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
- BROADBAND ISP MEASUREMENTS
- INTER-AZ MEASUREMENTS
- INTER-REGION MEASUREMENTS
- AWS GLOBAL ACCELERATOR
- MULTI-CLOUD CONNECTIVITY
End User Measurements

98 Global Locations

95 Regions

15

15

25

19

21

aws

Microsoft Azure

Google Cloud

Alibaba Cloud

IBM Cloud
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
Two Types of User <-> Region Connectivity

Internet Intensive
- Frankfurt USER
- Internet
- HongKong

Backbone Friendly
- Frankfurt
- Backbone
- Singapore DATA CENTER
But Three Approaches

- Backbone Friendly
  - Microsoft Azure
  - Google Cloud
- Hybrid
  - IBM Cloud
- Internet Intensive
  - Alibaba Cloud
  - AWS
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

USER LOCATION

ms
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?

No direct route from Europe to India
2019 Measurements Still Show High Latency

HOSTING REGION: MUMBAI, INDIA

BI-DIRECTIONAL LATENCIES

2018
AWS
Azure
GCP

2019
AWS
Azure
GCP

GCP has 3x the network latency

USER LOCATION

Asia
Europe
North America
Oceania
South America

ms
GCP Infrastructure Updated in 2019

GCP infrastructure updated in 2019
WARNING: Low Cloud Visibility Ahead
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

Packet Loss
Hosting Region: India

- AliCloud
- AWS
- Azure
- GCP
- IBM

Packet Loss (%)

User Locations

- Africa
- Asia (ex. China)
- China
- Europe
- North America
- Oceania
- South America
Viable Hosting Locations Outside of China

Bidirectional Packet Loss from China

- AliCloud
- AWS
- Azure
- GCP
- IBM

<table>
<thead>
<tr>
<th>Packet Loss (%)</th>
<th>Hosting Region: Hongkong</th>
<th>Hosting Region: Singapore</th>
<th>Hosting Region: India</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5%</td>
<td>3.5%</td>
<td>5.5%</td>
<td>5.0%</td>
</tr>
<tr>
<td>2.0%</td>
<td>4.0%</td>
<td>4.5%</td>
<td>4.0%</td>
</tr>
<tr>
<td>3.0%</td>
<td>3.0%</td>
<td>3.5%</td>
<td>3.0%</td>
</tr>
<tr>
<td>5.0%</td>
<td>5.0%</td>
<td>6.0%</td>
<td>6.0%</td>
</tr>
<tr>
<td>7.0%</td>
<td>7.0%</td>
<td>7.0%</td>
<td>7.0%</td>
</tr>
</tbody>
</table>
Viable Hosting Locations Outside of China

Bidirectional Latency from China
Multiple Hosting Regions

- AliCloud
- Azure
- IBM
- AWS
- GCP

Bidirectional Latency (ms)

Hosting Region: Hongkong
Hosting Region: Singapore
Hosting Region: India
US Broadband to Cloud Connectivity
US Broadband to Cloud Connectivity Is Strong

Bidirectional Latency to Azure East

- AT&T
- CenturyLink
- Charter
- Comcast
- Cox
- Verizon

USER CITIES:
- Chicago, IL
- Dallas, TX
- Los Angeles, CA
- San Jose, CA
- Seattle, WA
But, Exceptions Can Occur

Bidirectional Latency to GCP LA (us-west2)

- Los Angeles, CA
- San Jose, CA
- Seattle, WA

User Cities:
- AT&T
- CenturyLink
- Charter
- Comcast
- Cox
- Verizon
Routing Anomaly: SJC to LA via NJ

Traffic hairpins via East Coast
Fast MTTR from GCP – Routing Anomaly Resolved!

Routing Anomaly Resolved
AWS Global Accelerator

Your Mileage May Vary

Data collection Period: 10/05/2019 - 11/01/2019
Optimizations in effect by AWS
## Continuous Optimizations in Progress

<table>
<thead>
<tr>
<th>Vantage Points</th>
<th>LATENCY</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline Internet Connection</td>
<td>Global Accelerator Connection</td>
<td>Global Accelerator Connection</td>
</tr>
<tr>
<td>Bangalore, India (Reliance)</td>
<td>219.74</td>
<td>323.27</td>
<td>243.98</td>
</tr>
<tr>
<td>Los Angeles, CA</td>
<td>74.92</td>
<td>74.45</td>
<td>59.43</td>
</tr>
</tbody>
</table>

Wrap Up
Summary Findings

• Cloud routing preferences continue to vary
  o Backbone vs. Internet-centric vs. hybrid

• Inter-Region connectivity stays within the cloud provider network
  o Exception: Alibaba Cloud

• Inter-AZ latency < 2ms

• AWS Global Accelerator
  o 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

@archana_k7