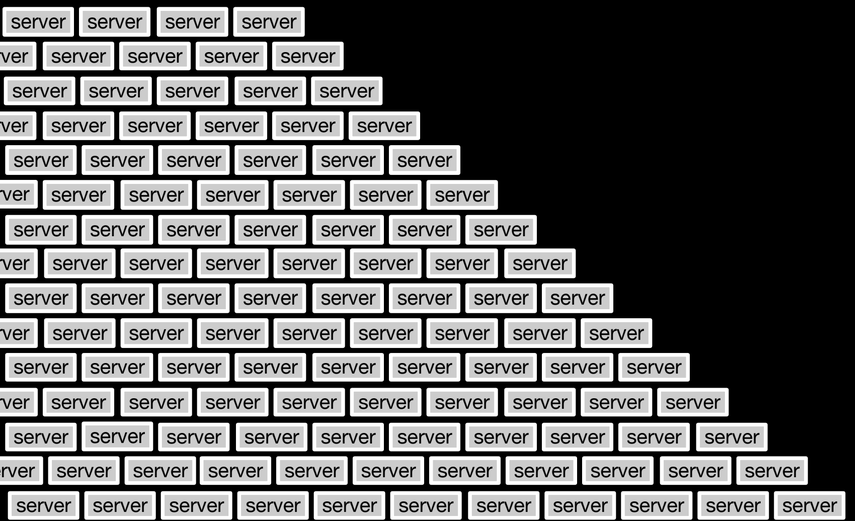
Local network controllers

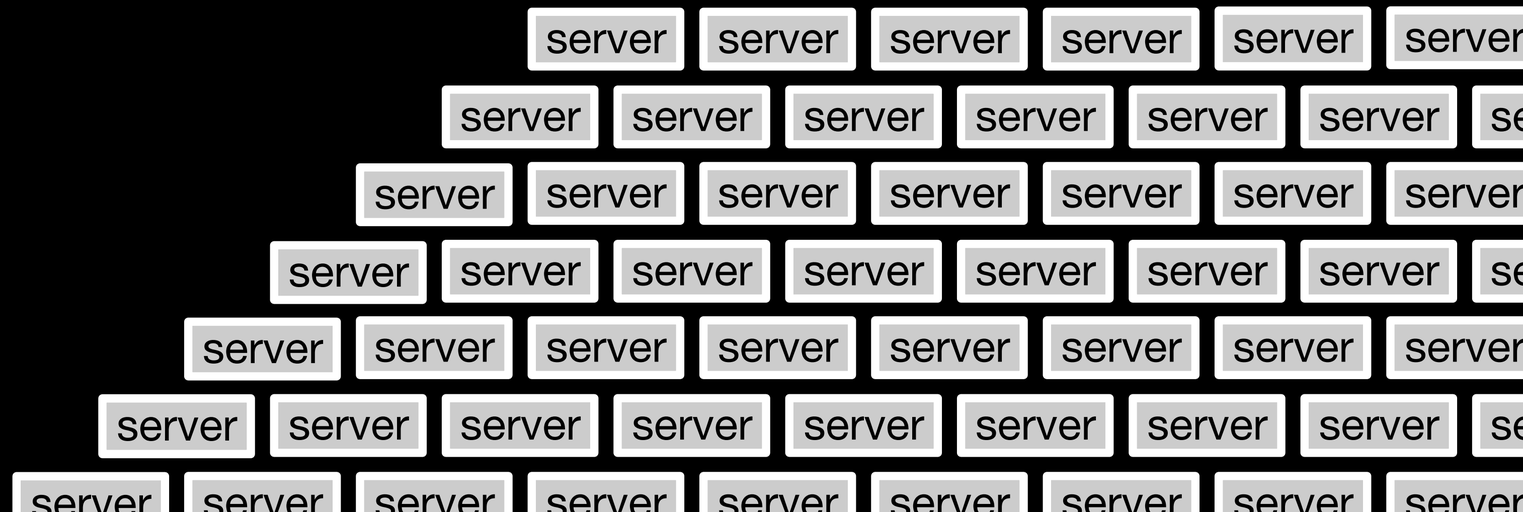
Host routing!

- Looks great in slides
- Not so simple in practice...
  - Synchronization across hosts
  - A *lot* of hosts
  - Non-local traffic sources
  - And which encapsulation?

Complete control!





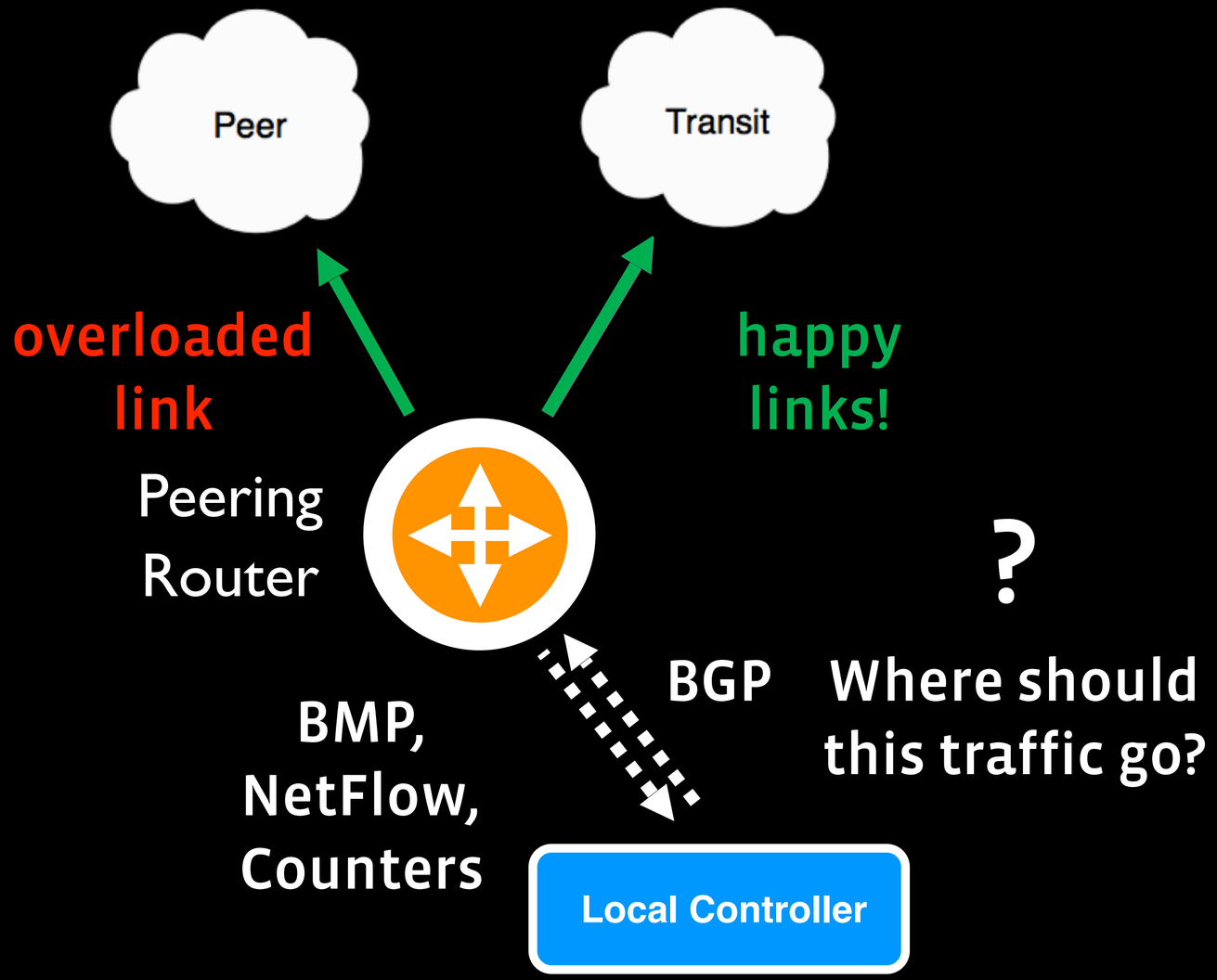


How do the servers encapsulate!?

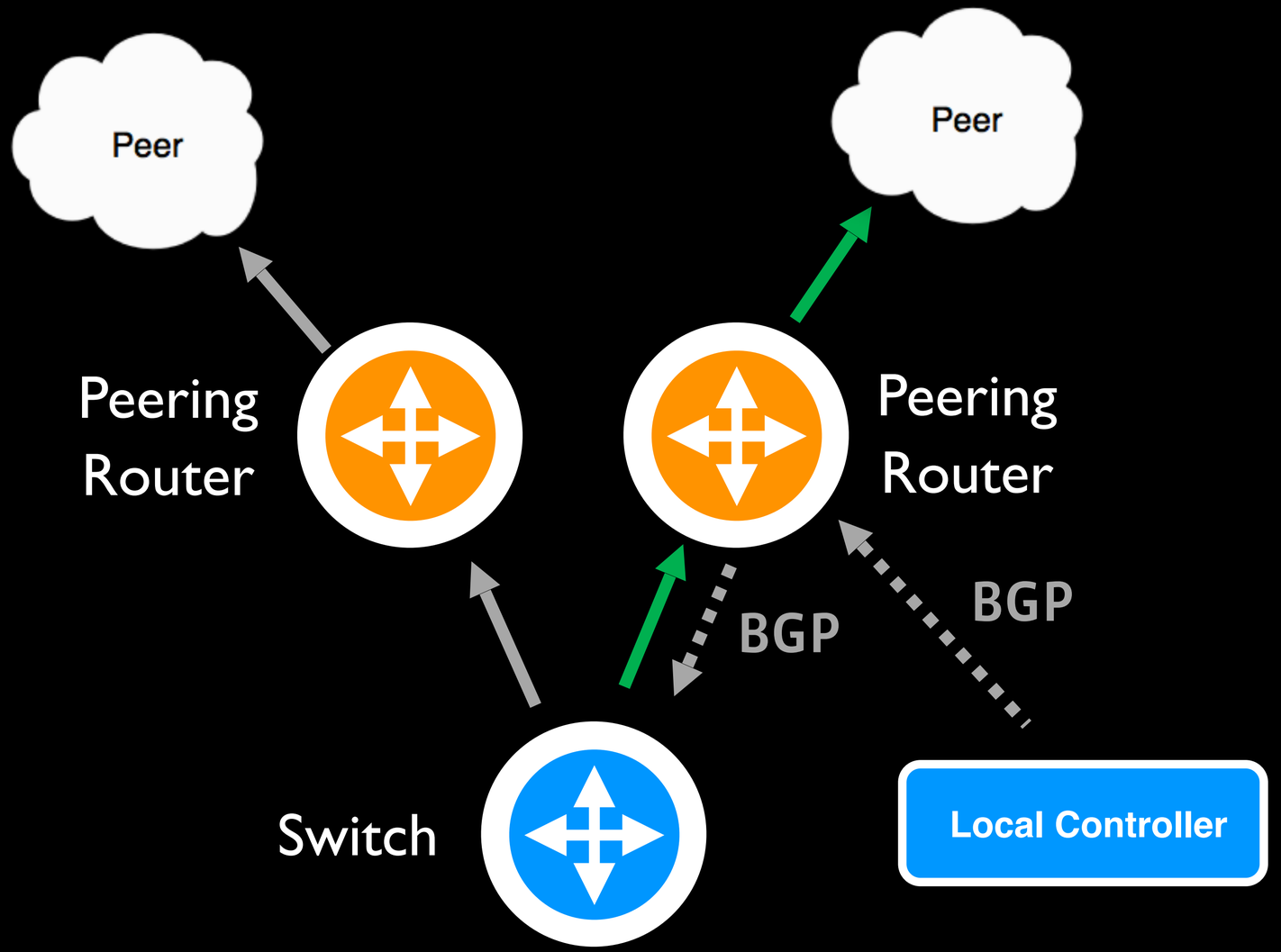
- MPLS – Kernel support problems, router features
- PBR - limits platform diversity, router configuration state
  - DSCP – very coarse range
  - GRE key – repeated router bugs

What we learned.... keep it simple

Evolving beyond BGP ...with BGP

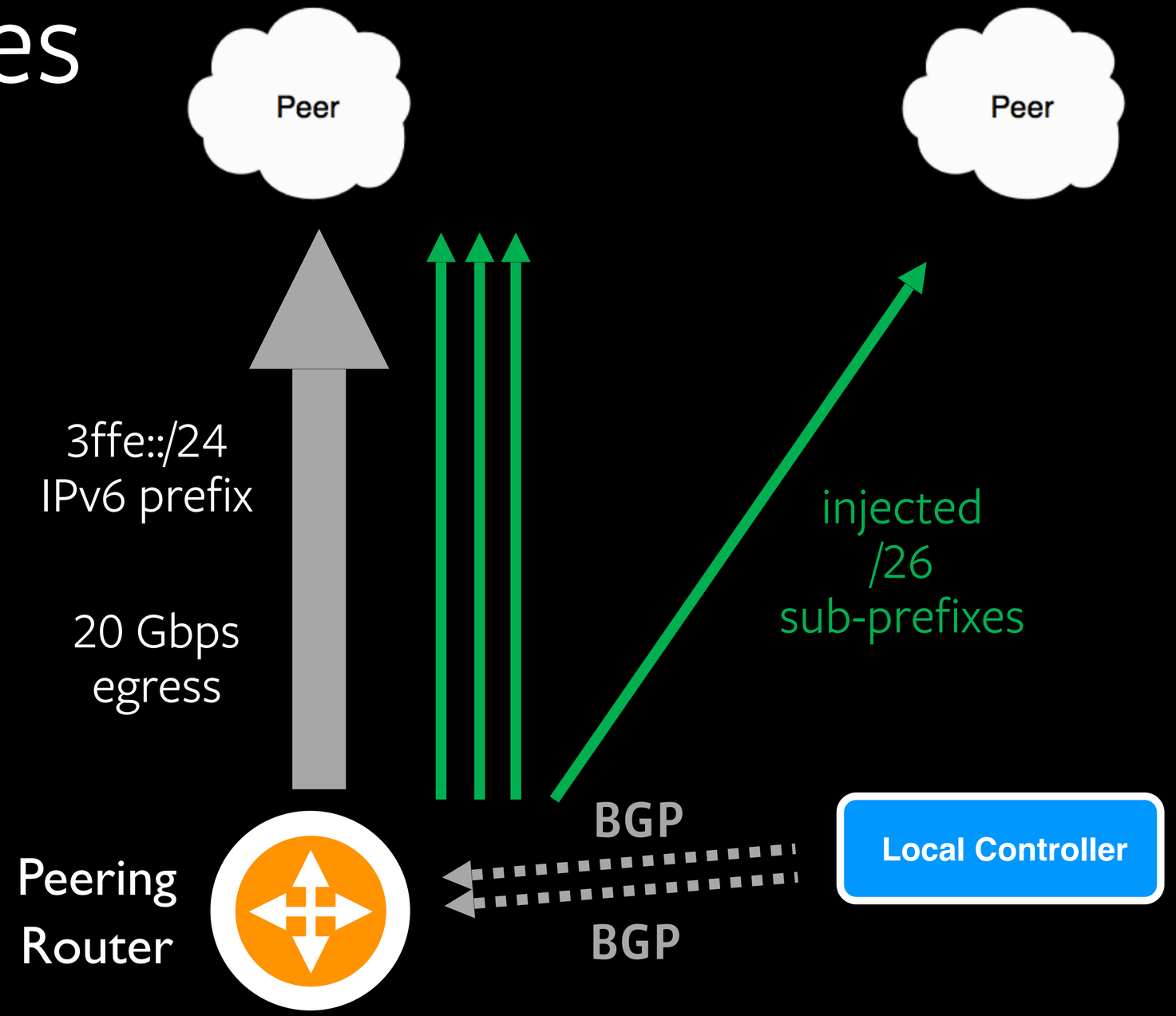


Metro Traffic Engineering



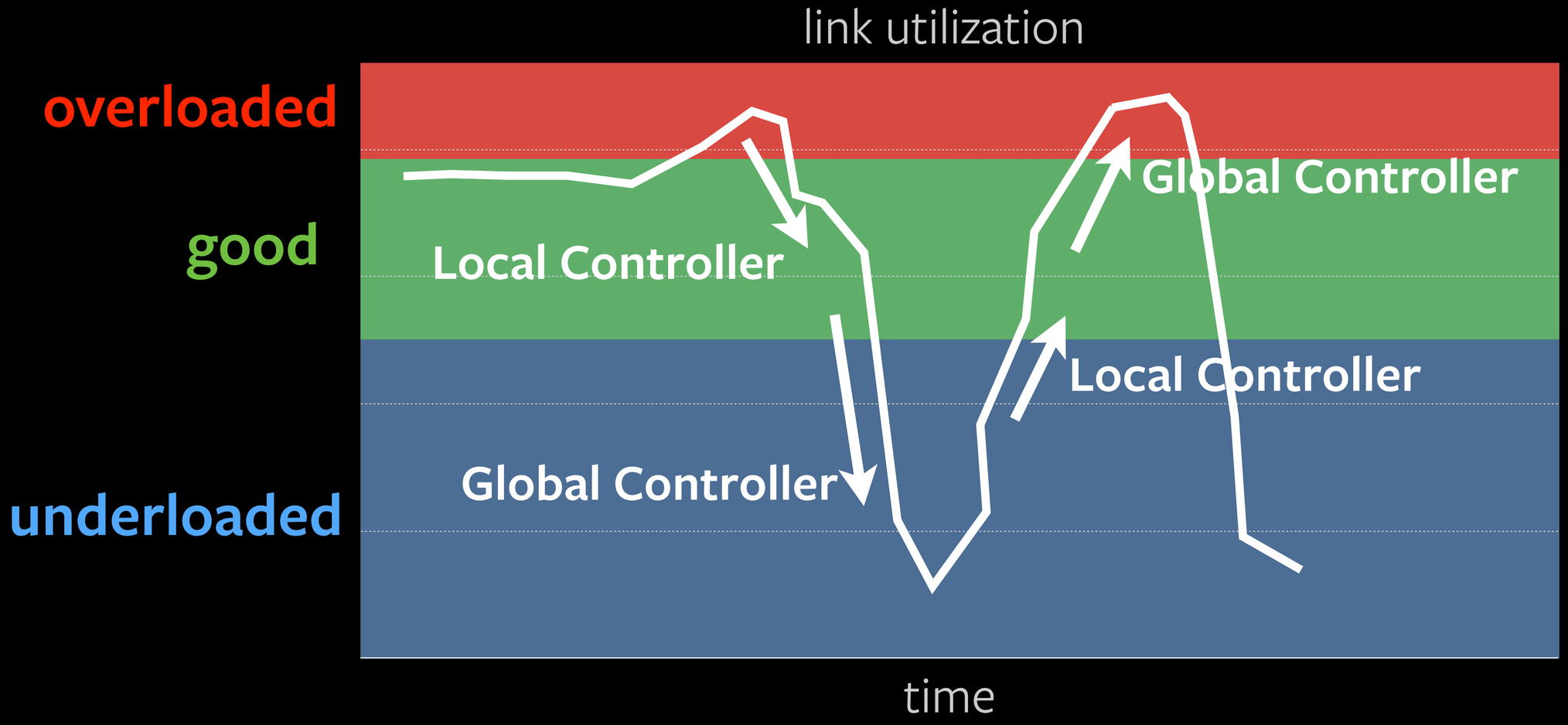
Awesome! So it’s all solved?

Huge Prefixes

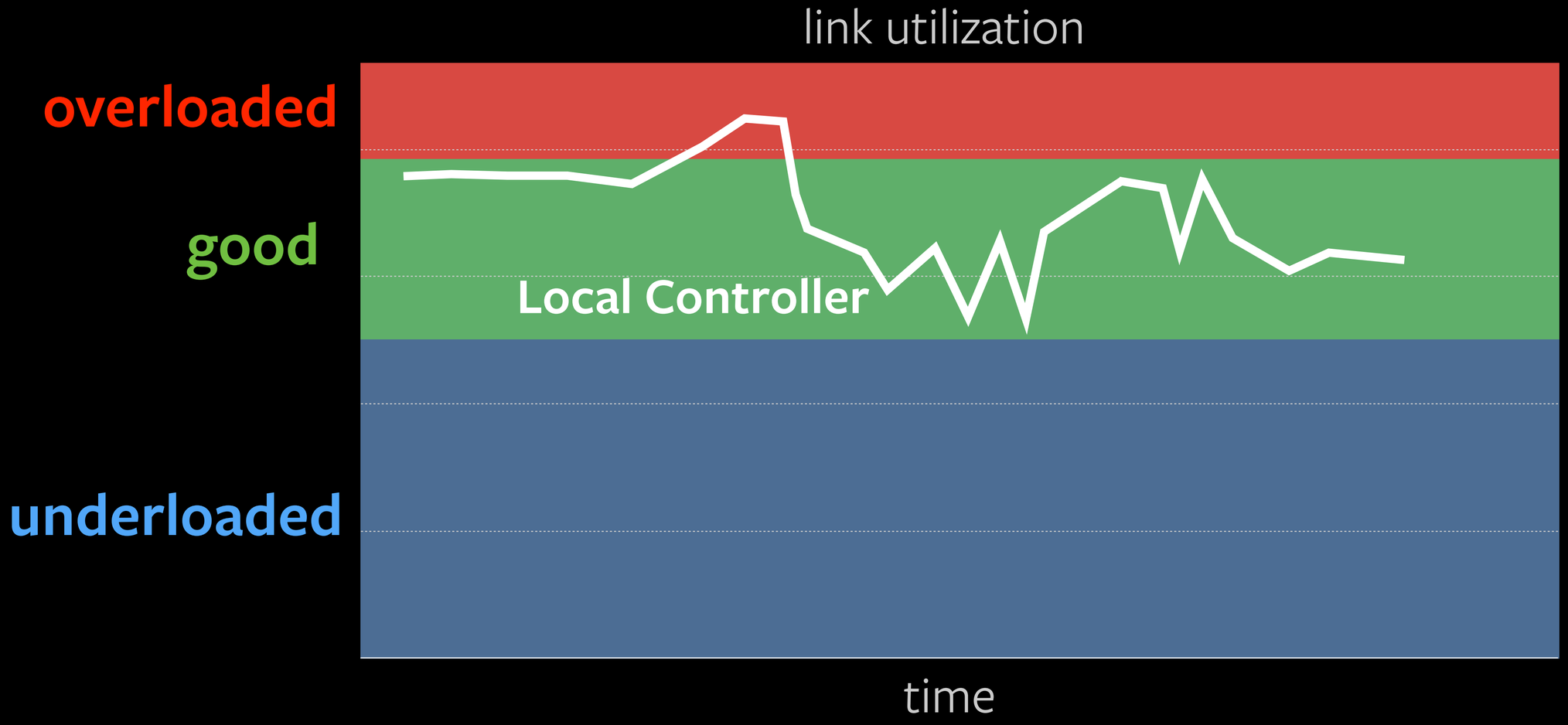


Oscillation between controllers

Local and Global Controller interaction



Clear roles so they don’t oscillate



Oscillation from imprecise data

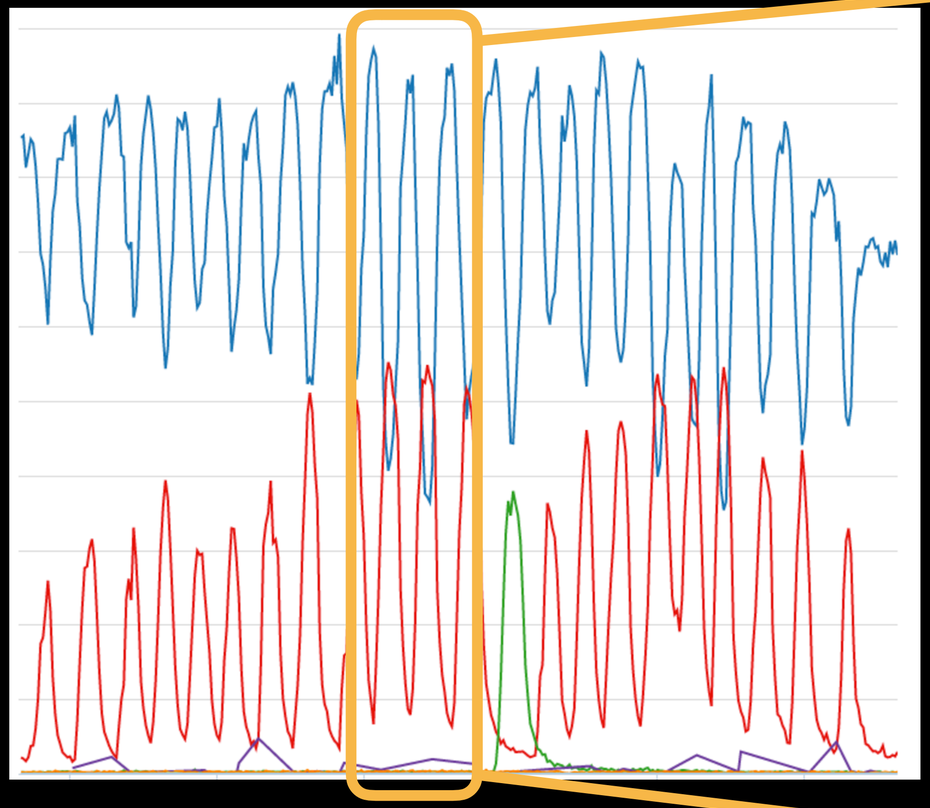
Lack of Precise Bandwidth Estimation

20 Gbps

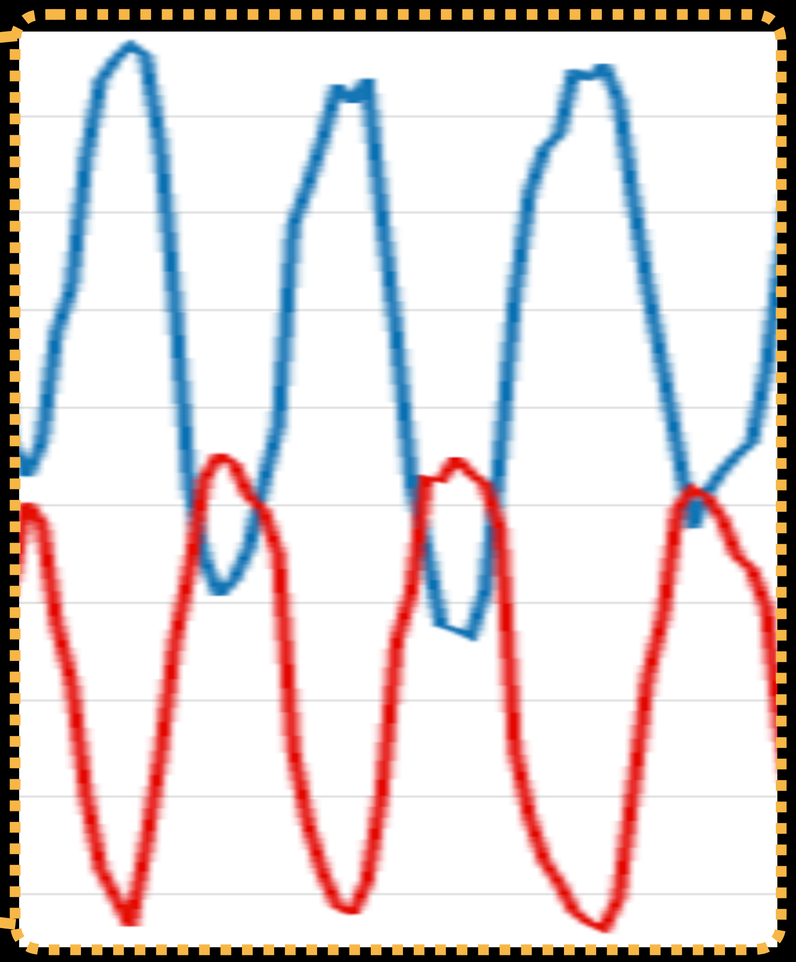


10 Gbps





w/o precision



Oscillation from imprecise data

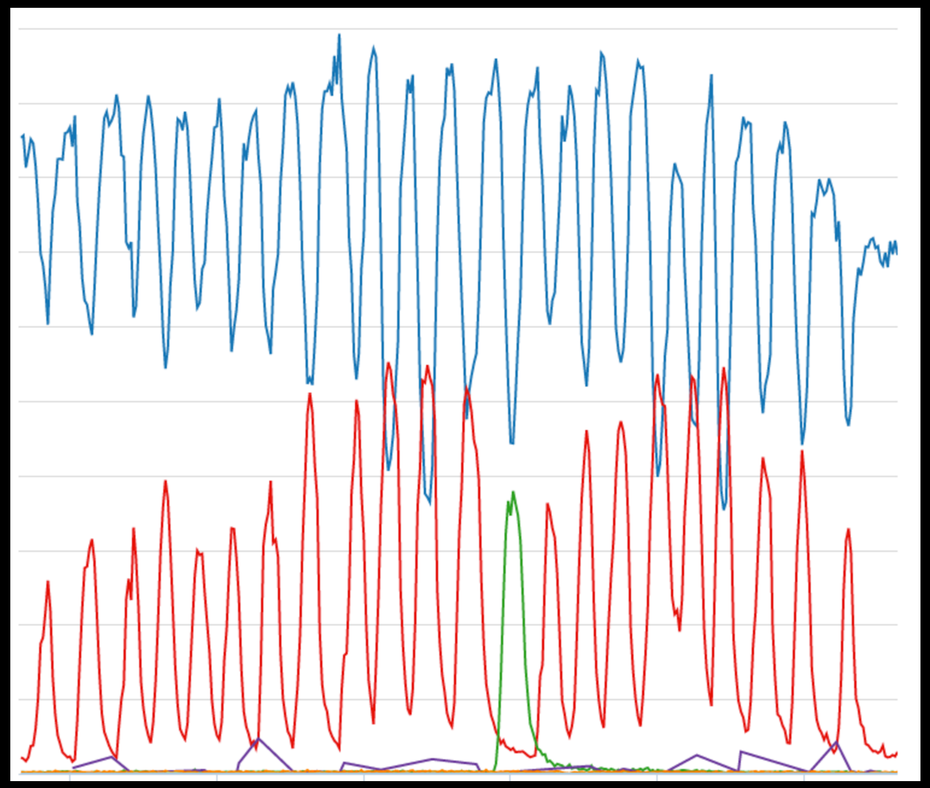
Lack of Precise Bandwidth Estimation

20 Gbps

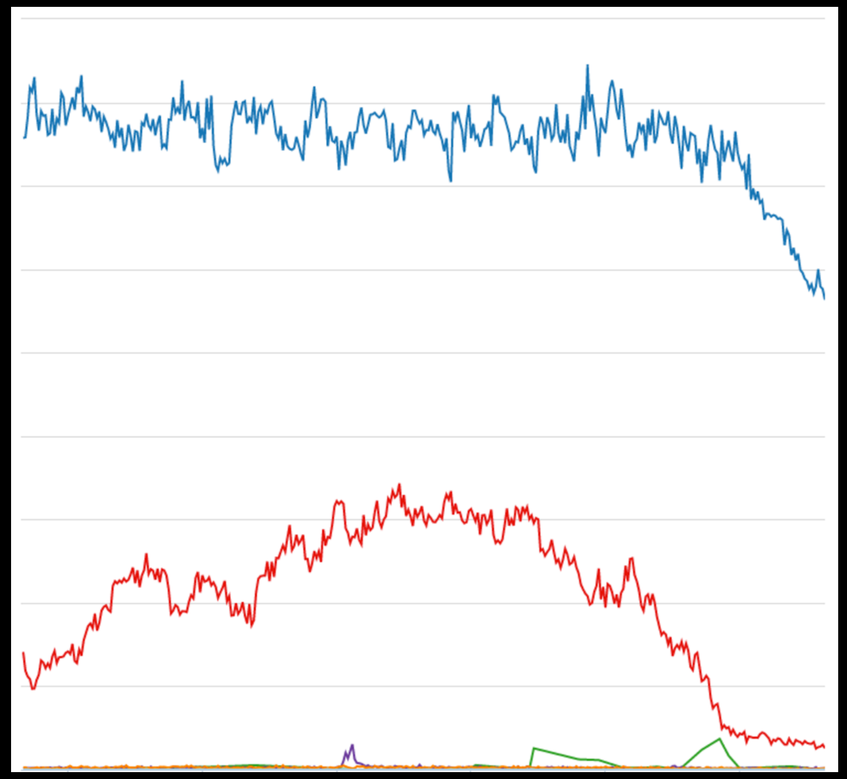


10 Gbps





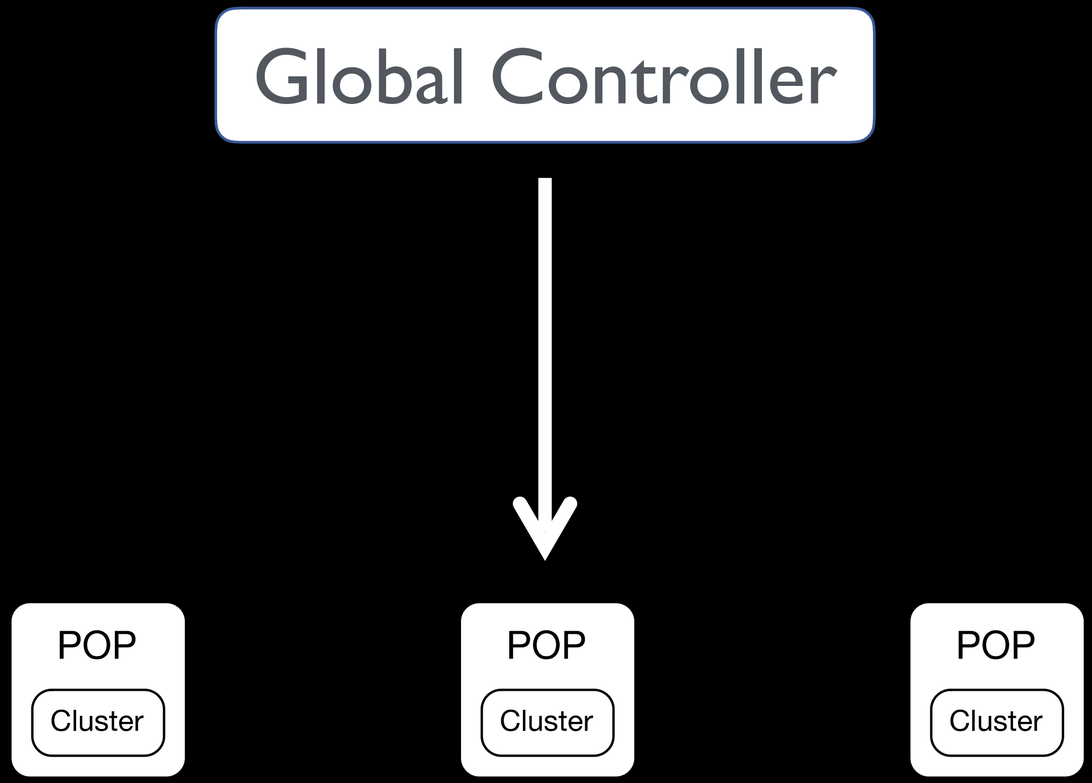
w/o precision



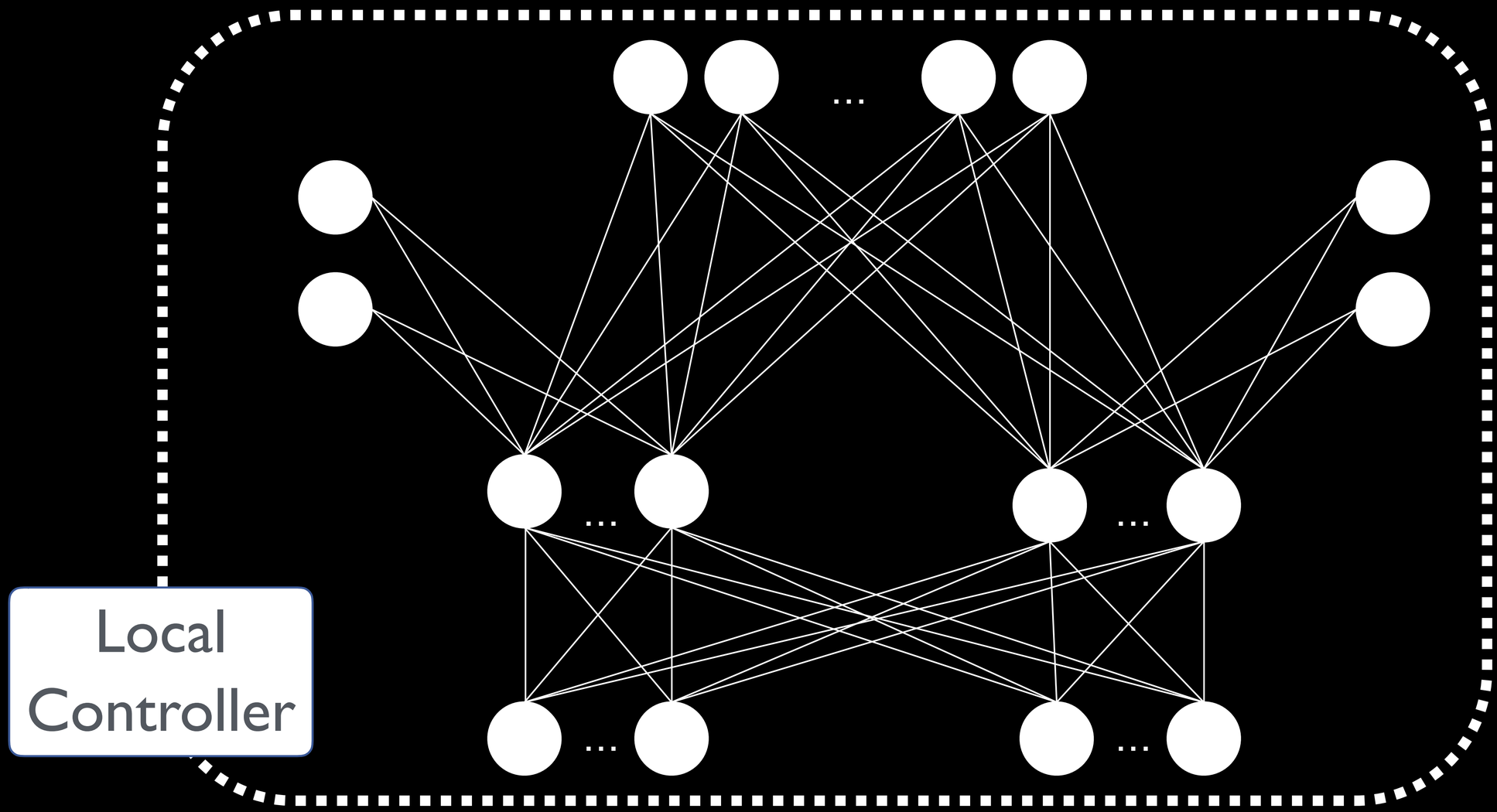
w/ precision

Confidential- Subject to NDA

All together...



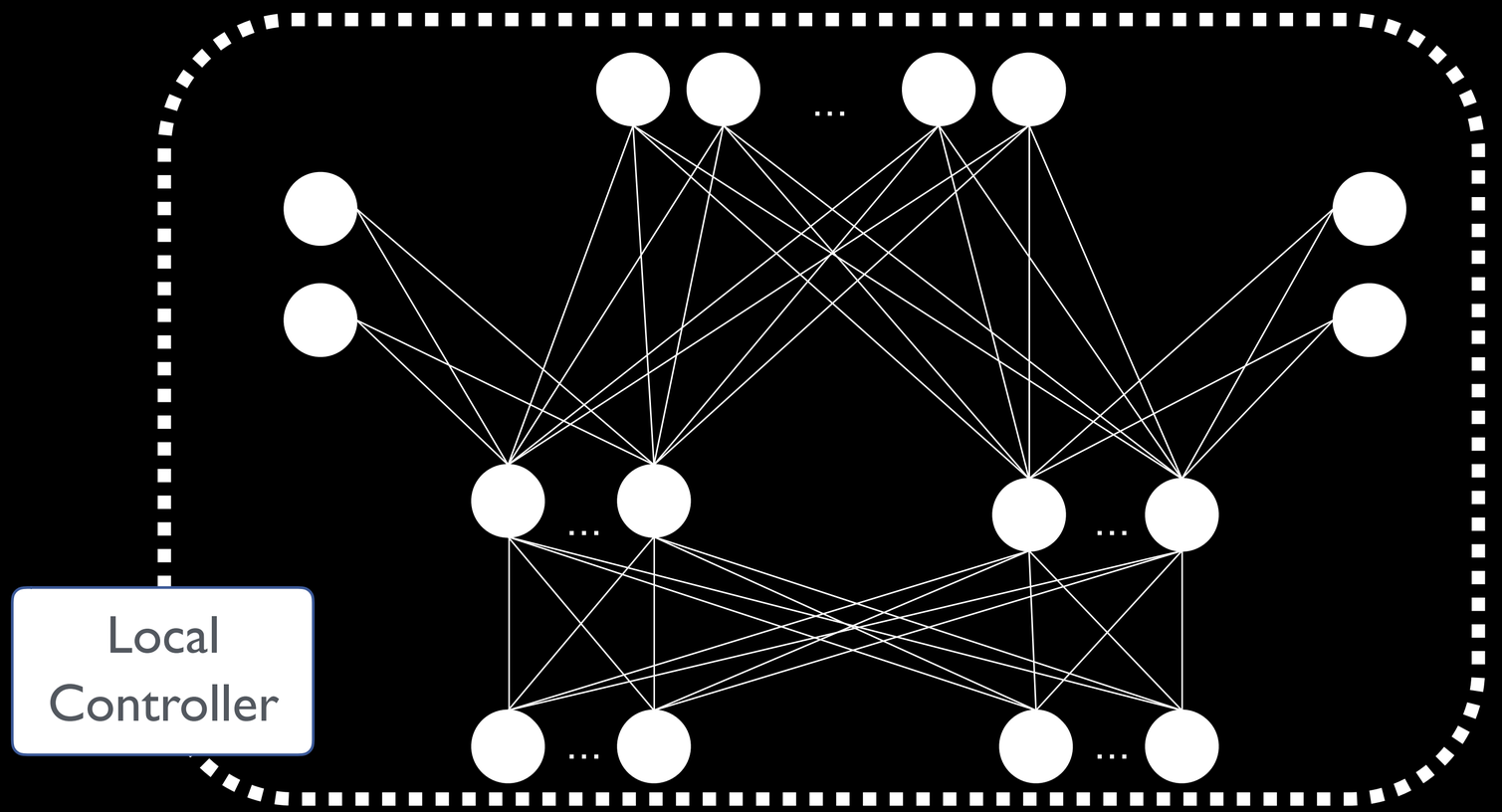
All together...



...and the future

Increasing:

- Modularity
- Programmability
- Performance



Summary

Questions?


facebook
INFRASTRUCTURE

Photo Credits

http://www.flickr.com/photos/27587002@N07/5170590074

http://www.flickr.com/photos/yaketyyakyak/7001664846

http://www.flickr.com/photos/hinnosaar/3778903507

http://www.flickr.com/photos/eamoncurry/8698726494

http://www.flickr.com/photos/43158397@N02/4514113429

http://www.flickr.com/photos/nobusue/6876280595

http://www.flickr.com/photos/29487672@N07/14760573314

http://www.flickr.com/photos/joyosity/3595242078

http://www.flickr.com/photos/kyntharyn74/3262089319

http://www.flickr.com/photos/rexipe/826987087

http://www.flickr.com/photos/lablasco/6815671096