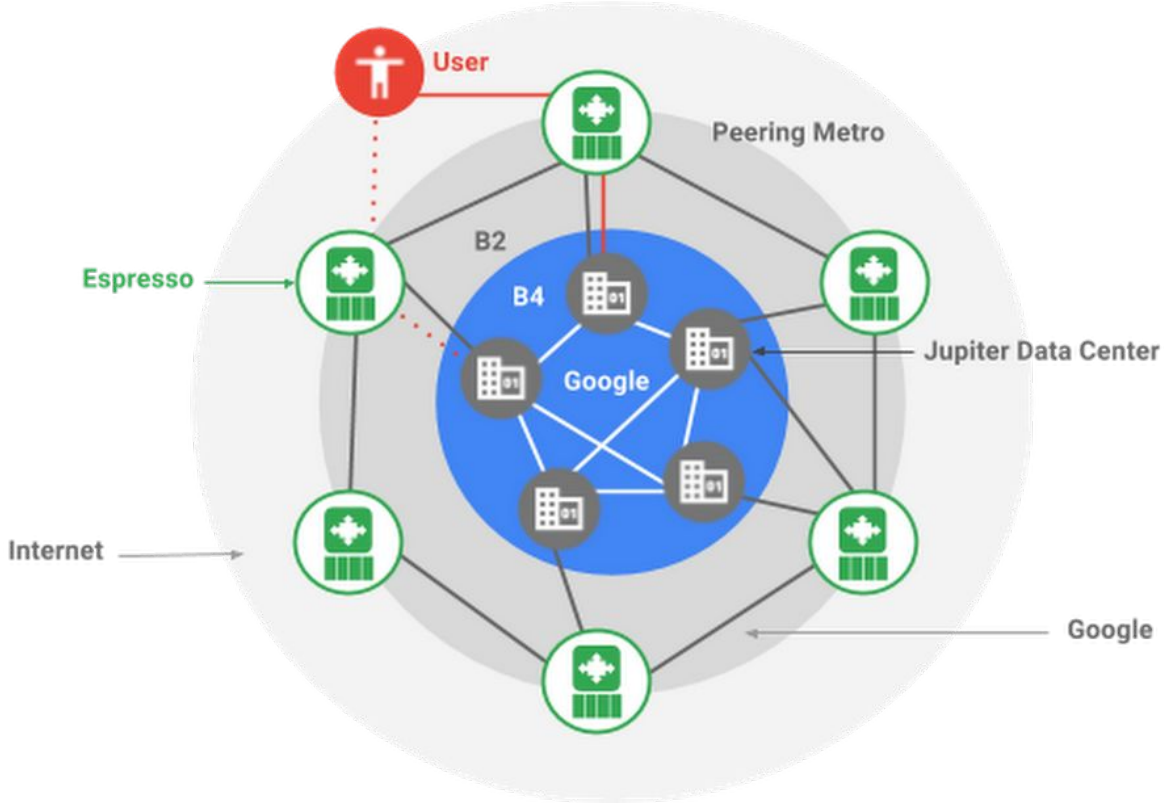
Google

Buffer sizing experiments at Google

Lincoln Dale

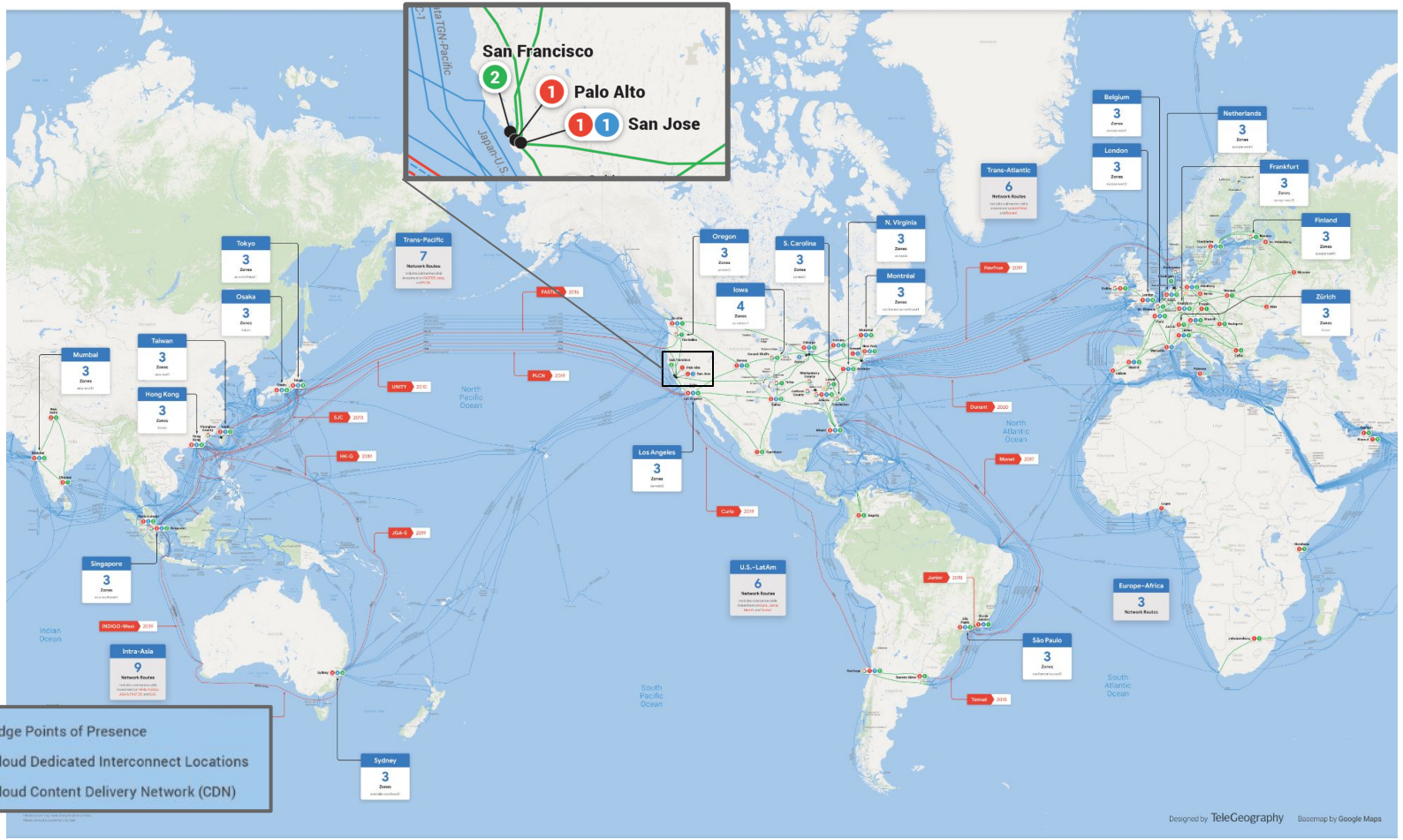
With Amin Vahdat, Yuchung Chen, Christophe Diot

Google’s Networks



Google Cloud

2018



1 Google Edge Points of Presence

1 Google Cloud Dedicated Interconnect Locations

1 Google Cloud Content Delivery Network (CDN)

Google

Status

B2 is our customer facing network built with vendor hardware

- “Default” buffer size available on these routers

We do a bunch of things to "be nice" to the internet

- Pace traffic
- Modern TCP implementation with good congestion behavior (BBR)
- Avoid NIC Offloads like TSO don’t work across WAN

Default buffer utilization approach

- Historically many routers had large buffers by virtue of requirement for multiple DRAM channels
- We deliberate limit it to smaller amounts

Experiment

Edge caches (10 & 100G ports) send video streams

Measurements made on peering routers

1. Queue-length monitoring on larger buffer devices
    - Gives us the largest queue per pipeline that exceed configured threshold (all others, if any, not logged)
2. Tune max buffer size down until we start to see drops

Buffer size, drop rates are later correlated with logs and network performance

Observations

From the experiments, 10ms buffer should be sufficient for all our workloads

We could get by with between 5-10ms on not more than 10% of the queues

Ultimately, buffer size is a function of how Google manages traffic delivery outbound

- Network configuration
- Edge caches management
- Congestion control
- Video coding

Note that in all cases, we don’t have control of one or more endpoints - e.g. cloud

Google

BACK-UP

Xkcd’s view of Google Networks

https://xkcd.com/2105/
(we had no input into this)