

Comparing the Network Performance of AWS, Azure, GCP, IBM Cloud, and Alibaba Cloud

Angelique Medina

Cloud Research Journey



2018

- Microsoft Azure
- AWS
- Google Cloud



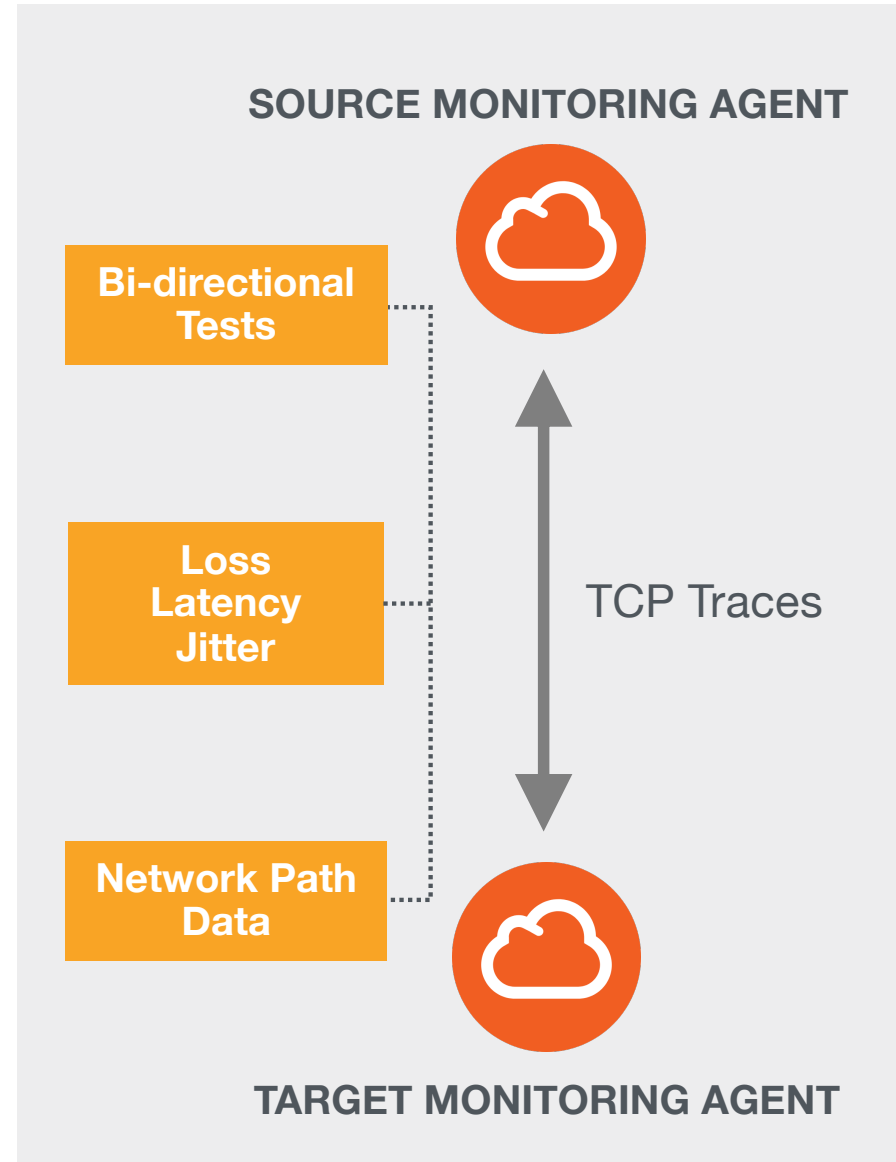
2019

- YoY changes (Azure, AWS, GCP)
- AWS Global Accelerator
- Broadband providers
- China
- Alibaba Cloud
- IBM Cloud

Research Methodology

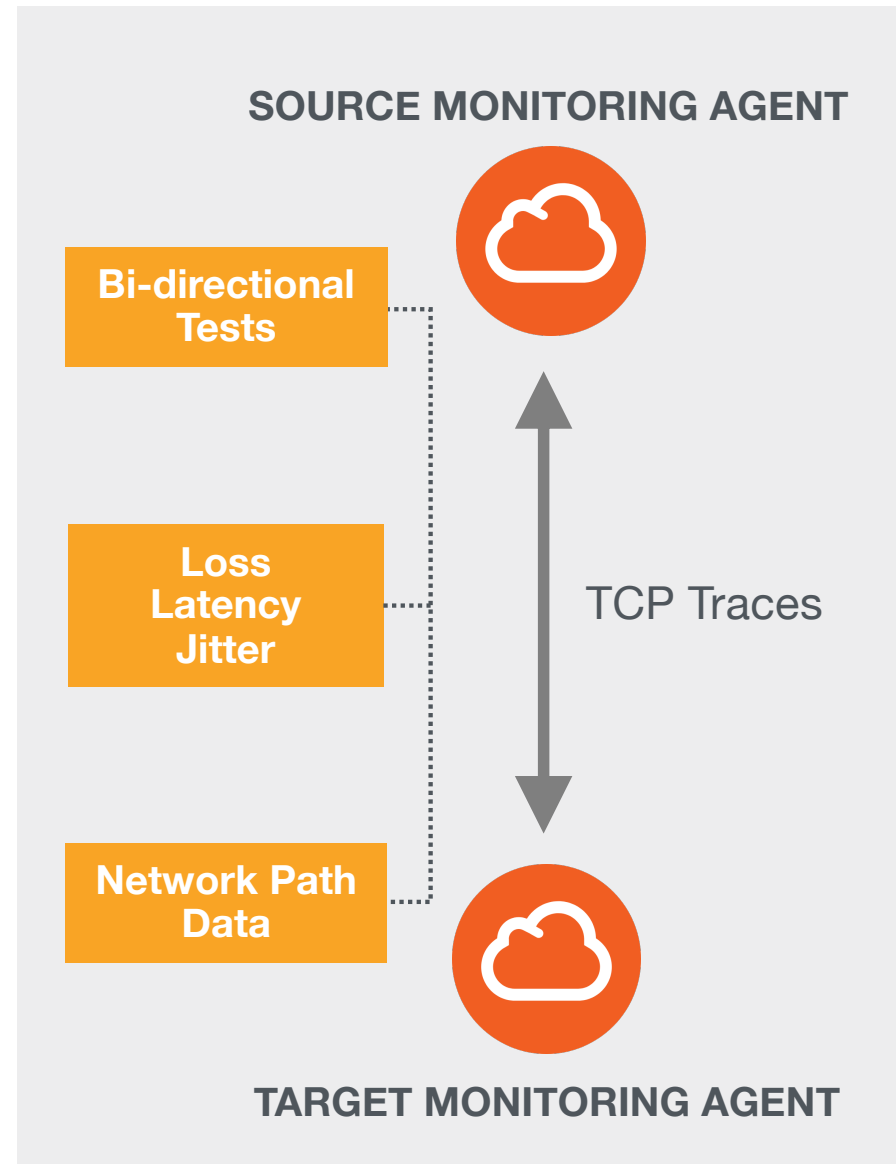
Data Collection

- Customized traceroute
 - TCP-based
- Bidirectional measurements
 - Accounts for variation in forward and return routes
 - Unidirectional metrics computed at target agent
- Network path
 - Layer 3 per-hop AS path, metrics, QoS re-markings
- End-to-end network metrics
 - Latency, loss, jitter

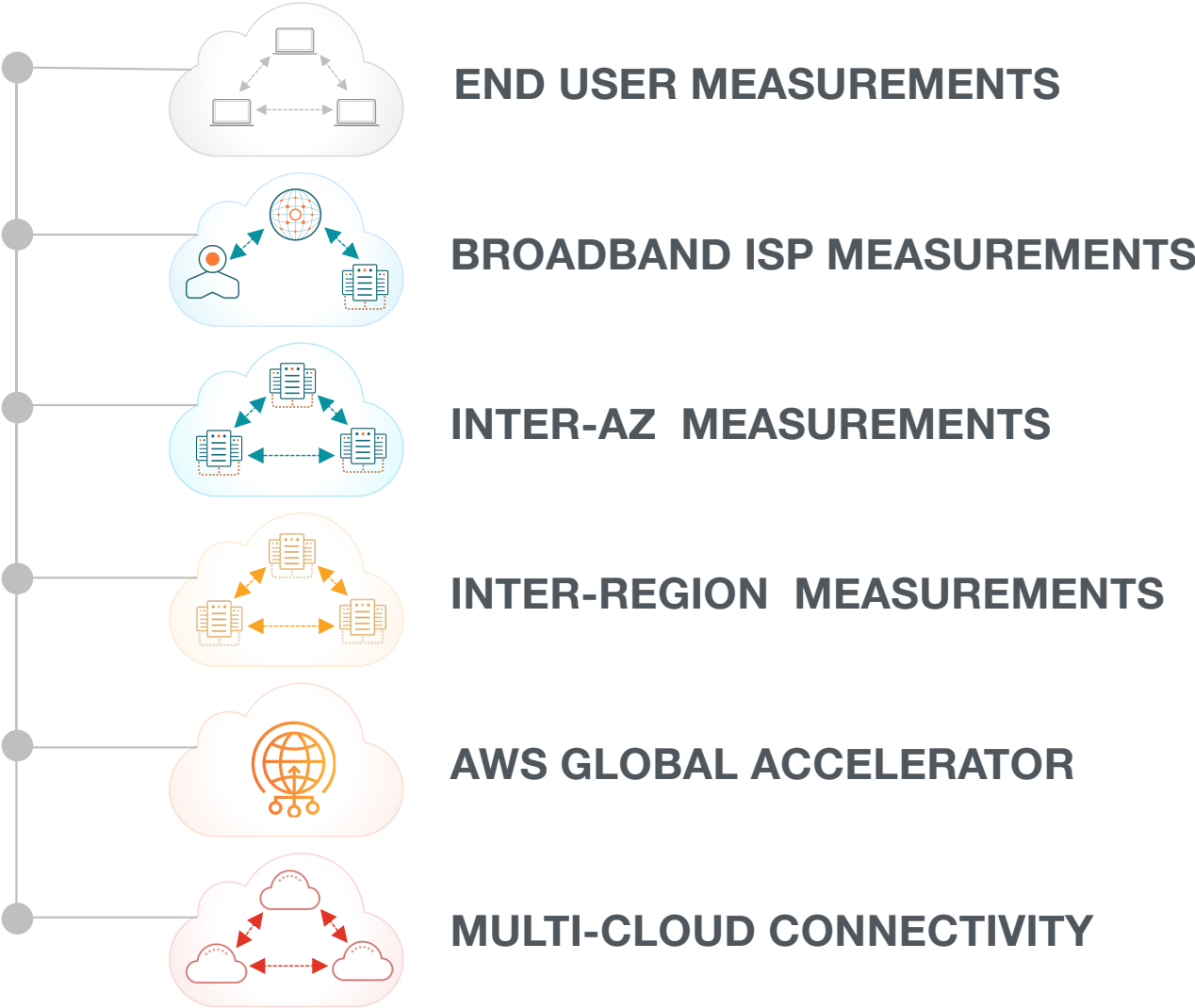


Data Processing

- Managed vantage points
 - Consistent measurement
 - Hosted in data centers
- 10 minute testing interval
- 30 day data collection
 - Data collection spans 09.01.2019 - 10.30.2019
 - Zero outages occurred during the data collection timeframe
 - Loss and Jitter negligible
 - Latency differences most prominent



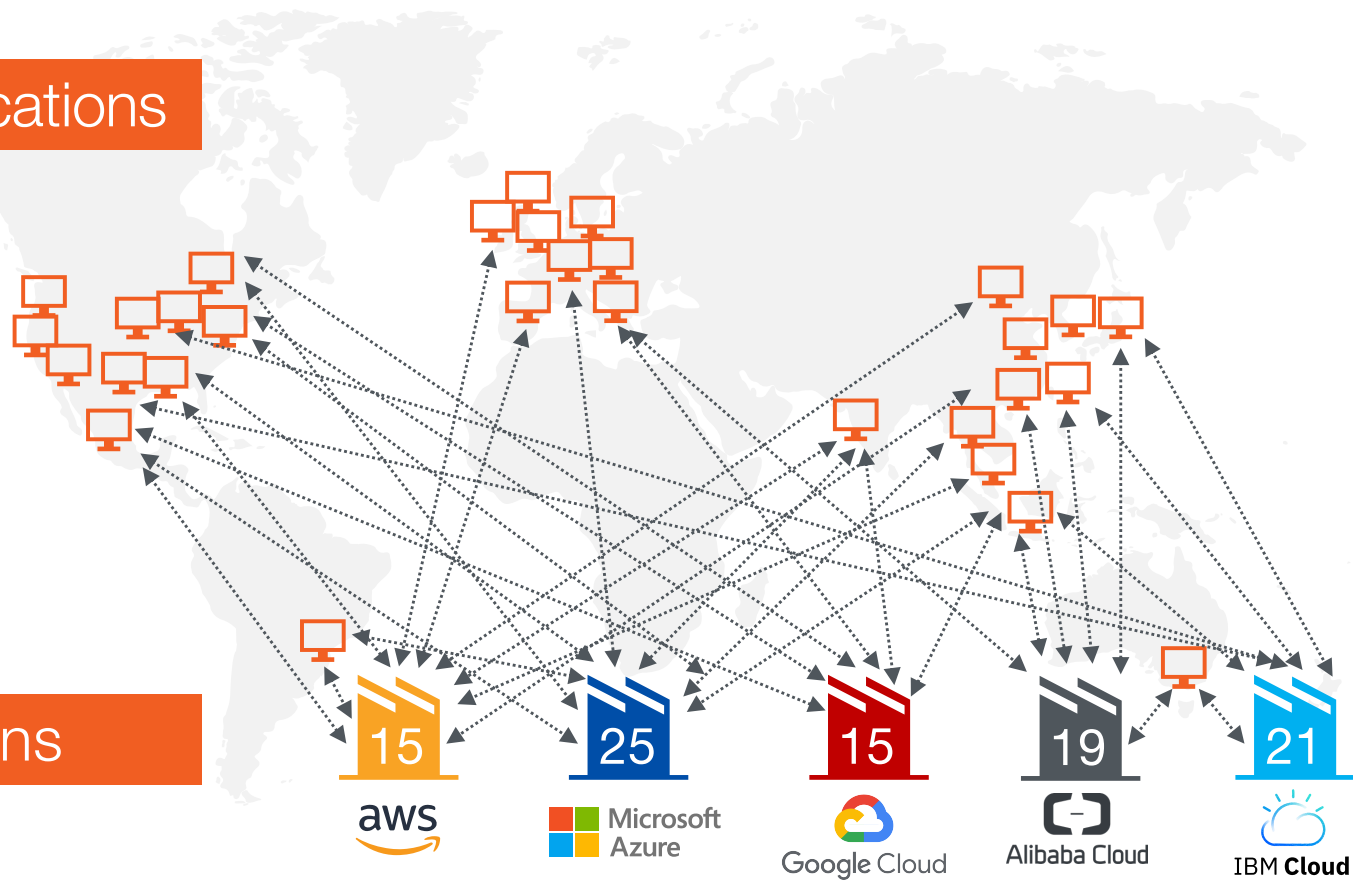
Research Scope



End User Measurements

98 Global Locations

95 Regions



Broadband Measurements



Ashburn, VA

US East



Chicago, IL

US Central



Dallas, TX

US West



Los Angeles, CA

US West



San Jose, CA

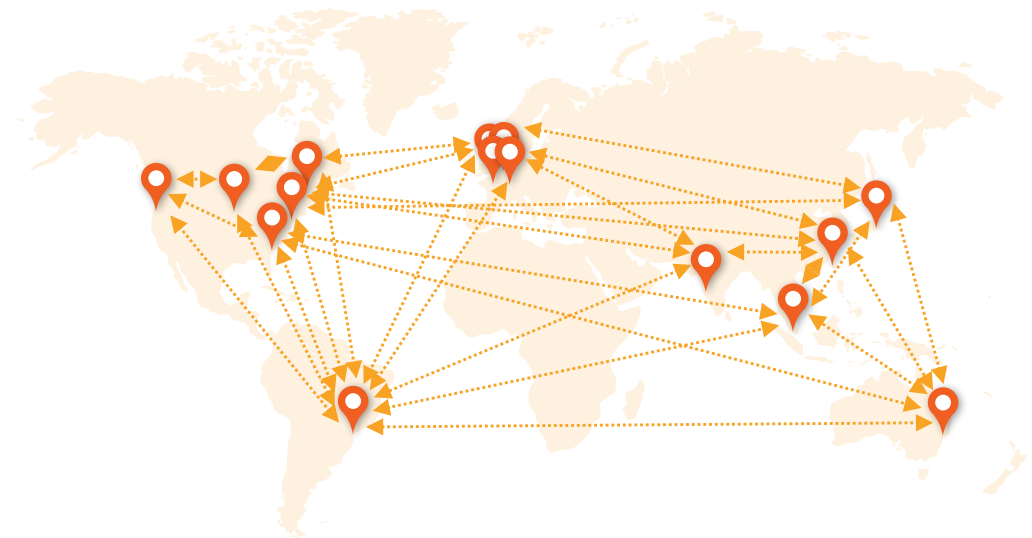
Canada



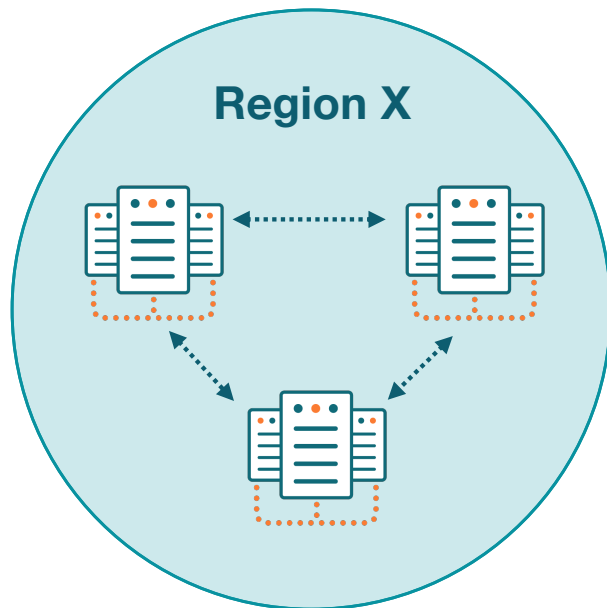
Seattle, WA

Inter-Region Measurements

- Inter-region performance measured per-provider
- Provider performance compared against baseline latency metrics derived from vantage points proximal but external to each cloud provider



Intra-Region (Inter-AZ) Measurements



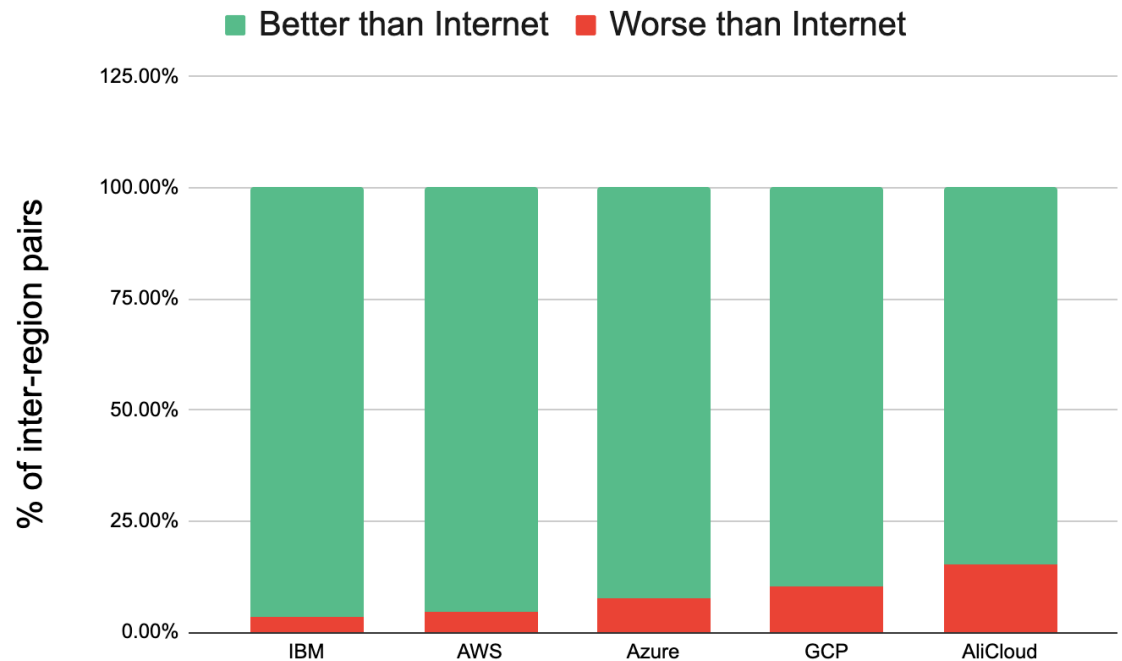
- AWS: 6 regions
- Azure: 6 regions
- GCP : 6 regions
- Alibaba Cloud: 7 regions
- IBM Cloud: 4 regions



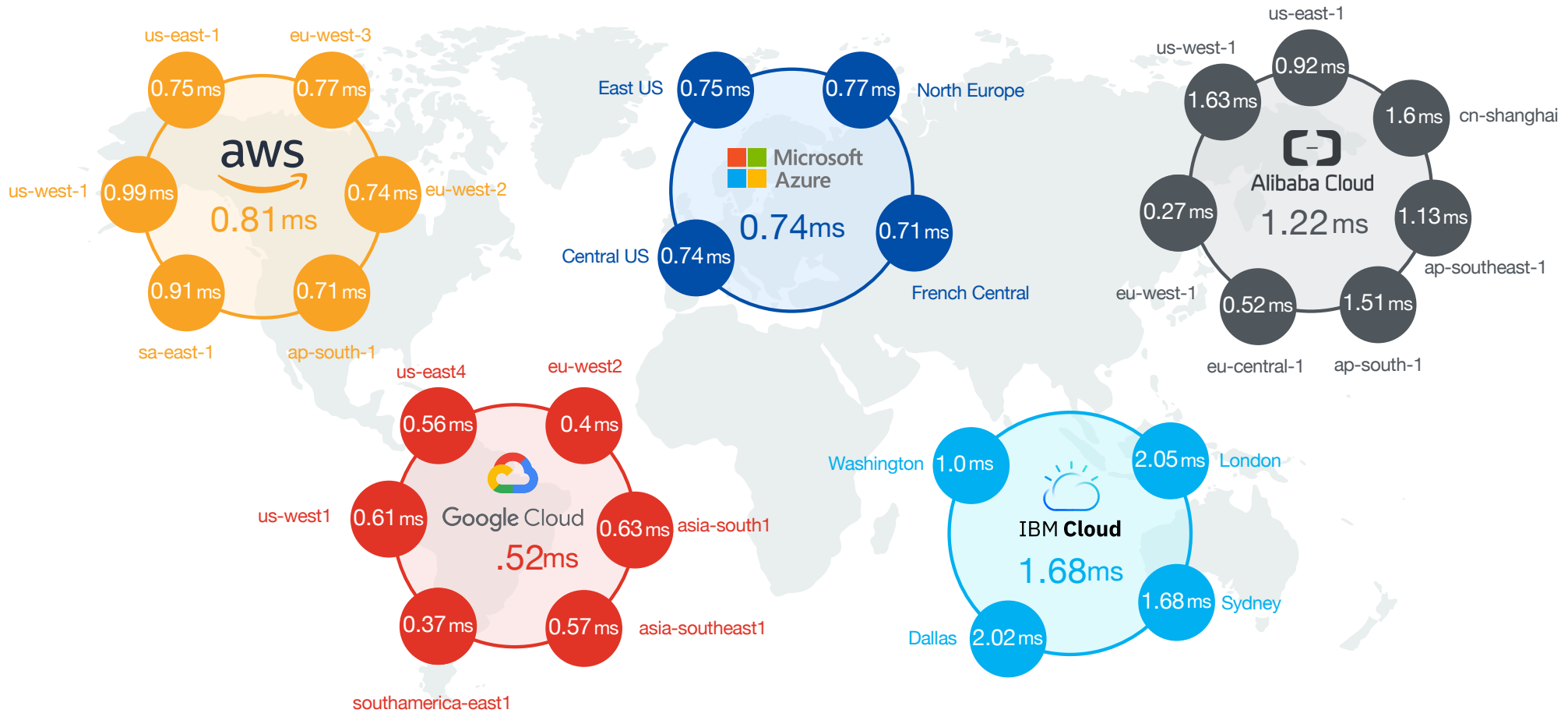
Are Cloud Backbones Created Equal?

Inter-Region Performance Reveals Connectivity

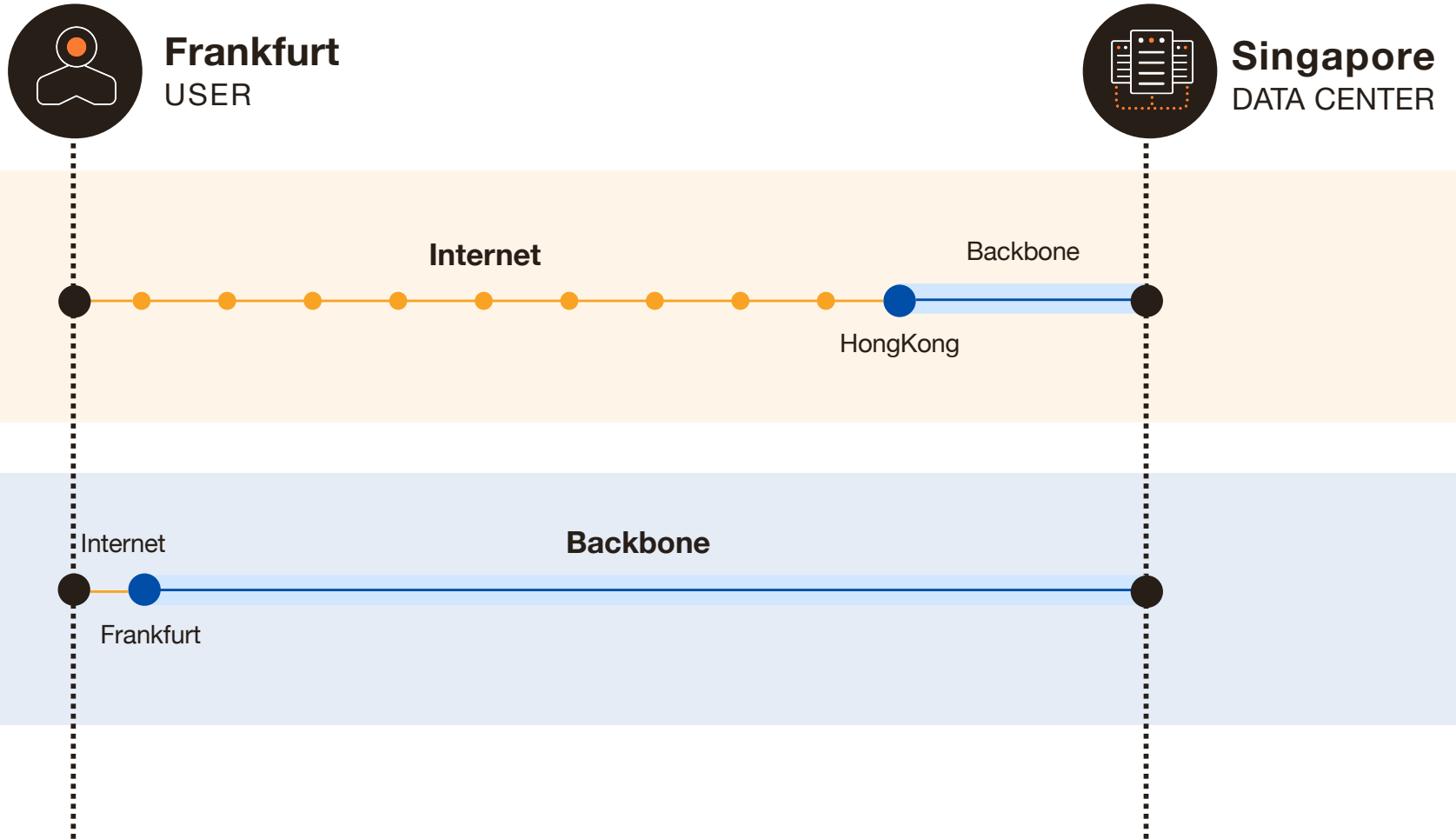
- Performance baselined with Internet averages to reflect relative performance
- 97% of IBM inter-regions pairs performed better than the Internet
- Alibaba Cloud had ~ 15% of inter-regions pairs perform worse than the Internet
- Inter-Region traffic for compute resources typically stays within the cloud provider network
 - Except for Alibaba Cloud



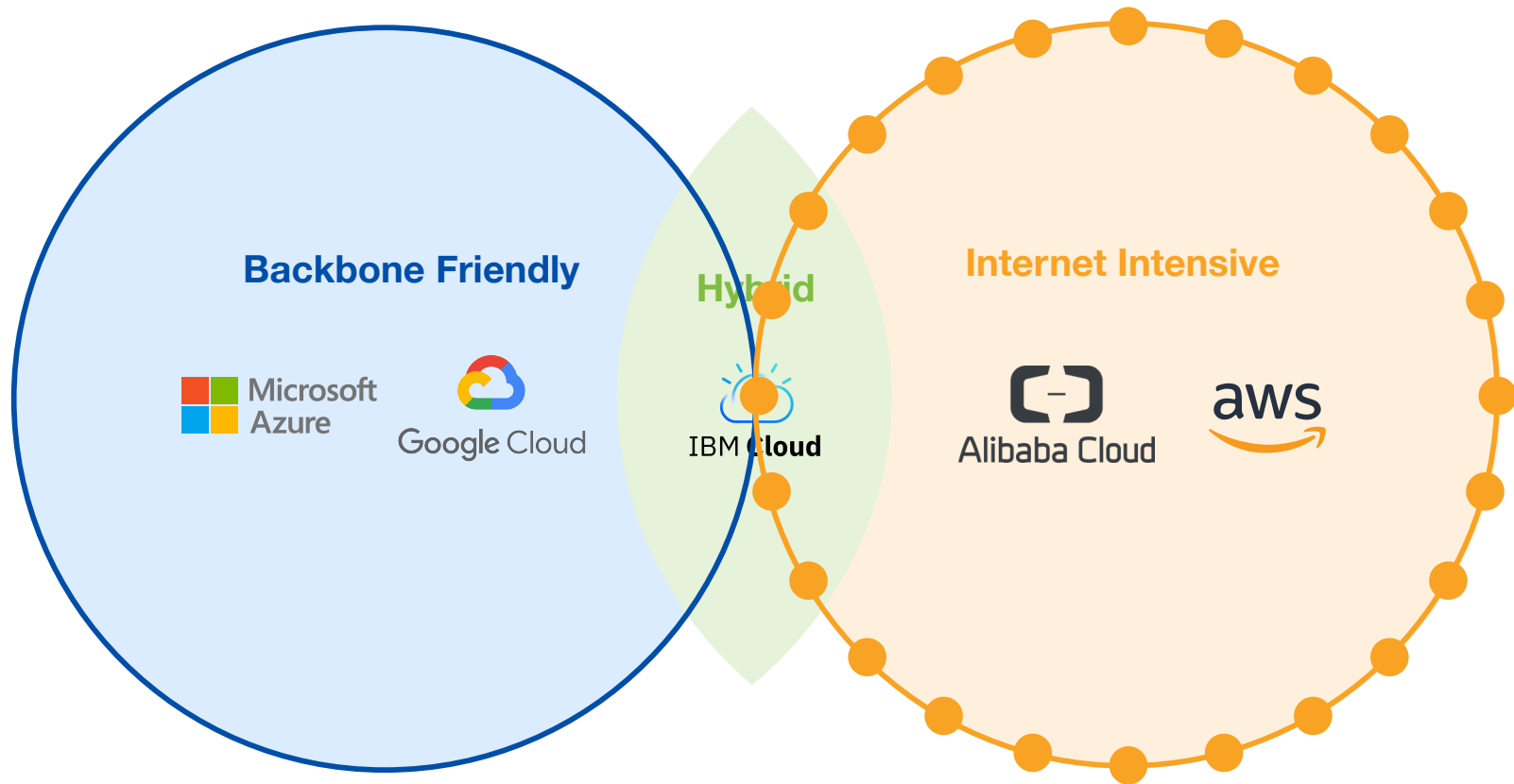
What Does Strong Inter-AZ Performance Mean?



Two Types of User <-> Region Connectivity



But Three Approaches



2018 vs. 2019 Performance

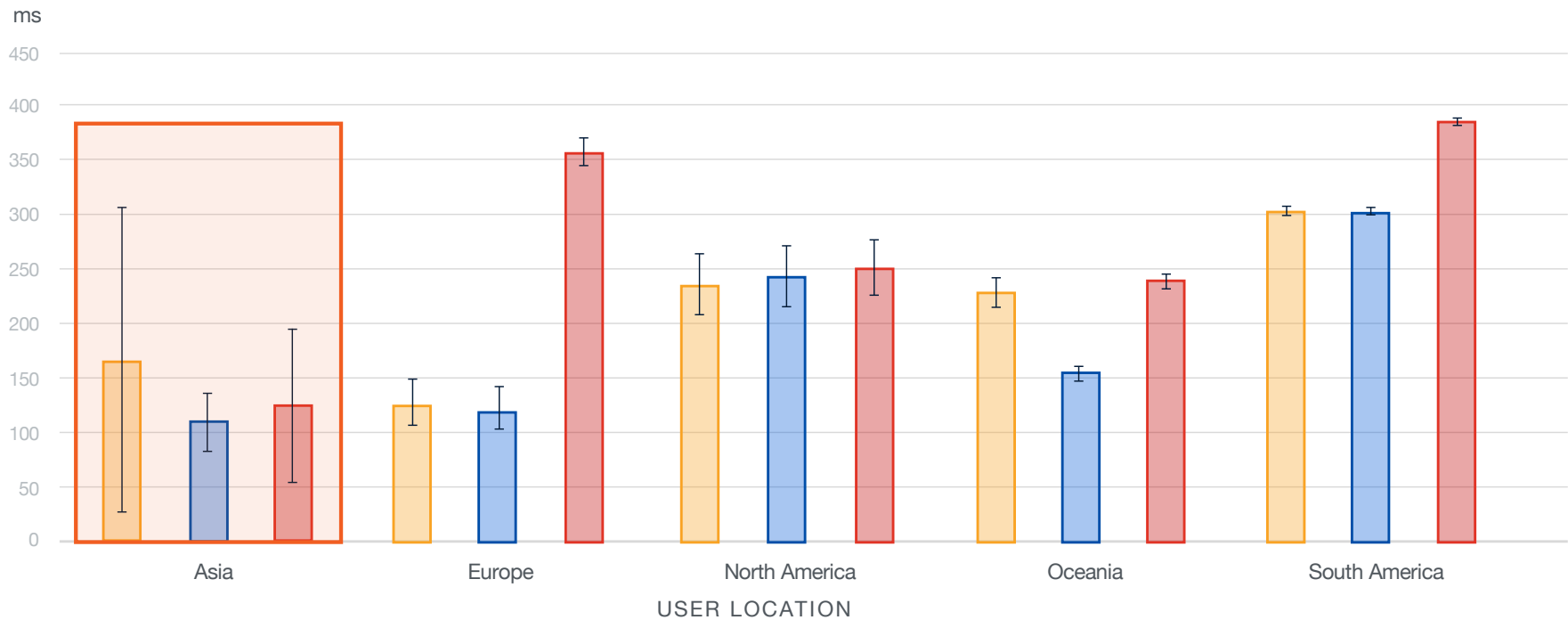
2018: AWS had High Performance Variability

HOSTING REGION: MUMBAI, INDIA

BI-DIRECTIONAL LATENCIES

END-USER MEASUREMENTS

2018 AWS Azure GCP

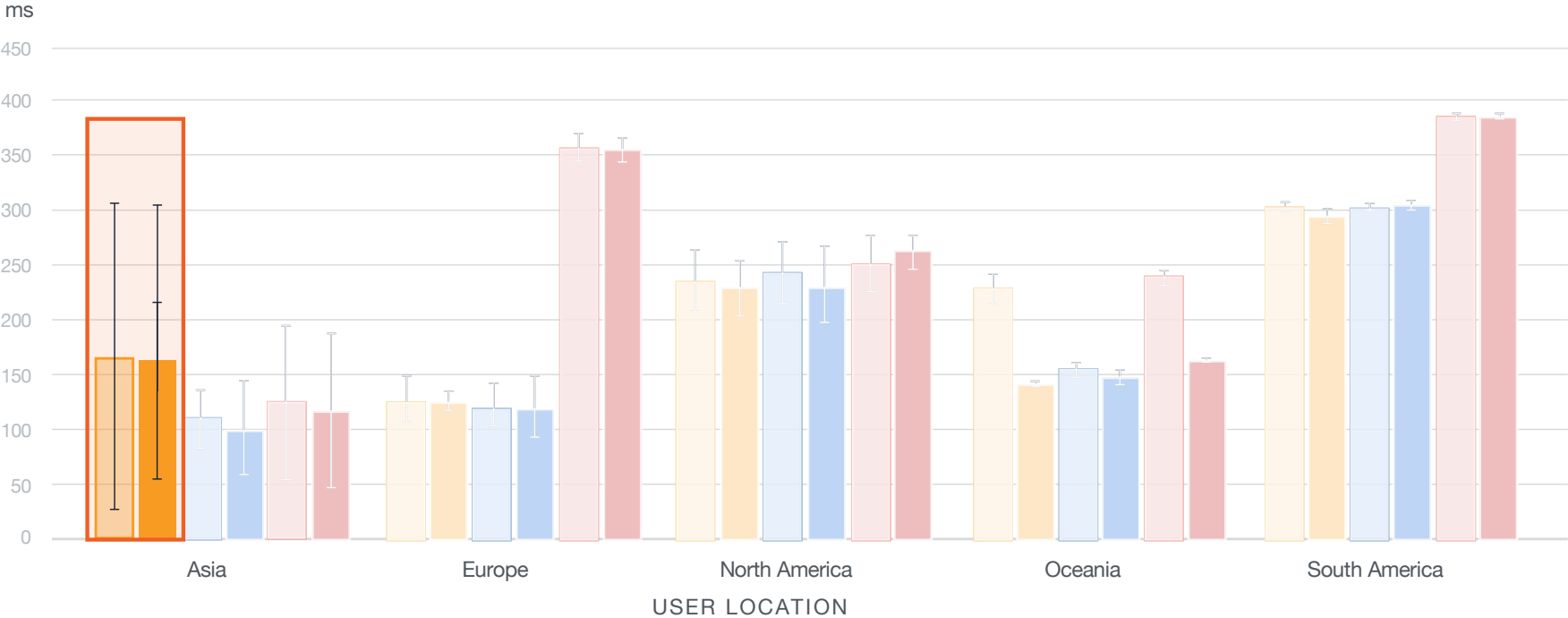


2019: AWS Improved Performance Variability

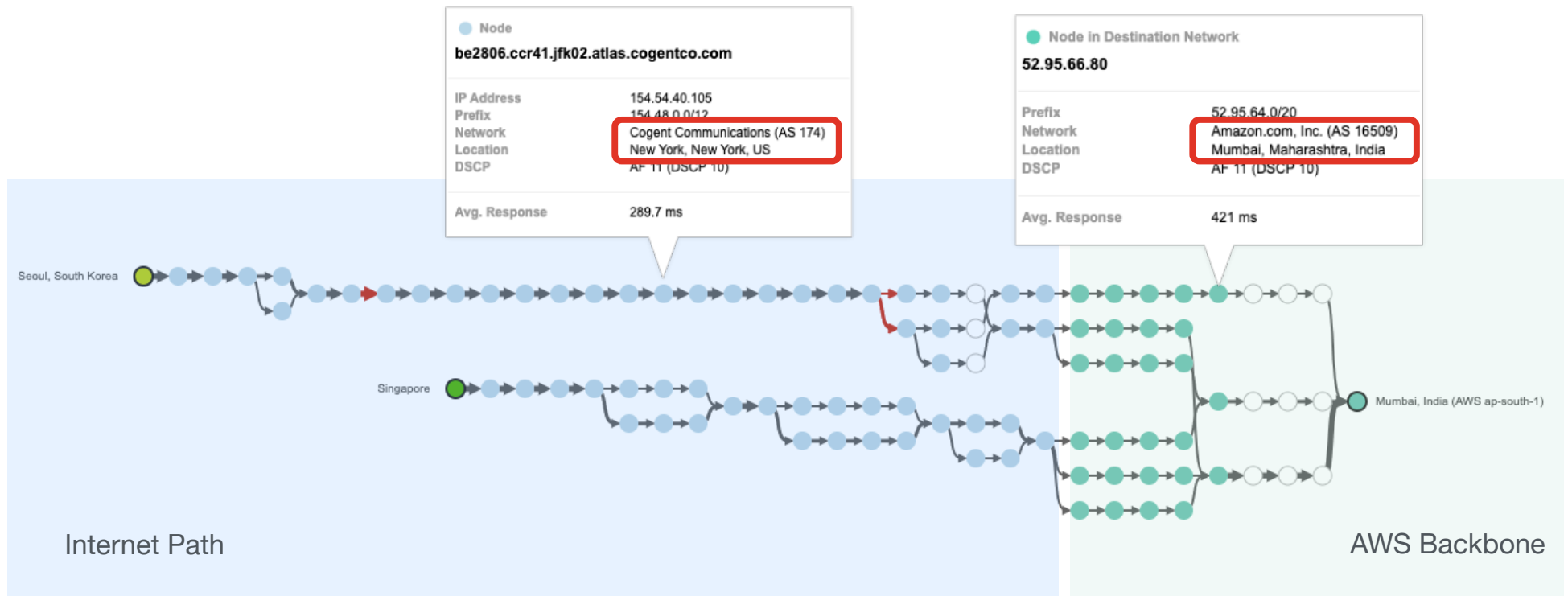
HOSTING REGION: MUMBAI, INDIA

BI-DIRECTIONAL LATENCIES

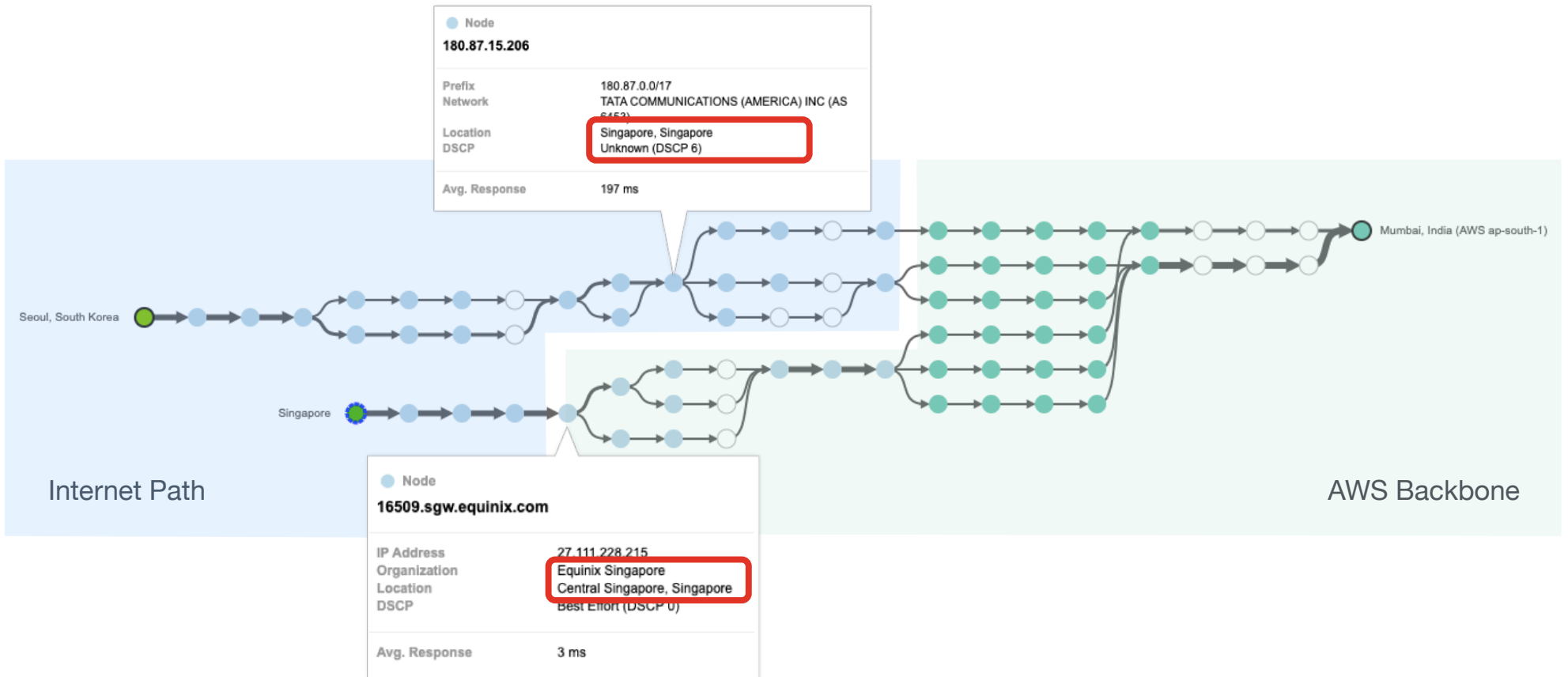
2018 AWS Azure GCP
2019 AWS Azure GCP



2018: Sub-Optimal Connectivity to AWS India



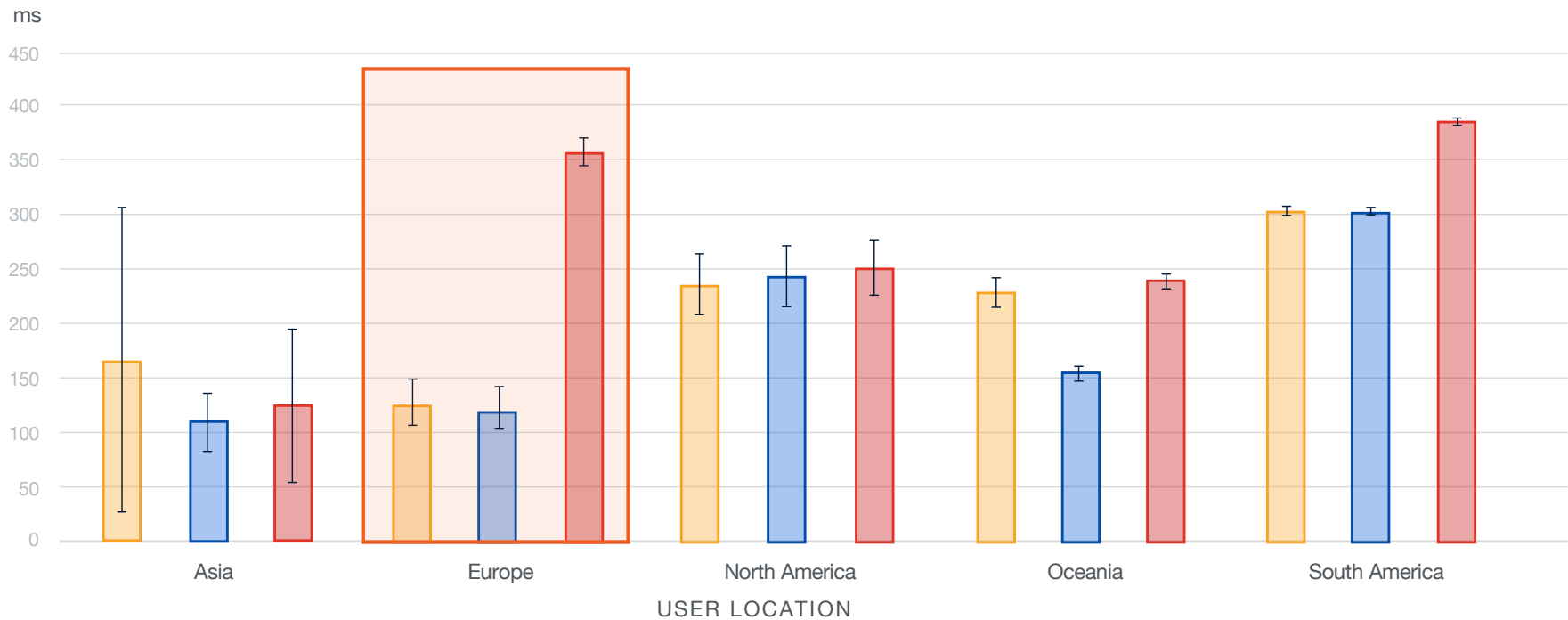
2019: AWS Optimized Routing



2018: GCP is 3x Slower From Europe to India

HOSTING REGION: MUMBAI, INDIA
BI-DIRECTIONAL LATENCIES

2018 AWS Azure GCP



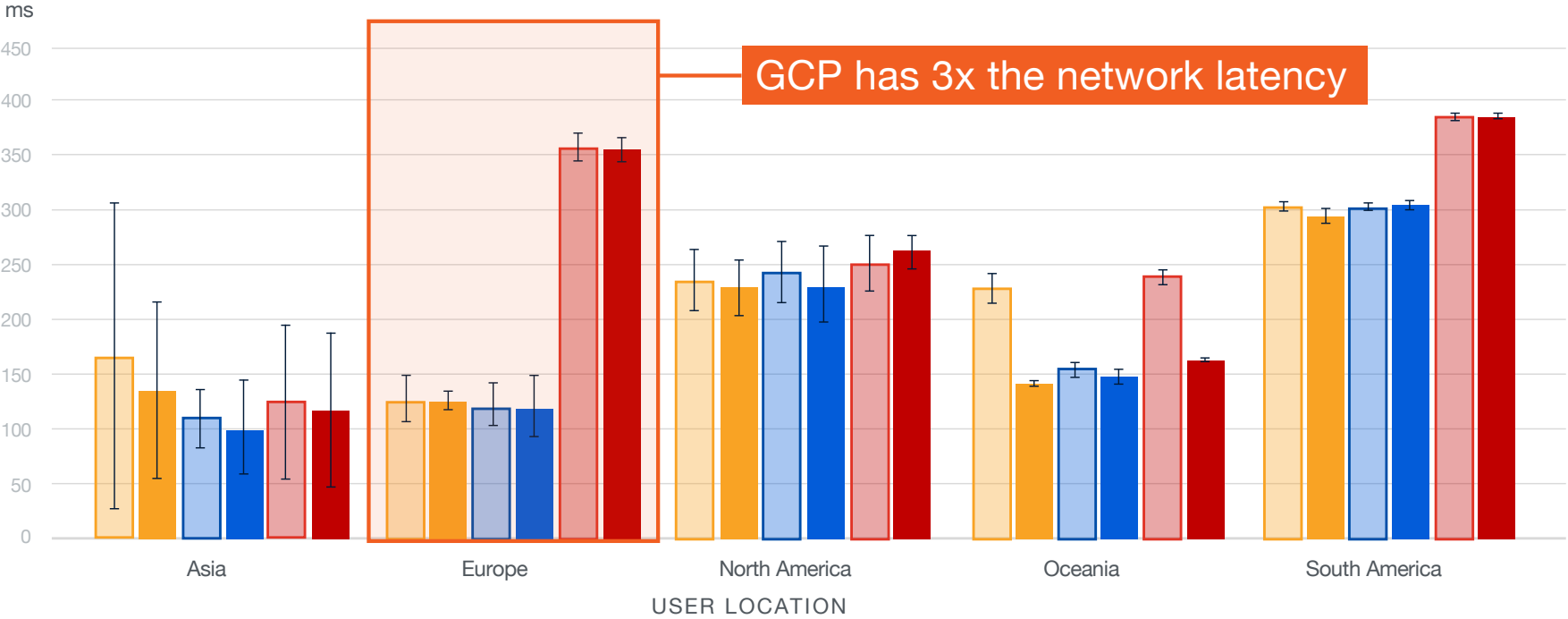
Why was GCP 3x Slower to Asia in 2018?



2019 Measurements Still Show High Latency

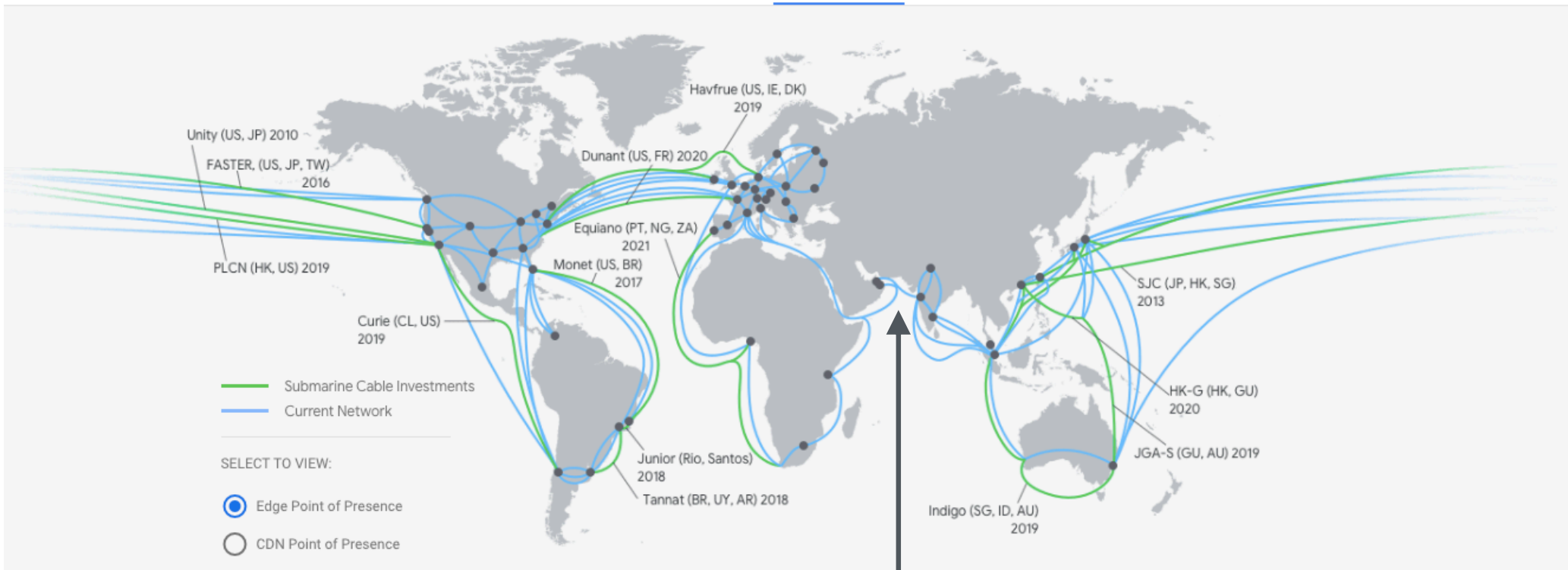
HOSTING REGION: MUMBAI, INDIA
BI-DIRECTIONAL LATENCIES

2018 AWS Azure GCP
2019 AWS Azure GCP



GCP Infrastructure Updated in 2019

REGIONS NETWORK

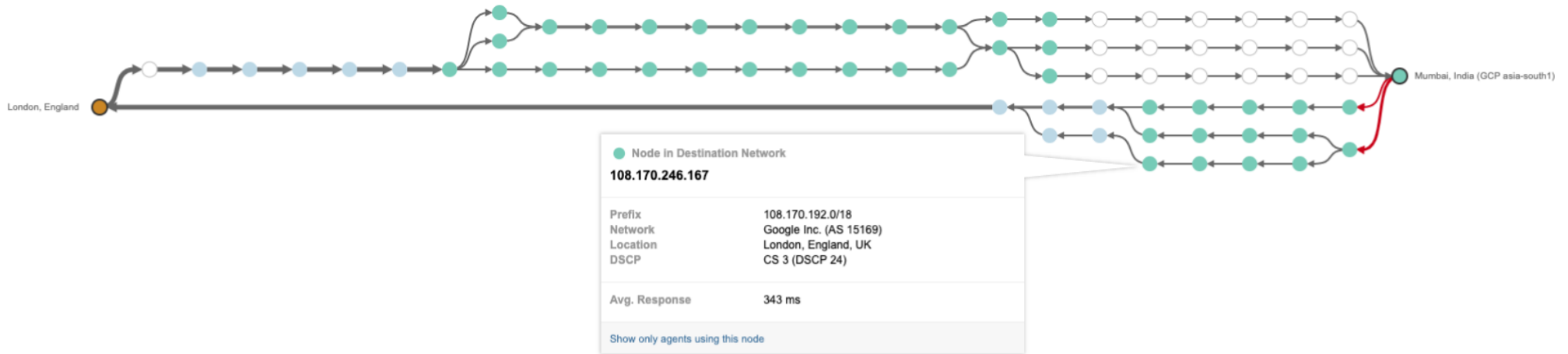


GCP infrastructure updated in 2019

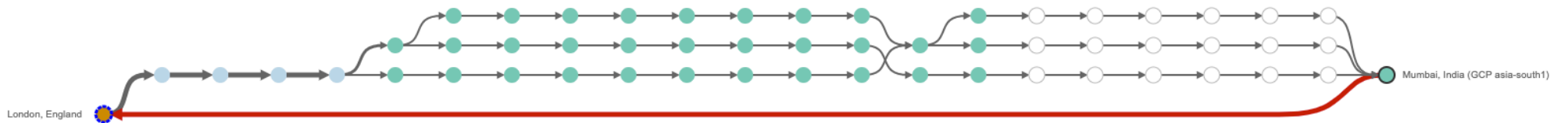
WARNING:
Low Cloud Visibility Ahead

GCP

2018: Reverse Path Visibility



2019: Reduced Visibility



Root cause: GCP Changes in TTL Handling

Traceroute to internet-bound destinations

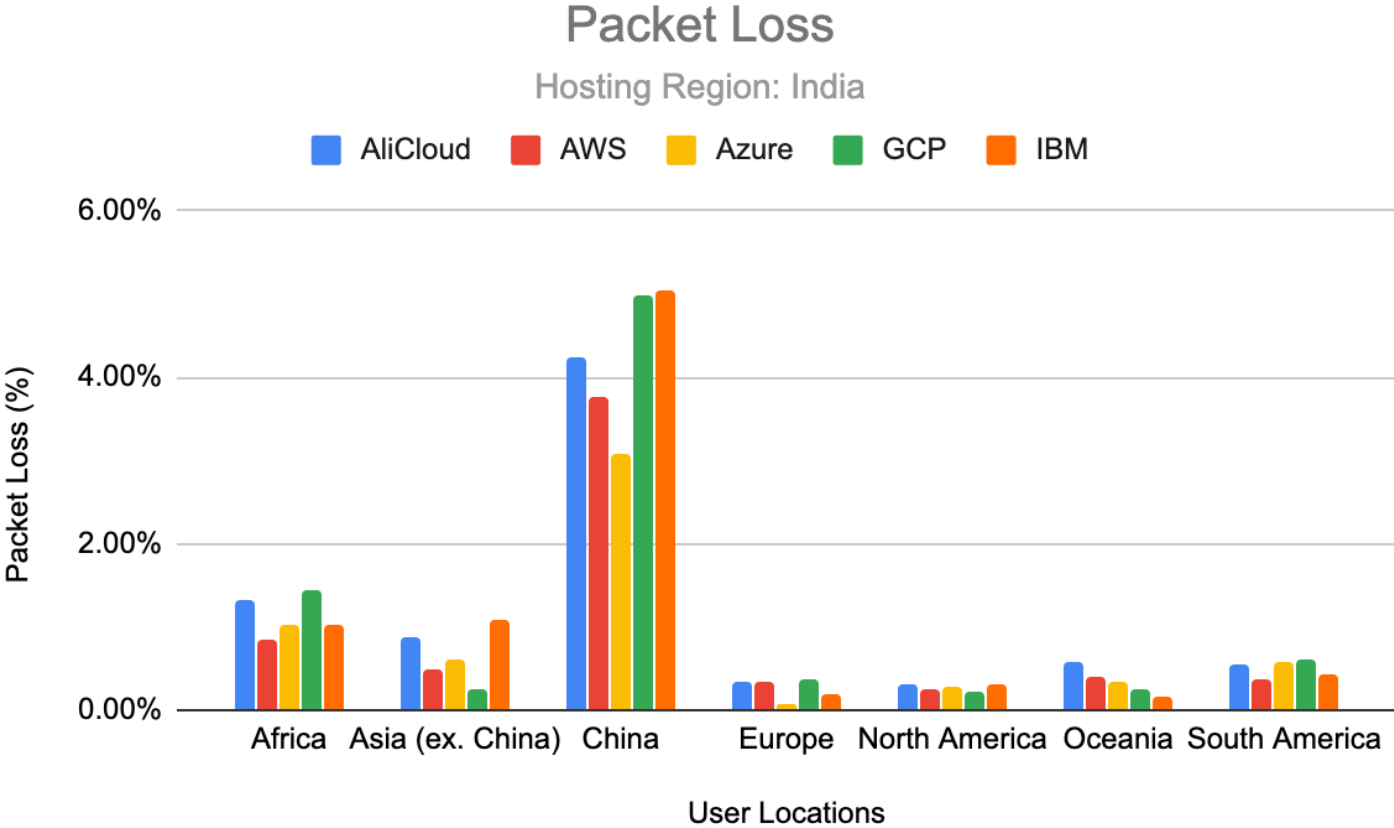
For internal reasons, GCP increases the TTL counter of packets leaving Compute Engine instances for the internet. Tools like `traceroute` might provide incomplete results because the TTL doesn't expire on some of the hops. Hops that are inside and outside of Google's network might be hidden.

The number of hidden hops varies based on the instance's Network Service Tier, region, and other factors. If there are only a few hops, it's possible for all of them to be hidden. Missing hops from a `traceroute` result don't mean that outbound traffic is dropped.

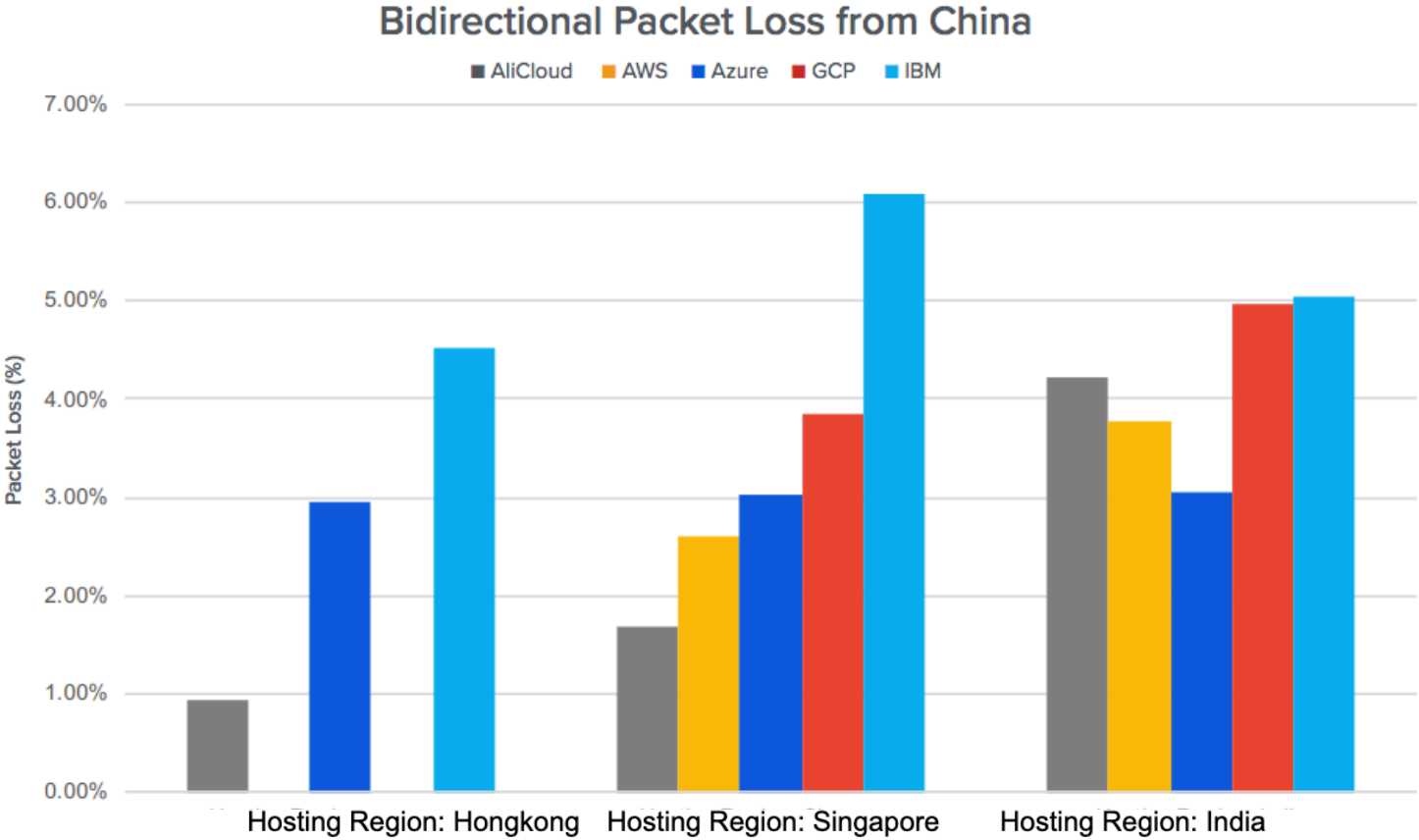
There is no workaround for this behavior.

The Performance Impact of China's Great Firewall

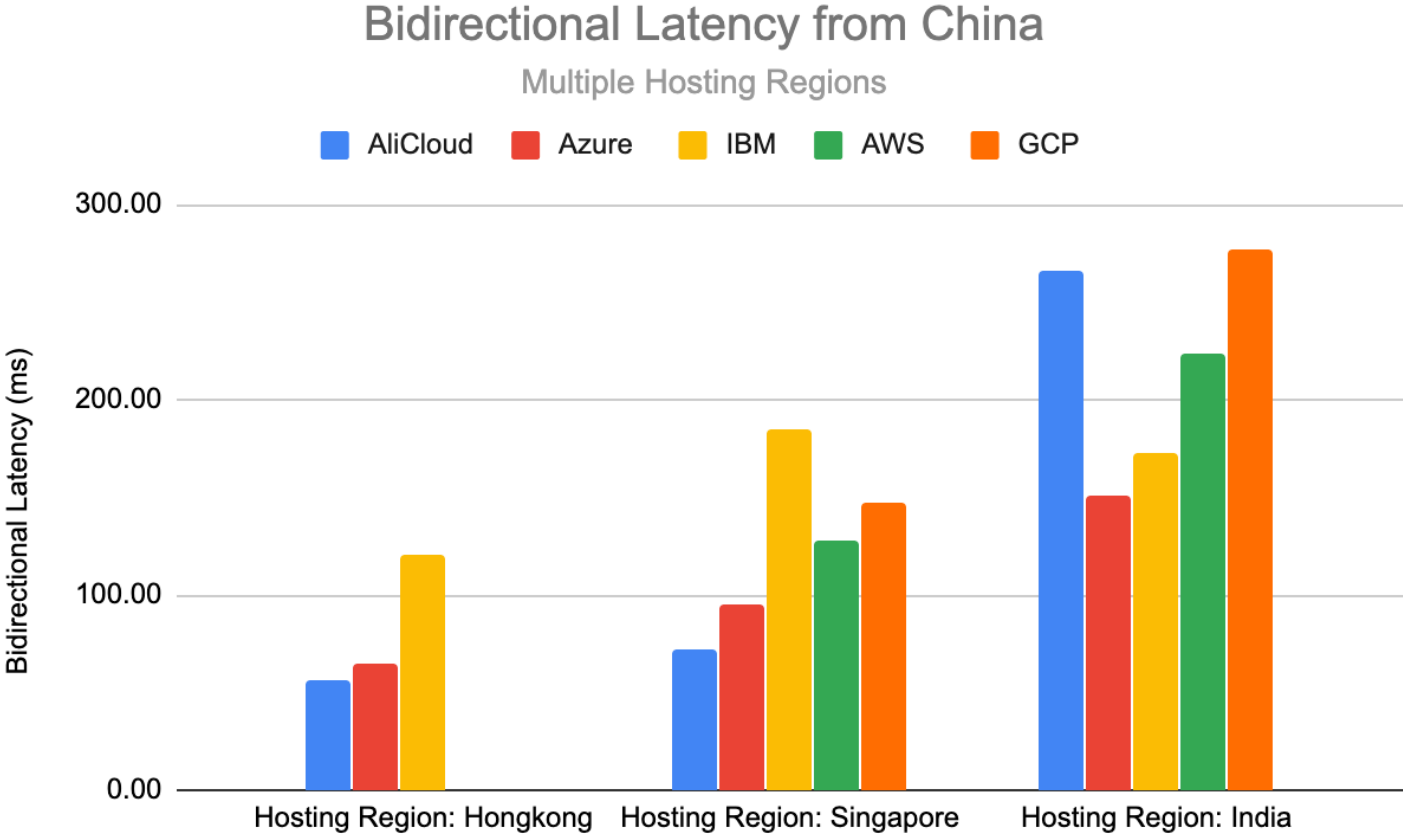
All Cloud Providers Pay Performance Toll



Viable Hosting Locations Outside of China

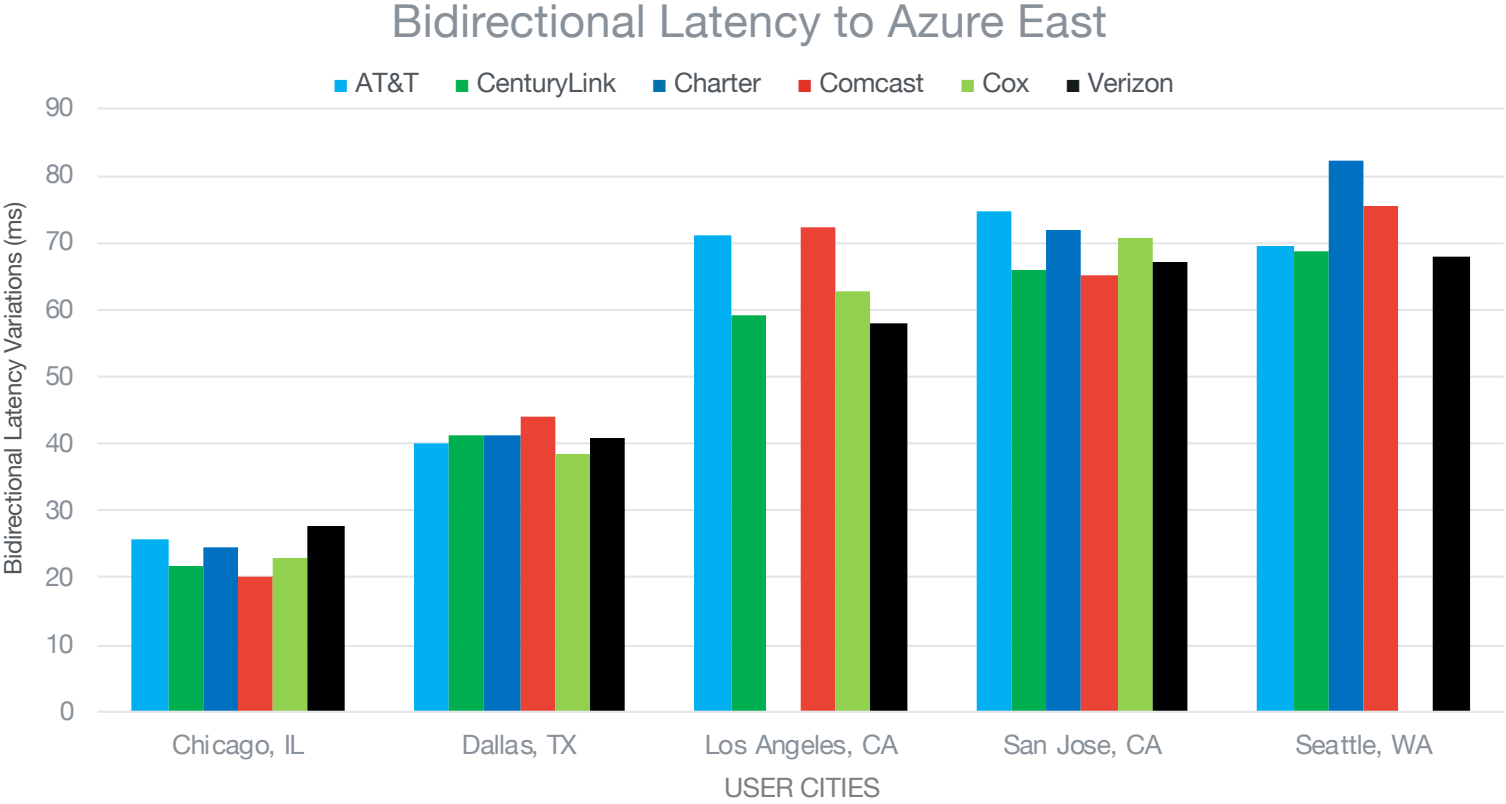


Viable Hosting Locations Outside of China



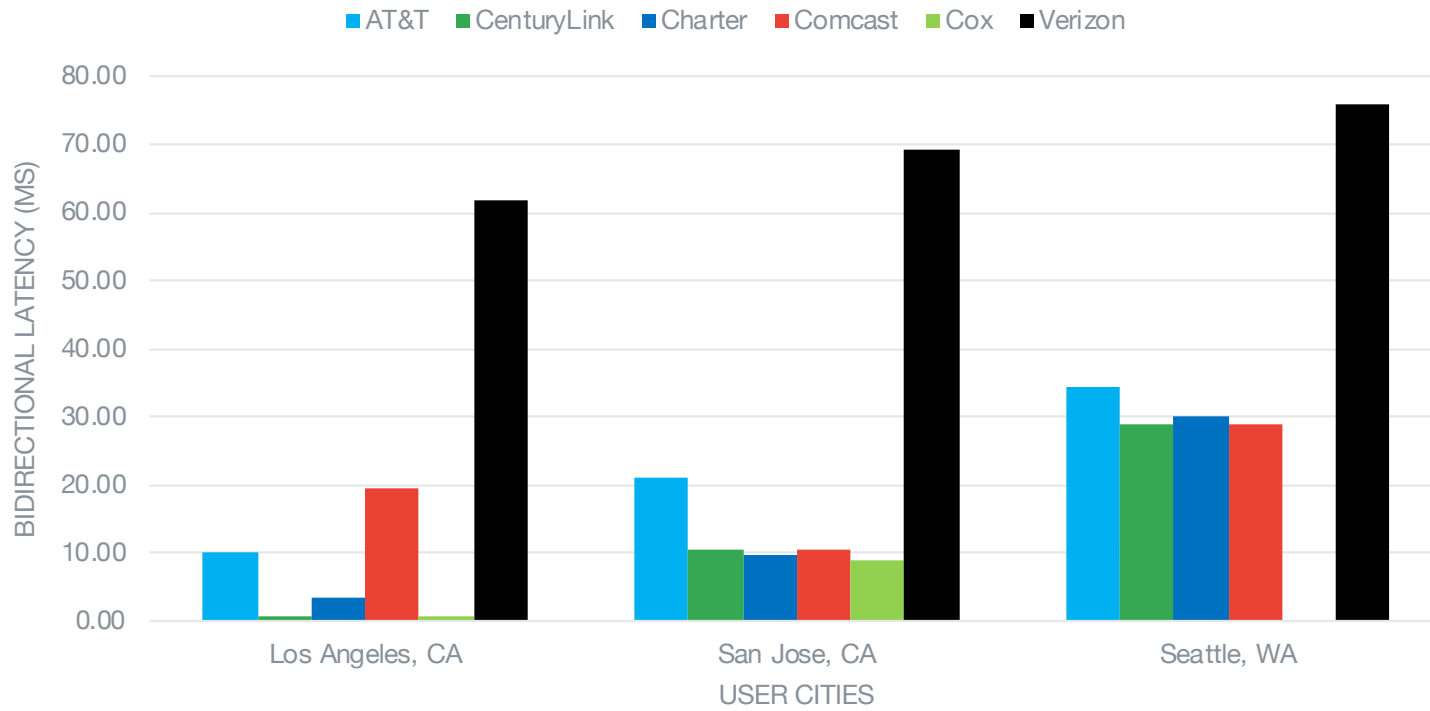
US Broadband to Cloud Connectivity

US Broadband to Cloud Connectivity Is Strong

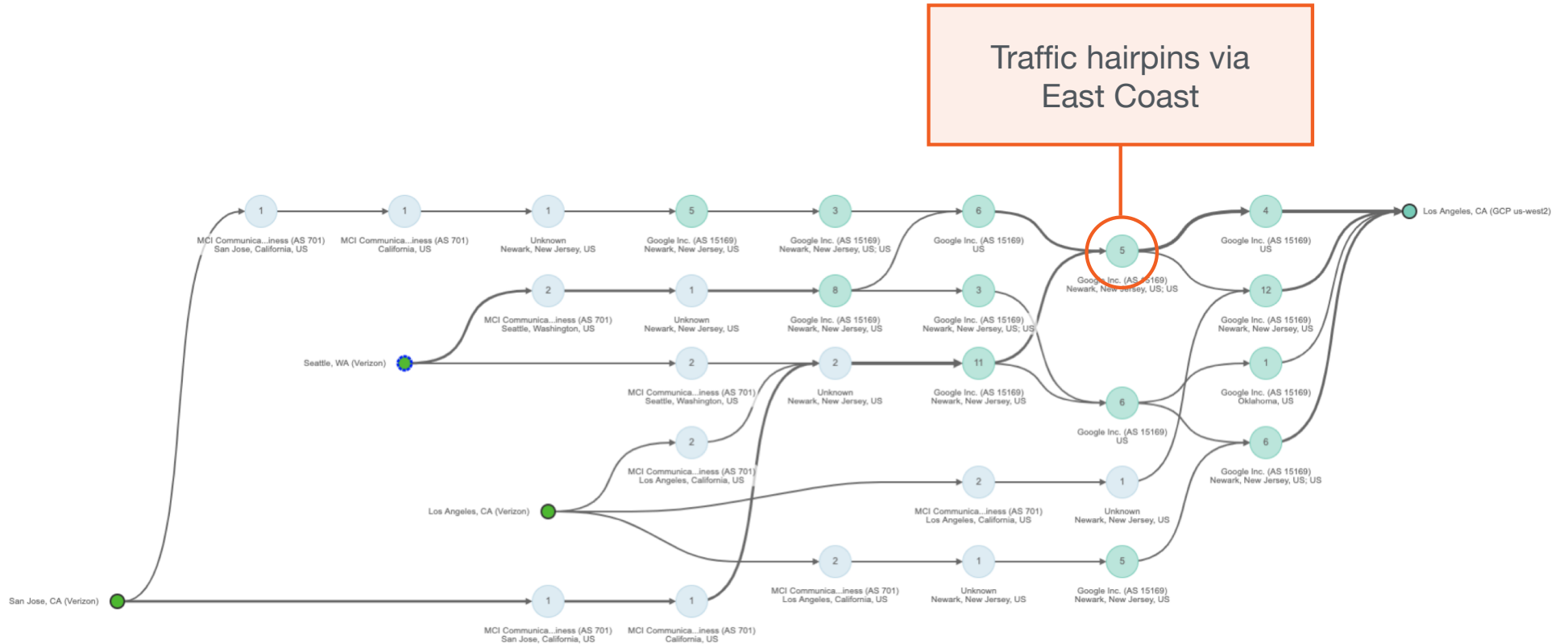


But, Exceptions Can Occur

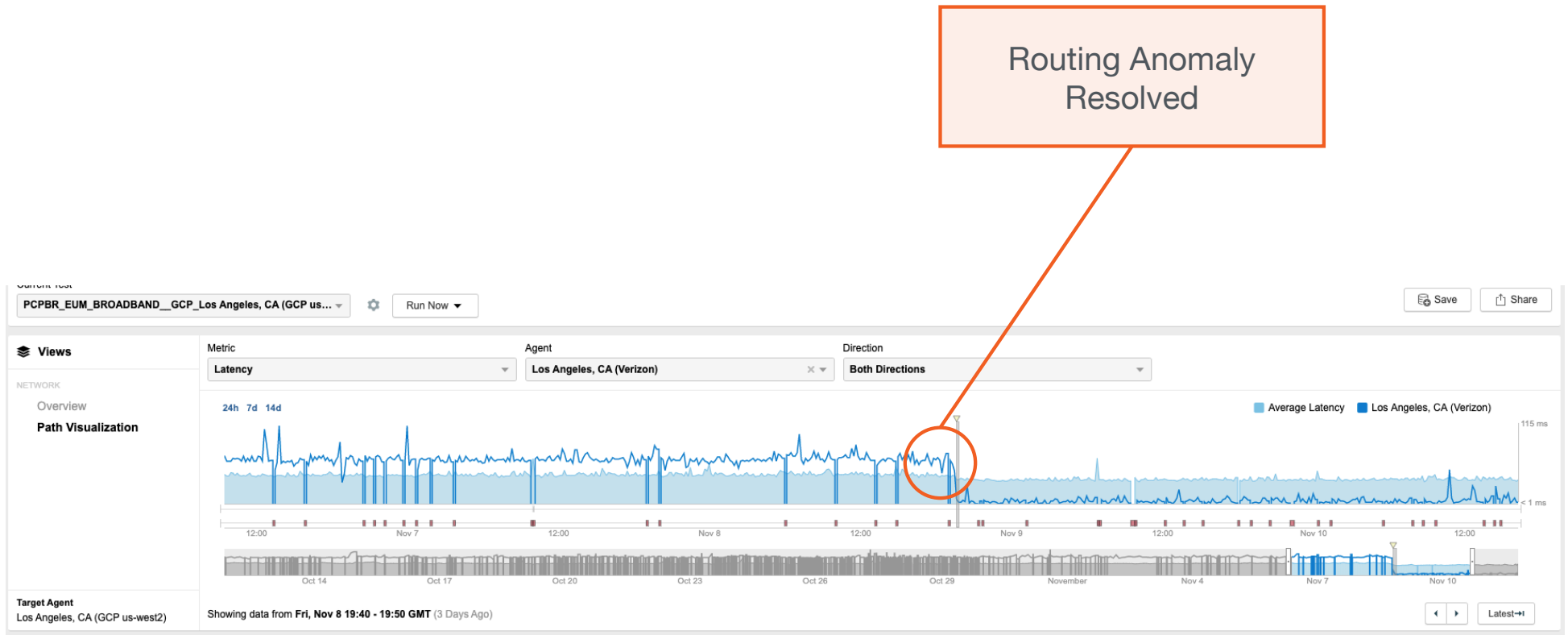
Bidirectional Latency to GCP LA (us-west2)



Routing Anomaly: SJC to LA via NJ

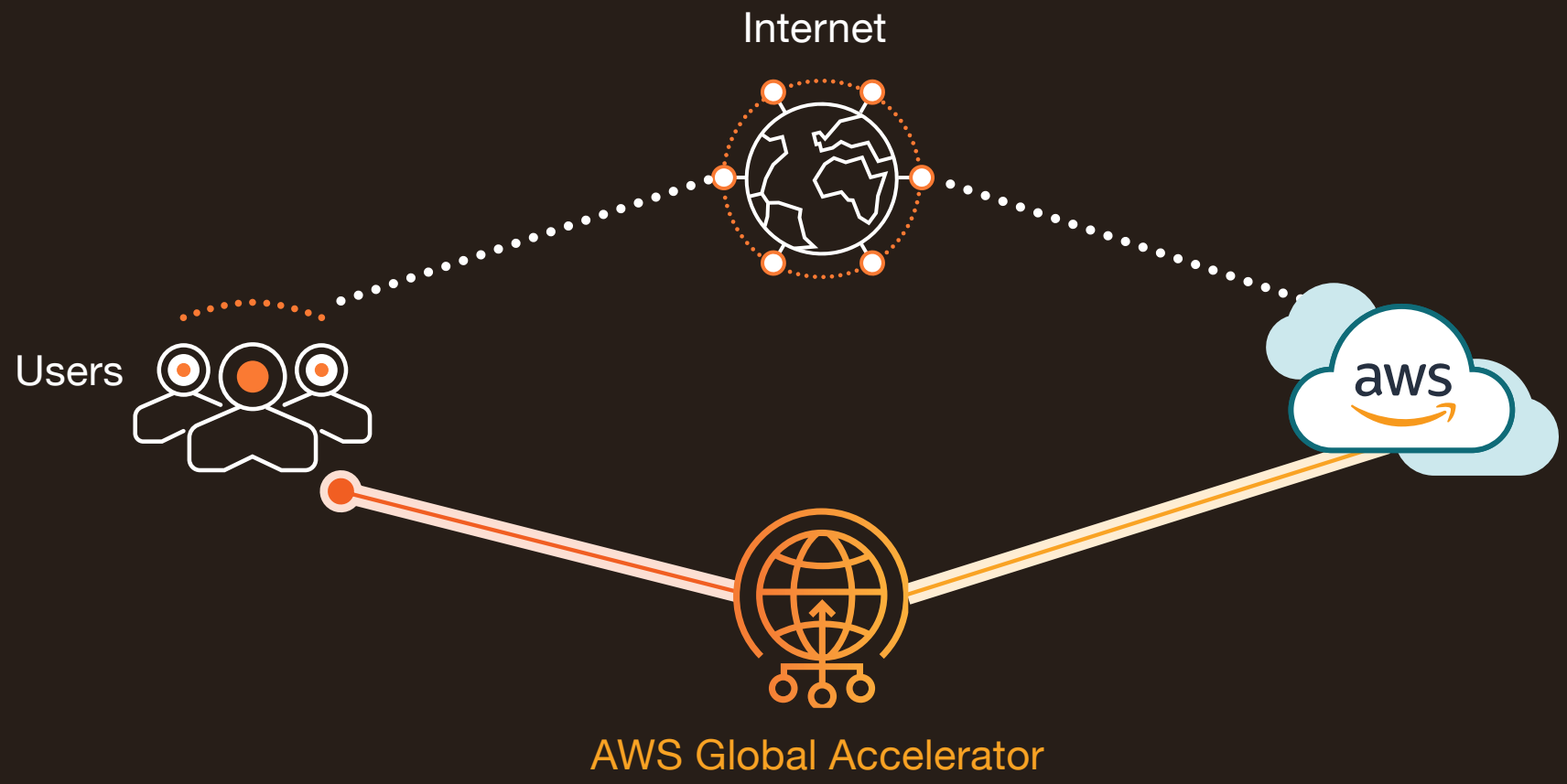


Fast MTTR from GCP – Routing Anomaly Resolved!



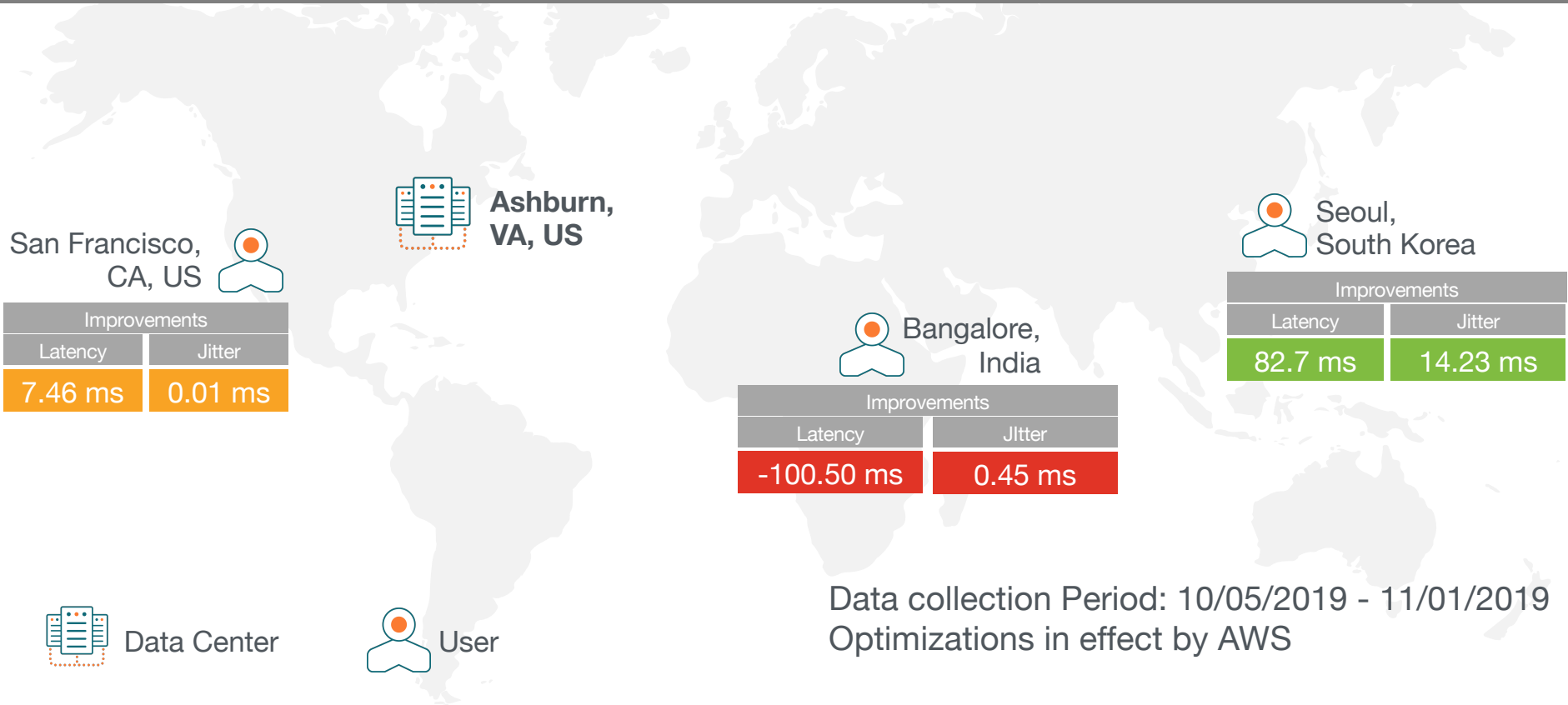
AWS Global Accelerator Performance

AWS Global Accelerator



AWS Global Accelerator

Your Mileage May Vary



Continuous Optimizations in Progress

| Vantage Points | | LATENCY | |
|-----------------------------|------------------------------|-------------------------------|-------------------------------|
| Sources | Baseline Internet Connection | Global Accelerator Connection | Global Accelerator Connection |
| Bangalore, India (Reliance) | 219.74 | 323.27 | 243.98 |
| Los Angeles, CA | 74.92 | 74.45 | 59.43 |

Oct. 2019

Dec. 2019

Wrap Up

Summary Findings

- Cloud routing preferences continue to vary
 - Backbone vs. Internet-centric vs. hybrid
- Inter-Region connectivity stays within the cloud provider network
 - Exception: Alibaba Cloud
- Inter-AZ latency < 2ms
- AWS Global Accelerator
 - Performance varies but optimization continues
- GCP Europe-to-India backbone route still pre-rollout for most geos

Takeaways

- No steady state in the cloud
- Trust, but verify performance and routing expectations
- Visibility key to provider oversight and accountability

Download the full report: ThousandEyes.com/cloud



[@bitprints](https://twitter.com/bitprints)



[@archana_k7](https://twitter.com/archana_k7)