



CLOUD PERFORMANCE BENCHMARK

2019 – 2020 Edition

2019 – 2020 EDITION

The Cloud Performance Benchmark measures and compares network performance between the top five cloud providers: Amazon Web Services, Google Cloud Platform, Microsoft Azure, IBM Cloud and Alibaba Cloud.

CONTENTS

3	EXECUTIVE SUMMARY
4	METHODOLOGY
5	The ThousandEyes Platform
6	End-User Measurements
8	Broadband Measurements
9	Inter-AZ and Inter-Region Measurements
11	Amazon Global Accelerator Measurements
12	Multi-Cloud Connectivity
13	Important Disclaimers
14	FINDINGS AND TAKEAWAYS
15	2018 vs 2019 Comparison
25	Understanding Cloud Connectivity Architectures
31	Global End User Network Performance
38	Network Performance in China
42	Broadband ISP Performance
45	Inter-Region Performance
54	Inter-AZ Performance
57	AWS Global Accelerator
61	Multi-Cloud Connectivity
62	CONCLUSION AND RECOMMENDATIONS
63	APPENDIX
64	Updates NEW
65	Global End-User Network Performance
68	AWS Global Accelerator

EXECUTIVE SUMMARY

The 2019 Cloud Performance Benchmark measures and compares network performance between five top public cloud providers: Amazon Web Services (AWS), Microsoft Azure (Azure), Google Cloud Platform (GCP), Alibaba Cloud and IBM Cloud. The measurements gathered benchmark the cloud providers against each other to discover what constitutes average, normative and best-in-class network performance.

As a follow-up to the inaugural 2018 study that covered AWS, Azure and GCP, the collected 2019 data makes comparisons with last year's data to showcase what has and hasn't changed with network performance year over year. The 2019 report adds substantive new findings on AWS' Global Accelerator service offering, cloud connectivity in and out of China, performance of US broadband provider connectivity to the cloud, and findings specific to the newly added providers Alibaba Cloud and IBM Cloud.

Enterprises making cloud choices rely heavily on comparative studies. Most previously available studies on cloud providers focus on services offered, pricing tiers and global data center presence. However, performance studies of public cloud providers have historically been missing in action. The few studies that existed lacked breadth of coverage, as well as both depth and duration of metric data. The Cloud Performance Benchmark provides a unique, unbiased third-party and metric-based perspective on public cloud performance as it relates to both end-user experience and back-end application architecture.

The report reveals comparable network performance data across all five public cloud providers. However, significant anomalies exist and public cloud provider connectivity approaches vary significantly, leading to geographical disparities in network performance and predictability. The report also highlights the performance toll that China's Great Firewall takes on Internet traffic to and from the country, and uncovers how US broadband ISPs impact cloud deployments.

Ultimately, it is imperative for enterprise IT leaders to understand that cloud architectures are complex and not to rely on network performance and connectivity assumptions or instincts while designing them. Enterprises relying heavily on the public cloud or considering a move to the cloud must arm themselves with the right data on an ongoing basis to guide the planning and operational stages. Every organization is different, cloud architectures are highly customized and hence these results must be reviewed through the lens of one's own business in choosing providers, regions and connectivity approaches.

The results and summary presented in this report are based on data gathered during the collection period. Because there is no steady state in the cloud, enterprises relying on the cloud should continuously monitor for changes and optimizations made by the cloud providers. The Appendix section of the report has been updated to capture some of the ongoing changes to cloud performance after its initial publication.

METHODOLOGY

The findings presented in this report are based on data gathered from periodically monitoring bi-directional network performance such as latency, packet loss and jitter to, within and between multiple global regions of the five public cloud providers over a four-week period.

Analysis of over 320M data points and in-depth path traces culminate as insights, trends and recommendations prescribed in this report. The insights uncovered in this report are categorized based on two guiding principles—the data collection methodology and common trends seen across the dataset. While the principles of data collection such as metrics gathered or frequency of testing remained consistent within the study, multiple test methodologies were deployed.



END-USER MEASUREMENTS

Network performance metrics gathered **from** global user location vantage points to global cloud regions across all five providers



BROADBAND ISP MEASUREMENTS

Network performance metrics **to** the five cloud providers' hosting regions in North America from six broadband service providers in six US cities



INTER-AZ AND INTER-REGION MEASUREMENTS

Inter-AZ and Inter-region performance **within** the same cloud provider



GLOBAL ACCELERATOR MEASUREMENTS

Network performance metrics for AWS' Global Accelerator service from global vantage points



MULTI-CLOUD CONNECTIVITY

Analysis of connectivity patterns **between** the five cloud providers

The data presented in this report is collected using the ThousandEyes platform and testing framework. Before we get into the specifics of the methodologies listed above, it helps to outline the common guiding principles of the data collection framework.

THE THOUSANDEYES PLATFORM

ThousandEyes uses an active monitoring technique to gather network metrics such as loss, latency and jitter along with in-depth path metrics with detailed layer 3 hops. ThousandEyes vantage point agents are deployed on both sides of the test measurement. These agents generate a stream of TCP probe packets at pre-configured intervals in each direction. This allows us to measure loss, latency and jitter per direction independently. For example, bi-directional latency is a combination of latency measurements from source to target agent and vice versa.



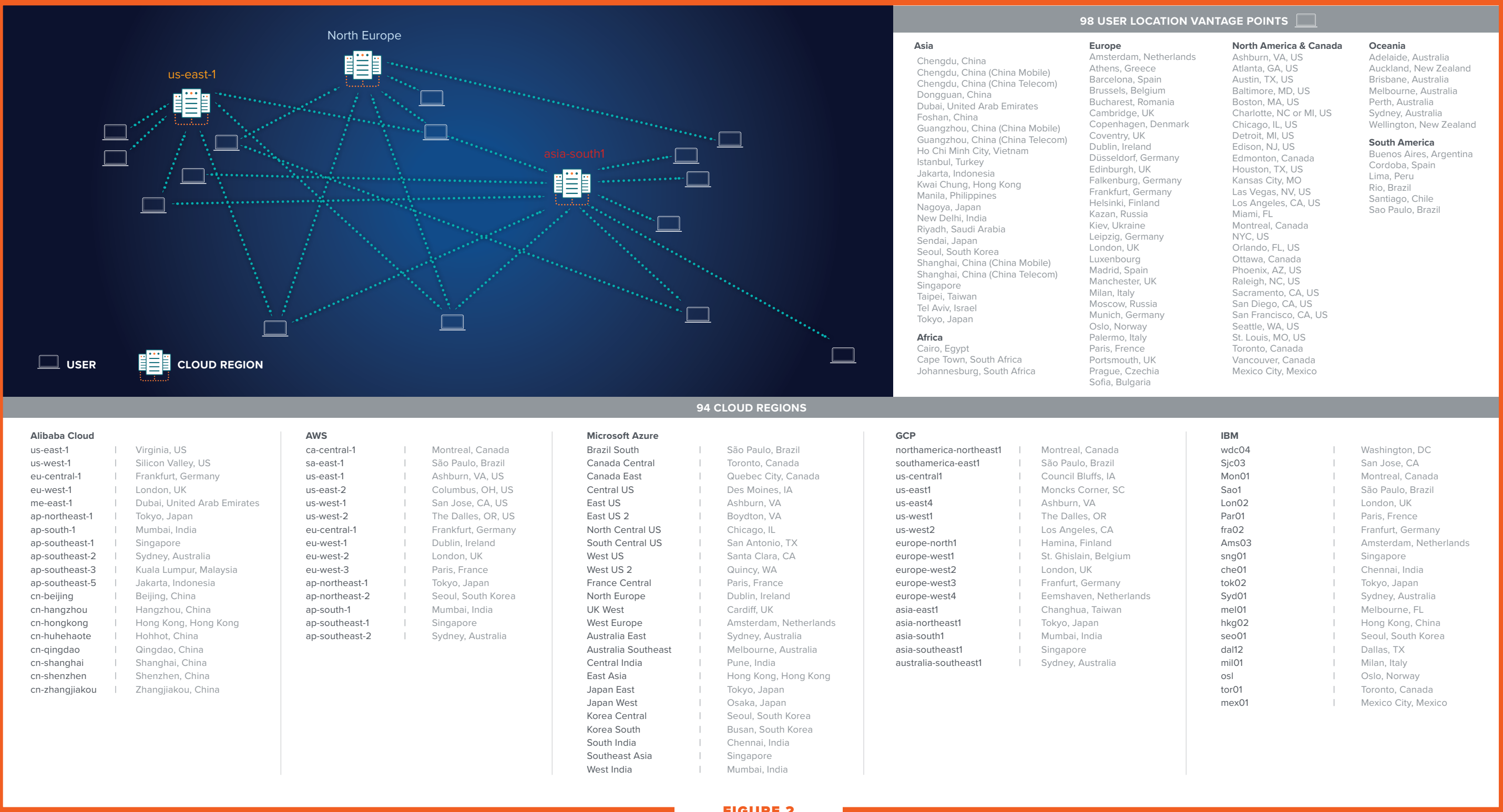
FIGURE 1

ThousandEyes active monitoring infrastructure powers the Cloud Performance Benchmark Report

END-USER MEASUREMENTS

End-User Measurements Data Collection Period: 09/01/2019 - 09/30/2019

Network performance metrics were gathered every 10 minutes from 98 user vantage points deployed in data centers around the globe to 95 cloud regions across all five providers. The 98 user vantage points are hosted in Tier 2 and Tier 3 ISPs and were picked to represent a uniform distribution around the globe. All user vantage points and the cloud hosting regions of the five cloud providers are listed in Figure 2. Network tests utilize TCP-based probes to collect hop-by-hop network path data along with network metrics like loss, latency and jitter. The data consists of bi-directional measurements and includes both forward and reverse path information.



While we tested to all 95 hosting regions, we only compared results across the regions identified in Table 1 below, as these data center locations provided the best opportunity to create an “apples to apples” comparison across the cloud providers. NA (Not Applicable) means that the provider doesn’t have a presence in that region. If you are interested in a dataset beyond the regions listed in the table, please contact ThousandEyes at cloudreport@thousandeyes.com.

	AMAZON WEB SERVICES	MICROSOFT AZURE	ALIBABA CLOUD	GOOGLE CLOUD PLATFORM	IBM CLOUD
United States East	us-east-1 Ashburn, VA	East US Ashburn, VA	us-east-1 Virginia, NV	us-east4 Ashburn, VA	wdc04 Washington, DC
United States West	us-west-1 San Jose, CA	West US Santa Clara, CA	us-west-1 Silicon Valley, CA	us-west2 Los Angeles, CA	sjc03 San Jose, CA
United States Central	us-east-2 Columbus, OH	Central US Des Moines, IA	NA	us-central1 Council Bluffs, IA	NA
Canada	ca-central-1 Montreal, Canada	Canada East Quebec City, Canada	NA	northamerica-northeast1 Montreal, Canada	mon01 Montreal, Canada
South America	sa-east-1 São Paulo, Brazil	Brazil South São Paulo, Brazil	NA	southamerica-east1 São Paulo, Brazil	sao1 São Paulo, Brazil
Europe – London / Cardiff	eu-west-2 London, UK	UK West Cardiff, UK	eu-west-1 London, UK	europa-west2 London, UK	lon02 London, UK
Europe – Paris	eu-west-3 Paris, France	France Central Paris, France	NA	NA	par01 Paris, France
Europe – Frankfurt	eu-central-1 Frankfurt, Germany	NA	eu-central-1 Frankfurt, Germany	europa-west3 Frankfurt, Germany	fra02 Frankfurt, Germany
Europe – Benelux	NA	West Europe Amsterdam, Netherlands	NA	europa-west4 Eemshaven, Netherlands	ams03 Amsterdam, Netherlands
Asia – Singapore	ap-southeast-1 Singapore	Southeast Asia Singapore	ap-southeast-1 Singapore	asia-southeast1 Singapore	sng01 Singapore
Asia – India	ap-south-1 Mumbai, India	West India Mumbai, India	ap-south-1 Mumbai, India	asia-south1 Mumbai, India	che01 Chennai, India
Apac – Tokyo	ap-northeast-1 Tokyo, Japan	Japan East Tokyo, Japan	ap-northeast-1 Tokyo, Japan	asia-northeast1 Tokyo, Japan	tok02 Tokyo, Japan
Apac – Australia	ap-southeast-2 Sydney, Australia	Australia East Sydney, Australia	ap-southeast-2 Sydney, Australia	australia-southeast1 Sydney, Australia	syd01 Sydney, Australia

Table 1: Common cloud provider hosting regions evaluated in the report



BROADBAND MEASUREMENTS

Broadband Measurements Data Collection Period: 09/10/2019 - 10/10/2019

Apart from testing to the cloud regions from the 98 locations listed above, a separate subset of measurements was gathered from agents connected to broadband ISP providers in the United States. Network performance metrics were gathered every 10 minutes from six broadband ISPs including AT&T, Verizon, Comcast, CenturyLink, Cox and Charter, from six cities (Ashburn, Chicago, Dallas, Los Angeles, San Jose and Seattle) in North America. As the tested broadband providers are located in North America, measurements from broadband-connected agents were limited to cloud hosting regions in North America to emulate the most realistic user scenarios.

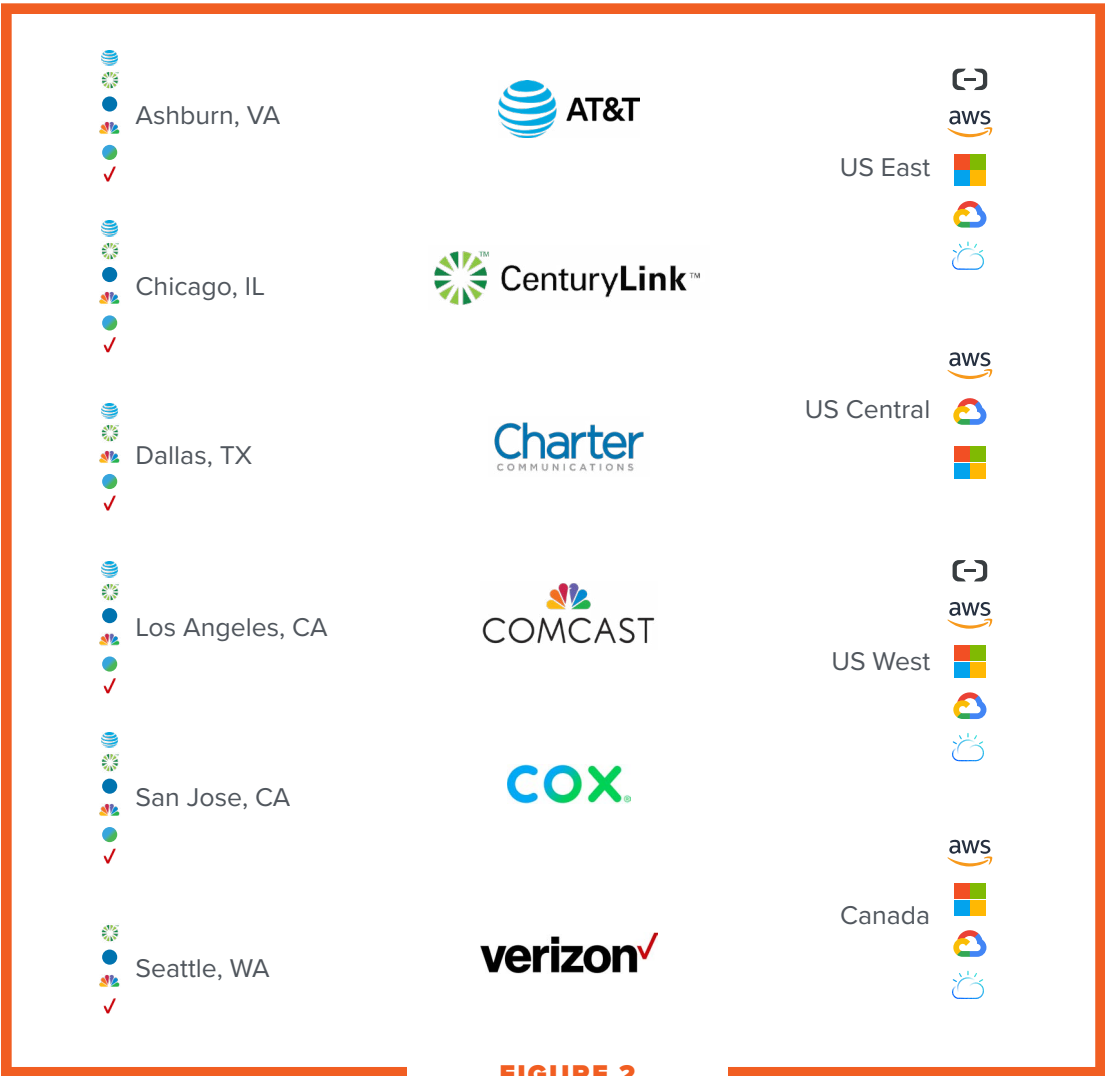


FIGURE 2
Broadband measurements were taken from six US cities (on the left) to North American hosting regions across the five cloud providers (on the right)

INTER-AZ AND INTER-REGION MEASUREMENTS

Inter-AZ and Inter-Region Data Collection Period: 09/10/2019 - 10/10/2019

As Availability Zones (AZ) and Regions are within a cloud provider, measurements in this category are limited to a single public cloud provider. Inter-AZ network performance metrics were collected every 10 minutes from 6 AWS regions, 6 GCP regions, 4 Azure regions, 7 Alibaba Cloud regions and 4 IBM Cloud regions. Since Availability Zones are assigned on a per account basis, multiple AZ pairs were analyzed to ensure ample coverage. Average inter-AZ latency metrics per provider are presented to assess relative performance between the five cloud providers.

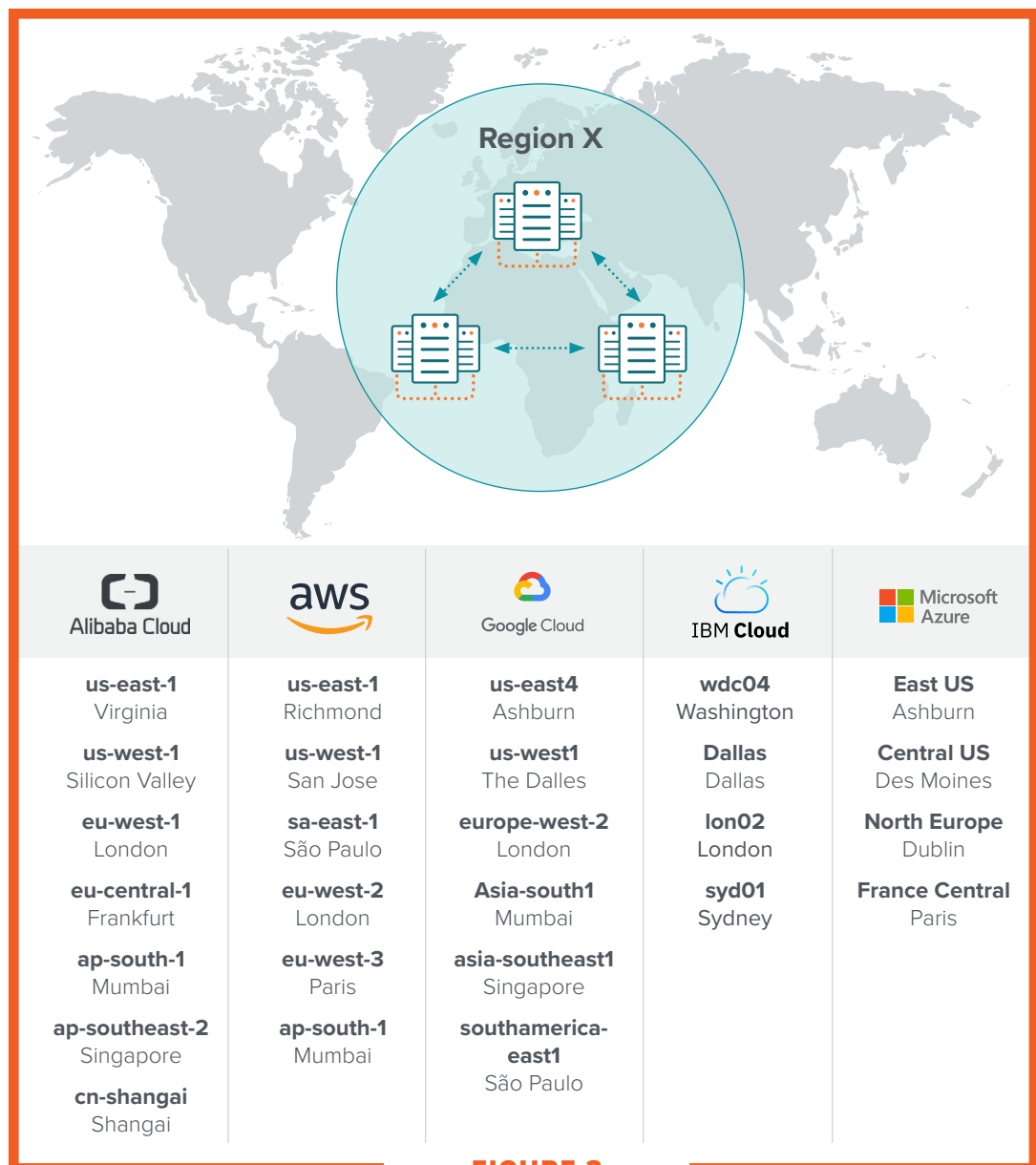


FIGURE 3

Cloud provider regions where inter-AZ network latency is measured

Inter-region measurements are also limited to individual cloud providers. Results are compared with baseline measurements taken from vantage points as near as possible to, but outside of, each cloud provider's regional data centers. These external locations will be specific to each cloud-provider. These metrics allow comparing inter-region latency with typical Internet latency rather than other cloud providers. As it is not always possible to match baseline locations to the exact location of the cloud provider's data center, round-trip latency measurements in a very few cases might need to be normalized.

To avoid large matrix combinations, we limited the number of inter-region combinations to 15 for AWS, Azure and GCP, 14 for IBM Cloud and 13 for Alibaba Cloud.

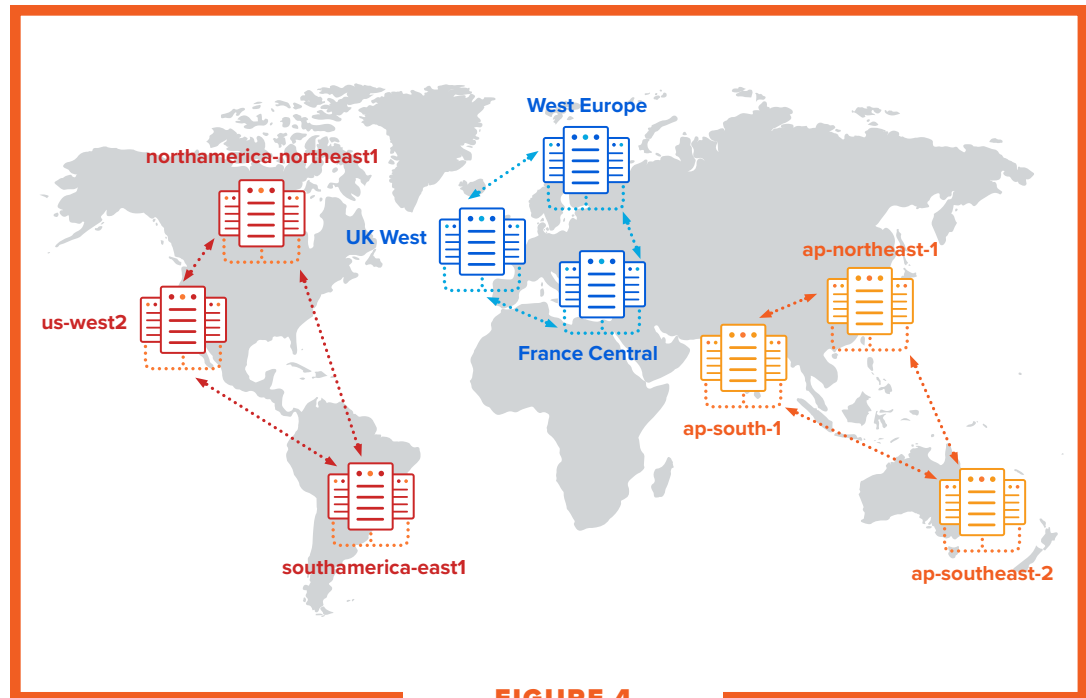


FIGURE 4

Inter-Region tests are within a single cloud provider

AMAZON GLOBAL ACCELERATOR MEASUREMENTS

Global Accelerator Data Collection Period: 10/05/2019 - 11/01/2019

Announced in November 2018, AWS Global Accelerator (GA) is a networking service that improves the availability and performance of applications hosted in AWS regions. By default, AWS does not anycast public routes associated with their regions from global edge locations, resulting in traffic being forced across the public Internet to their regions. The AWS Global Accelerator is a commercially available service that modifies this behavior by leveraging the AWS global backbone.

The report tests the difference in network performance (latency, jitter) between the default connectivity path to AWS regions and while using the AWS Global Accelerator. At the time of writing the report, the Global Accelerator is available from 14 of their regions. In this edition of the report, we compare the Global Accelerator performance for 5 AWS regions from 38 global locations, as seen in Figure 5. As with all our other tests, we gather performance metrics every 10 minutes.



FIGURE 5

Amazon Global Accelerator

MULTI-CLOUD CONNECTIVITY

Multi-cloud connectivity patterns were detected by testing to a subset of regions across all five providers. Given the scale of the test, with 95 hosting regions globally, multi-cloud performance metrics would have been unwieldy and hard to interpret. Sample tests across the providers were used to understand connectivity and peering patterns between the providers.



FIGURE 6

Multi-Cloud Connectivity

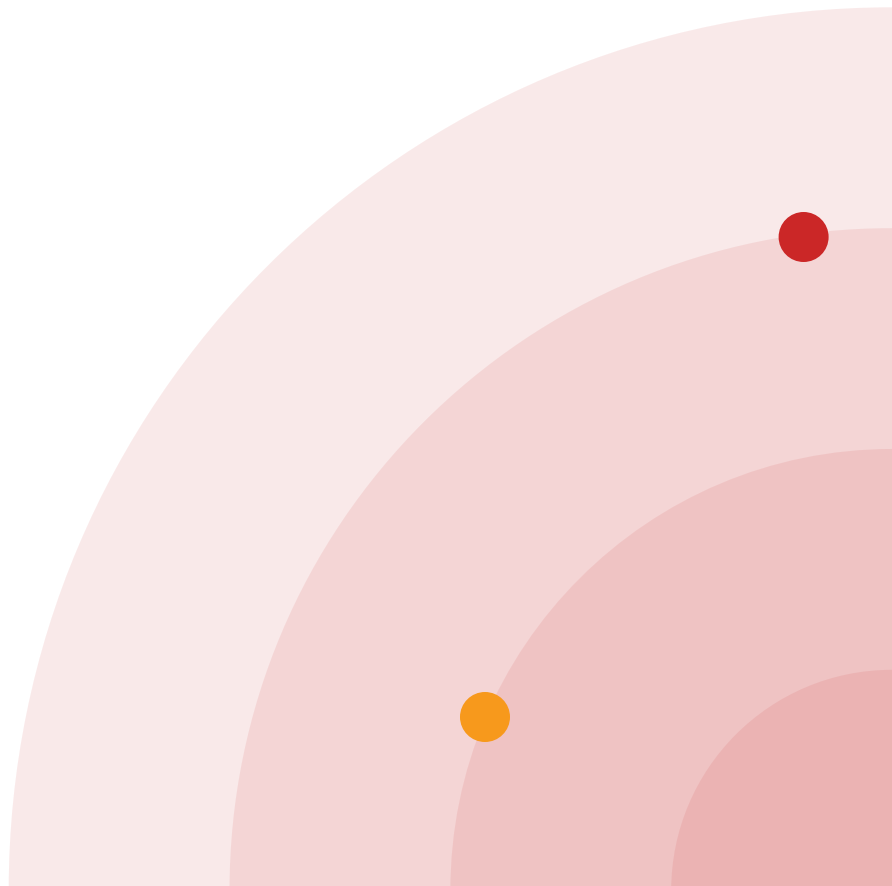
IMPORTANT DISCLAIMERS

Active performance measurements in this report are collected by ThousandEyes vantage point agents that are connected to Tier 2, Tier 3 and broadband ISPs, as well as cloud backbone networks in the case of agents in cloud provider hosting regions. Measurements taken from other locations or ISP connections in similar locations may yield different results. This highlights the importance and complexity of geo-location and network peerings as factors in network performance on a global basis.

The results presented in this report have been gathered during the timeframe mentioned in the methodology section. It does not reflect any changes made by cloud providers after the data collection period. Measurements taken before or after the documented time frames could yield different results, as there is no steady state in the cloud, which is why it is critical to continuously monitor and measure the cloud for changes.

ThousandEyes vantage points are used by hundreds of the world's largest enterprises, financial institutions, cloud and SaaS providers to actively monitor and provide real-time business and operational insights. ThousandEyes visibility data is trusted to automate service path remediation for large-scale cloud services. Vantage point agents and monitoring methodologies are continuously optimized for accuracy.

Enterprises looking to establish their specific performance baselines and operational metrics should utilize the data in this report as a guide and collect performance measurements from their own data center, office and VPC locations.



FINDINGS AND TAKEAWAYS

The Cloud Performance Benchmark provides a unique, unbiased third-party and metric-based perspective on cloud performance and cloud monitoring as it relates to both end-user experience and back-end application architecture. In this section, we discuss the in-depth findings from the study and provide recommendations for enterprises to consider when planning their cloud strategies.



2018 VS 2019 COMPARISON



UNDERSTANDING CLOUD CONNECTIVITY ARCHITECTURES



GLOBAL END USER NETWORK PERFORMANCE



NETWORK PERFORMANCE IN CHINA



BROADBAND ISP PERFORMANCE



INTER-REGION PERFORMANCE



INTER-AZ PERFORMANCE








AWS GLOBAL ACCELERATOR



MULTI-CLOUD CONNECTIVITY

2018 VS 2019 COMPARISON

In 2018, we examined cloud performance data across three public cloud providers—AWS, Azure and GCP—during a period of four weeks. The study found that there were significant architectural differences between the three providers that resulted in performance inconsistencies (related to latency, loss and jitter) depending on where in the world an end user was located. Because there is no steady state in the cloud, and because these providers are constantly making optimizations to their networks, a key question we strove to answer was: how have these cloud providers improved or changed over the past year? In this section, we will discuss the key differences we observed from 2018 to 2019.

	FINDING	EVIDENCE
	Cloud providers autonomously make architectural decisions that can impact the level of visibility into their internal networks, highlighting the ephemeral state of the cloud.	Google Cloud's network topology is obscured in 2019 when compared to the previous year.
	Azure and GCP saw the highest improvement in inter-AZ latencies from 2018.	GCP saw the highest improvement of 36.37% followed closely by Azure with 29.29% improvement from 2018.
	AWS improved performance predictability and network performance in Asia by optimizing peering and Internet routing to its data center in Mumbai.	Variation in network latency improved by 42.29% in Asia for AWS deployments.
	Despite a slight decrease year over year, Azure continues to lead in performance predictability in Asia when compared to the other two cloud providers.	Network latency fluctuations improved by 50% in Sydney but decreased by 31% in India.
	GCP continues to show weaker performance between Europe and India due to a lack of direct connectivity on the GCP backbone.	GCP continues to exhibit 2.5-3.0x the network latency in comparison to AWS, Azure, Alibaba Cloud and 1.75x higher than IBM from Europe to the respective cloud provider regions in India.

VARYING LEVEL OF VISIBILITY IN CLOUD PROVIDER NETWORKS

The first major difference that we observed over the past year is a significant reduction in visibility into the Google Cloud Platform (GCP) network. ThousandEyes active monitoring uses a continuous stream of probe packets with decreasing Time to Live (TTL) values to decipher the layer 3 hops along a service path. This is what enables us to understand the hop-by-hop path that users take as they traverse a network.

In our testing, it appears that GCP is modifying the TTL of its packets to ensure that it doesn't expire—which is counterproductive when gathering per hop metrics. The result of this change is that GCP regressed in the level of visibility observed through the ThousandEyes platform, particularly in the reverse path (as shown in Figures 7 and 8). This behavior was not observed to be consistent across all GCP hosting regions.

UNDERSTANDING PATH VISUALIZATION

Path Visualization traces the journey of traffic streams from source to destination, identifying hop-by-hop nodes with metrics such as path latency, forwarding loss at each node, link delays and Quality of Service (DSCP) remarkings along the way. Figure 7 represents the path from a user location on the left to the data center of the cloud provider and vice versa. Path Visualization shows the Layer 3 nodes along the path with detailed information on IP address, geo-located node location along with the Autonomous System Number (ASN) of the ISP network. Light blue nodes typically represent Internet path while teal nodes represent the network of the destination.

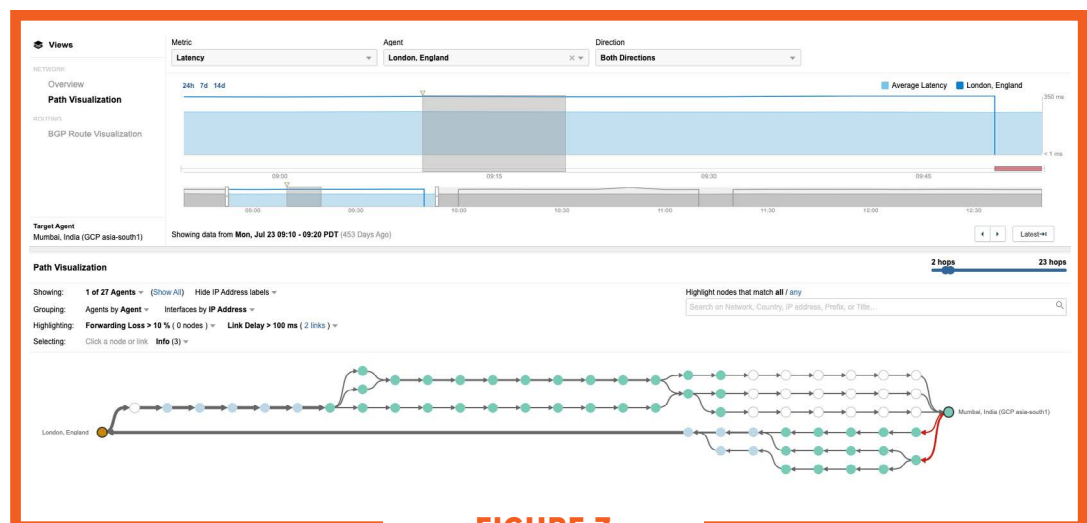


FIGURE 7

Reverse path from GCP India to London in 2018 reveals the connectivity path and layer 3 hops between the source and destination

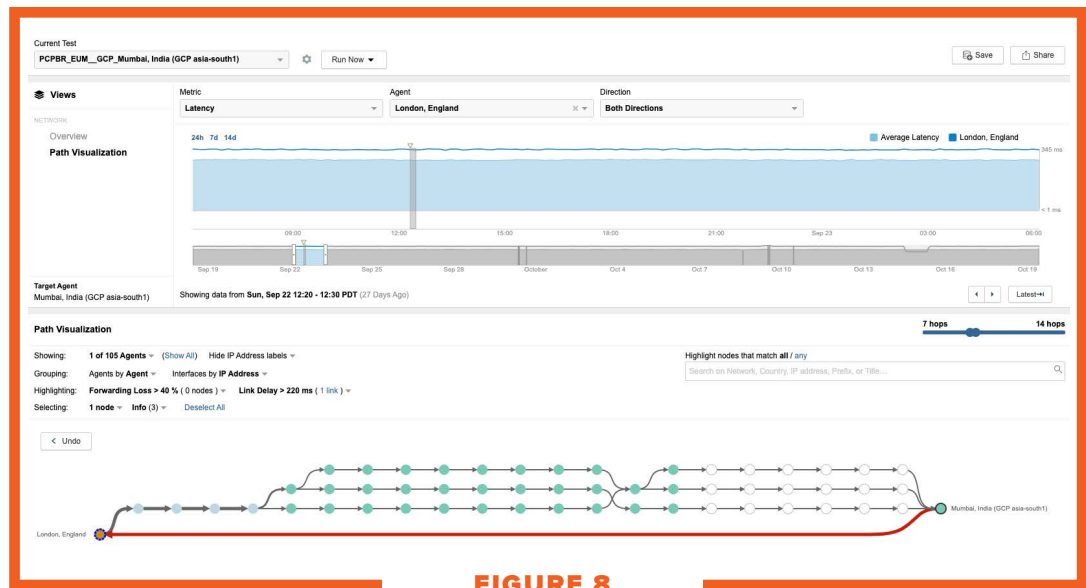


FIGURE 8

Reverse path from GCP India to London is seen as a single hop in 2019

While this type of change may be imperceptible from a performance perspective, the loss of granular visibility into the layer 3 hops can make it difficult to diagnose and resolve performance issues, if they happen, resulting in higher MTTR for cloud issues. Note that GCP's TTL modification doesn't just impact visibility within the boundaries of the GCP network, but through any hop on the reverse path. As Internet and cloud outages have become increasingly damaging, loss of visibility can ultimately have detrimental impacts on the digital experience of your customers and employees. This example with GCP demonstrates how a seemingly innocuous, behind-the-scenes change made by a public cloud provider within their network can have consequential impacts on the enterprises they serve.

IMPROVEMENTS TO INTER-AZ LATENCIES

When we analyzed bi-directional network latency, all three public cloud providers—AWS, Azure and GCP—showed an improvement in inter-AZ latency when compared to the 2018 results (as shown in Figure 9). The results revealed that GCP performed the best, with an overall average improvement in latency across global regions of 36.37%, and Azure followed closely with a 29.29% improvement. AWS, however, showed only marginal improvement in latency—less than 1% YoY. Bear in mind that absolute inter-AZ latency numbers for all three providers are strong, in the low to fractional milliseconds range. For 2019 inter-AZ latencies across providers and regions, please consult Figure 44.

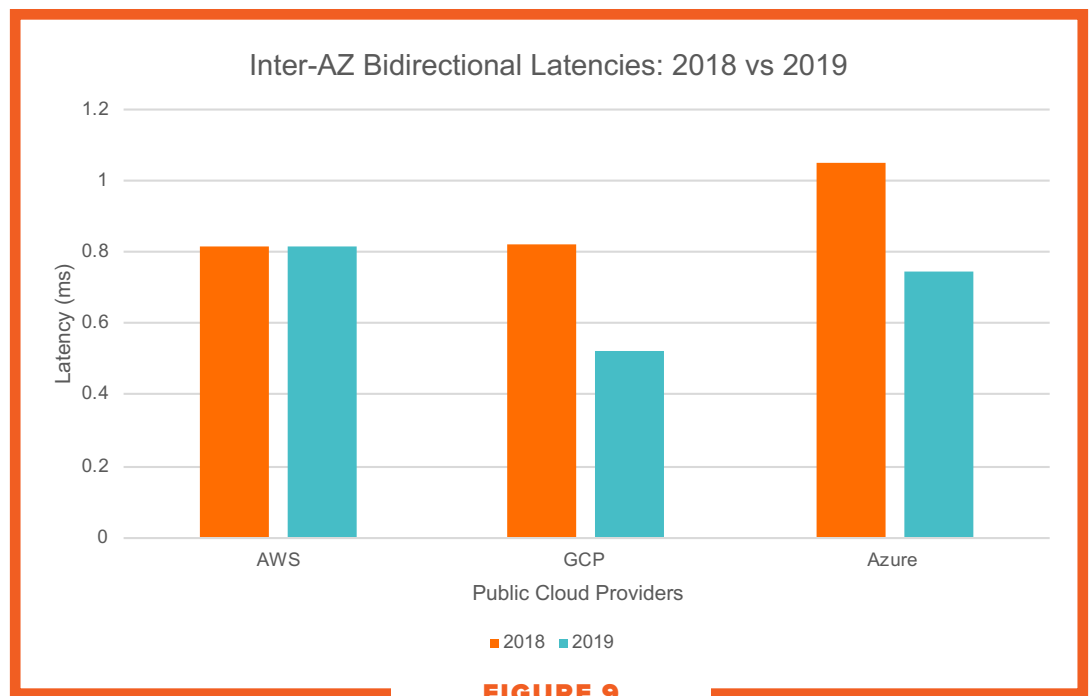


FIGURE 9

YoY improvements in Inter-AZ latencies across the three cloud providers

AWS IMPROVES PERFORMANCE PREDICTABILITY IN ASIA

The 2018 report highlighted a significant variation in bidirectional network latency (computed through standard deviation from mean latency in a four-week period) that reflected on the performance predictability of connections for users in Asia to AWS' hosting region in Mumbai (ap-south-1) as seen in Figure 10. The vertical black lines are a measure of the standard deviation of latency, in other words how far from the mean did latency measurements swing by. The higher the variation in latency, the lower the performance predictability of the end-to-end connection.

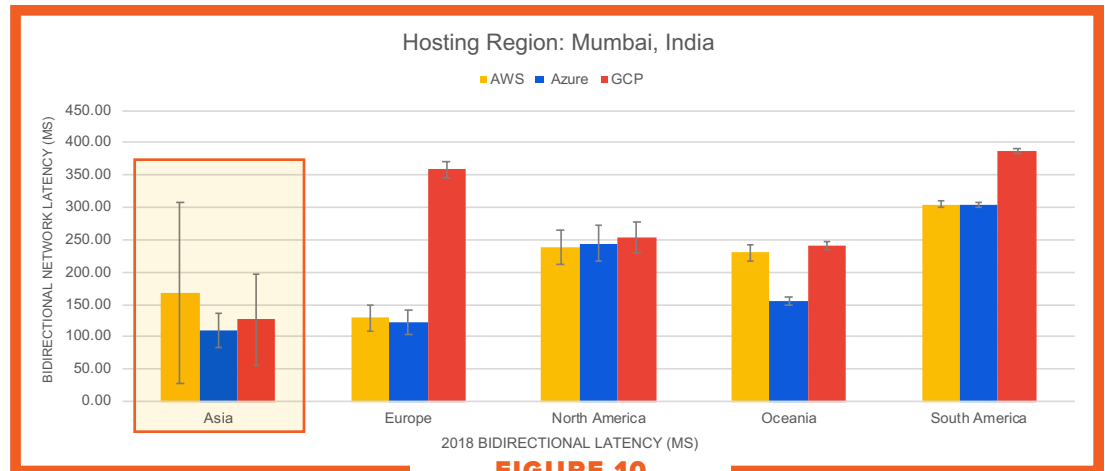


FIGURE 10

Users in Asia connecting to AWS' hosting region in Mumbai (ap-south-1) experienced a large swing in bidirectional latency in 2018

In 2019, we found that while AWS' connectivity architecture has not changed and that it still heavily relies on the Internet, there have been improvements in both network latency and performance predictability (Figure 11) between Asia and Mumbai, India. The results from our testing show that network latency variations improved by 42.29% YoY.

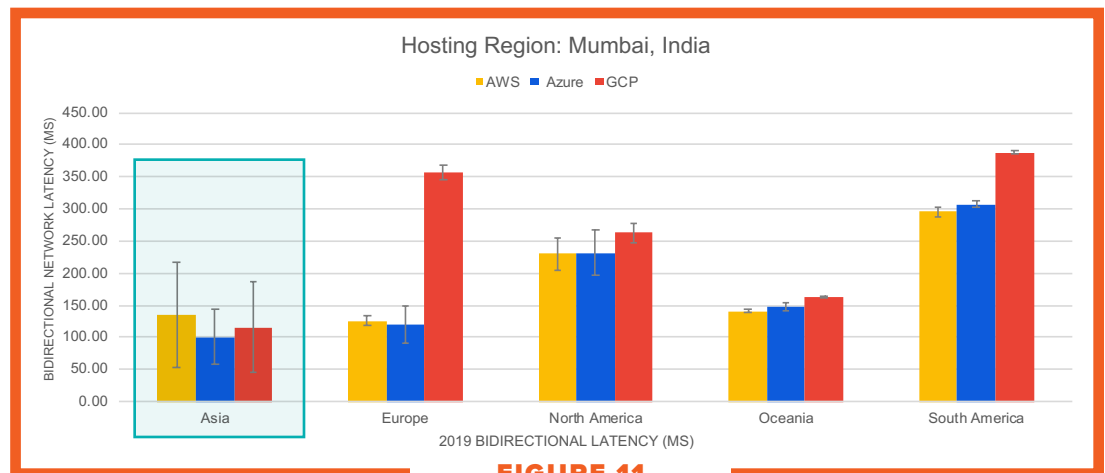


FIGURE 11

Users in Asia experience reduced latency variation while connecting to AWS Mumbai (ap-south-1) region in 2019

By comparing the network path from a few user locations in Asia to the Mumbai (ap-south-1) region, we observed that AWS has optimized its peerings. This change benefits AWS users as they are subject not only to lower variations in latency but also a much faster network path.

Figure 12 below shows the network path from users in Seoul and Singapore connecting to AWS ap-south-1 in 2018. Note that the path visualizations below only highlight the forward path for readability. Notice how traffic from Seoul goes across the Internet around the world, to the United States and Europe before entering AWS network in Mumbai. Users from Singapore also traversed multiple hops through NTT's network before entering AWS.

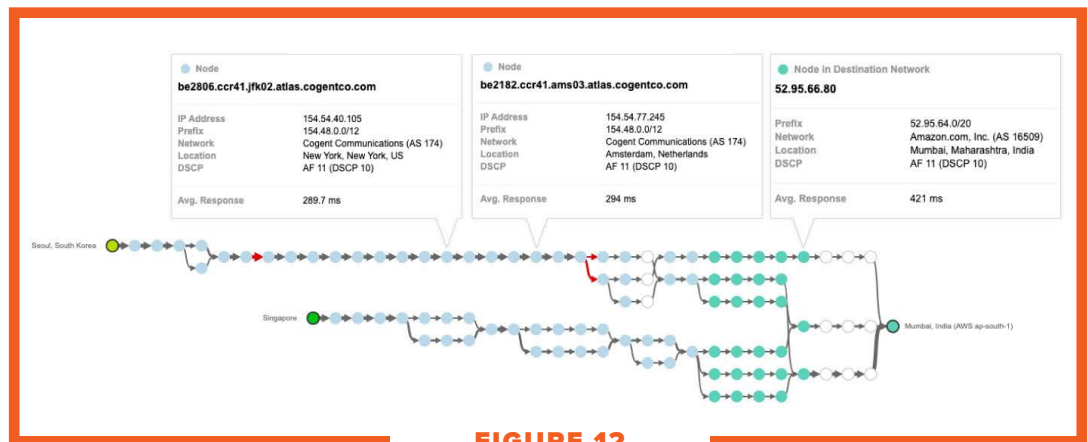


FIGURE 12

Traffic path from Seoul and Singapore to AWS ap-south-1 (Mumbai) in 2018

In 2019, the network paths from the same user locations are more optimized, as seen in Figure 13 below. User traffic avoids a round the world trip, stays in Asia and enters the AWS backbone sooner through an Equinix interconnect in Singapore.

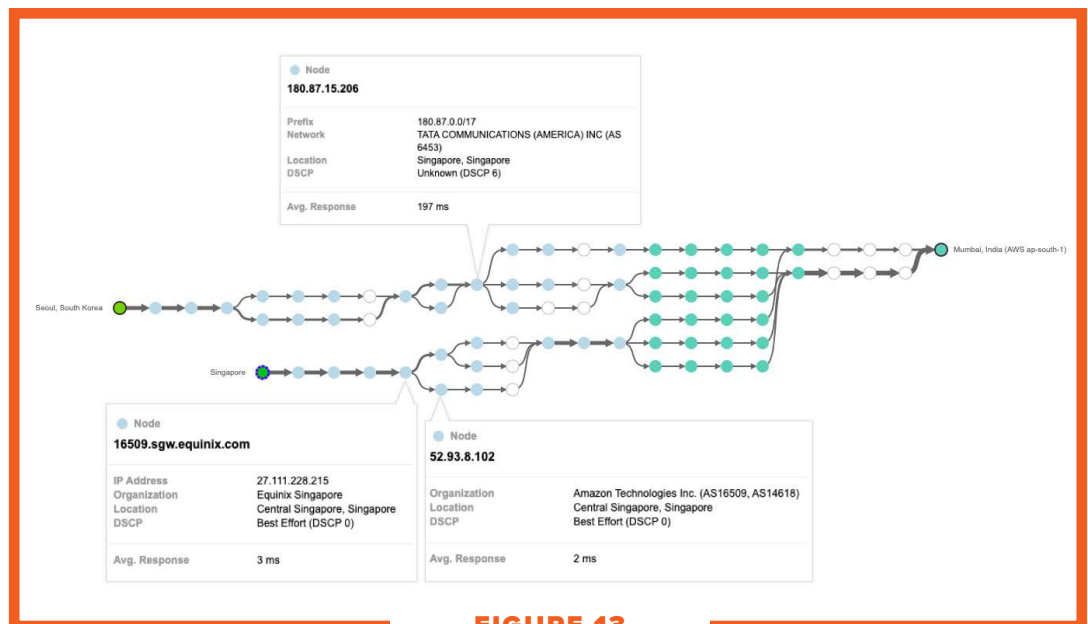


FIGURE 13

Traffic path from Seoul and Singapore to AWS ap-south-1 (Mumbai) in 2019

From 2018 to 2019, latency from Seoul to the AWS Mumbai region improved from 235 ms (Figure 14) to 140ms (Figure 15).

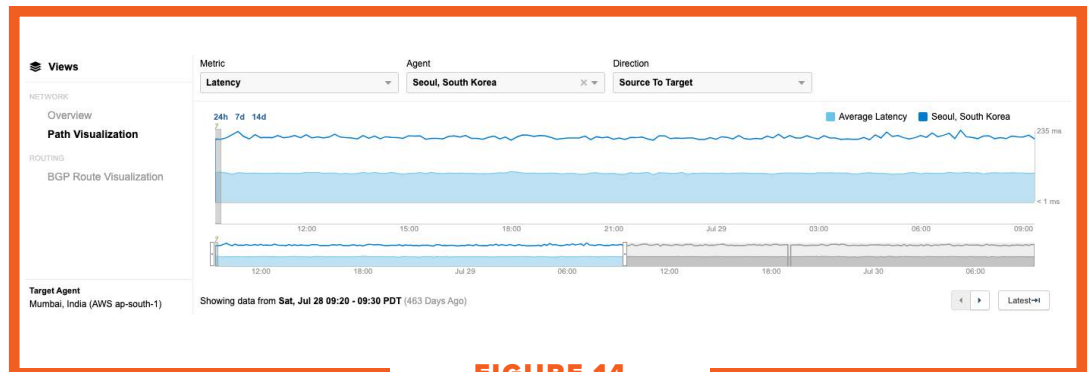


FIGURE 14

Forward path network latency from Seoul in 2018

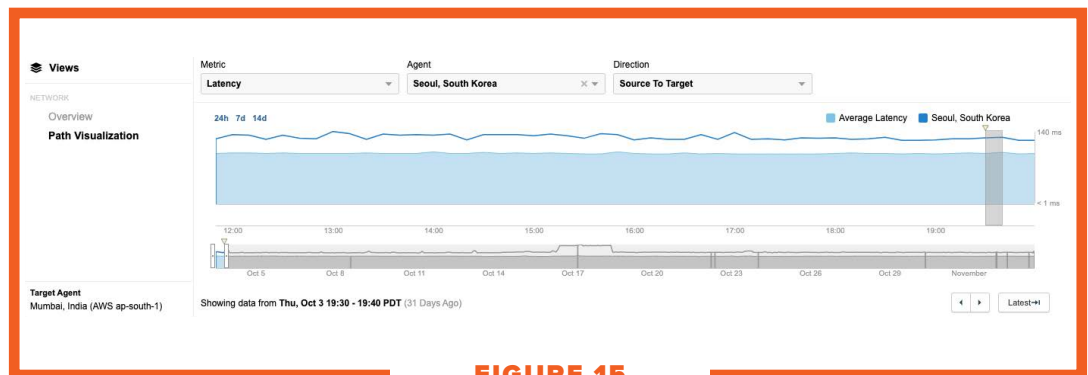


FIGURE 15

Forward path network latency from Seoul in 2019

We saw similar improvements for other AWS regions in Asia. However, despite that improvement, AWS continues to demonstrate the largest standard deviation in latency in Asia, as shown in Figure 16—likely a by-product of AWS' heavy reliance on the Internet to route user traffic.

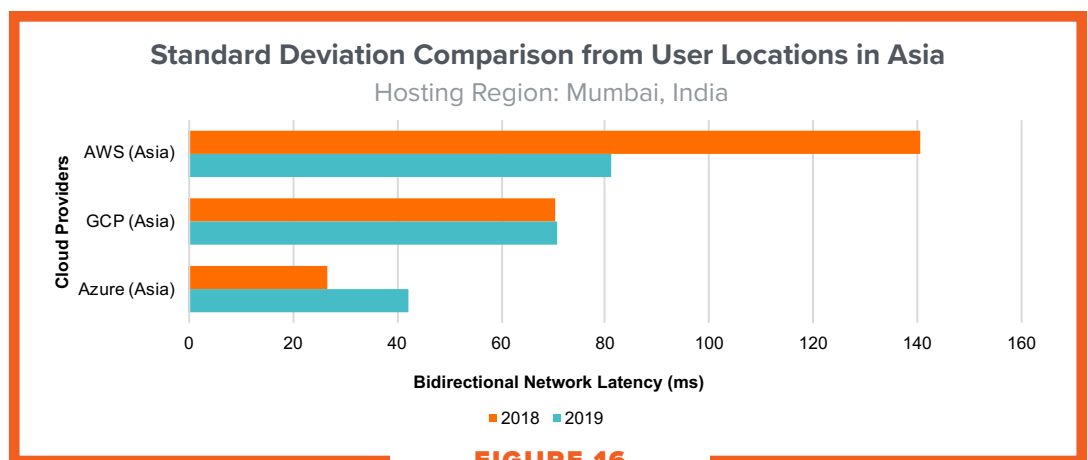


FIGURE 16

AWS improved in standard deviation measurements from 2018 to 2019, but still lags behind Azure and GCP in Mumbai and Asia generally

AZURE END-USER PERFORMANCE

Microsoft Azure continues to use its global private backbone to move traffic from user locations to its geographically diverse regions. Year over year, we noticed that Azure's performance predictability in some hosting regions improved. For instance, in Sydney, the performance variation over a four-week period improved by 50%. However, in other regions, such as India, performance predictability decreased by 30%.

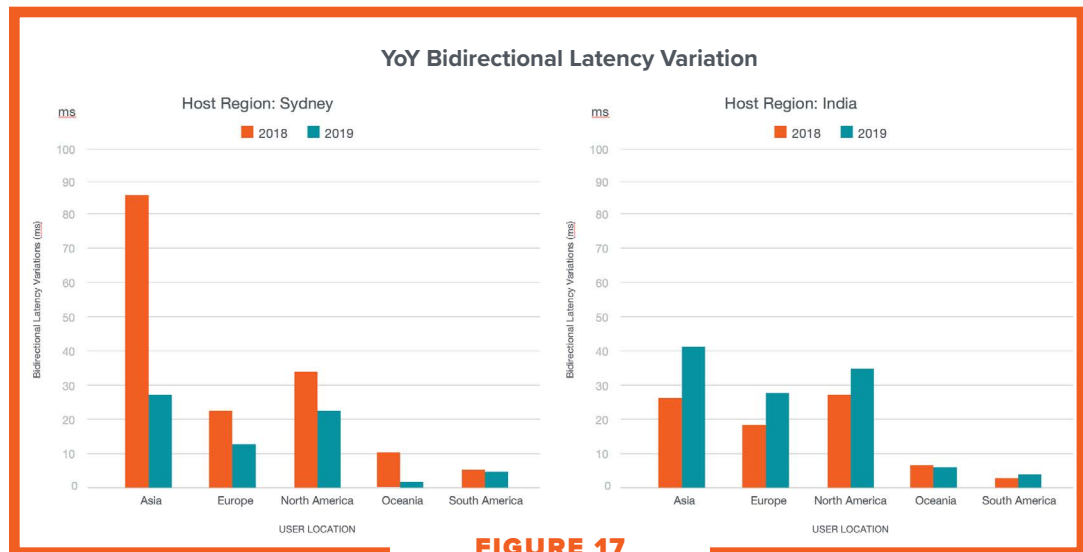


FIGURE 17

End-to-end latency variation improves by 50% in Sydney, but reduces by 30% in India

GCP'S CLOUD BACKBONE CONTINUES TO IMPACT EUROPE & INDIA

Users in Europe are subject to 2.5-3x the network latency while accessing compute engine workloads hosted in GCP's asia-south1 region in Mumbai, India. This pattern was initially observed in the 2018 Cloud Performance Benchmark report and attributed to the lack of connectivity on the GCP backbone between Europe and India (as seen in Figure 18). The lack of connectivity resulted in traffic from Europe going half way around the world to reach the hosting region in asia-south1 located in Mumbai, India.



FIGURE 18

GCP's infrastructure map shows no direct connectivity between Europe and India on the GCP backbone in 2018

Our data in 2019, gathered from the same vantage points used in 2018, indicated that bidirectional latency between Europe and India remained the same YoY. Figure 19 below compares the bidirectional network latencies observed in 2018 and 2019. We saw similar results in our Inter-Region tests as well. Jump to Inter-Region Measurements to look at GCP's network performance between their regions.

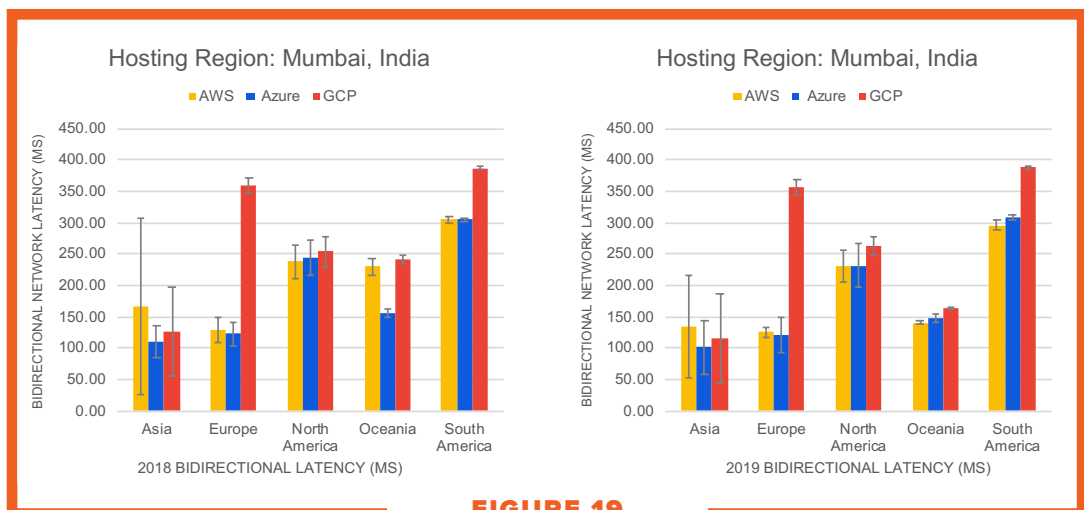


FIGURE 19

2019 measurements do not show any YoY improvements on network latency between Europe and India

While the latest update to GCP's infrastructure map indicates an update to their connectivity (observed as of November 12th, 2019) between Europe and India, the results from our end-user and GCP hosting region vantage points across Europe, and Africa did not reflect a change of routes or performance metrics that would correspond with those changes.

TAKEAWAY

There is no steady state in the cloud, and enterprises are subjected to the impacts of any architecture changes made at the discretion of public cloud providers.

RECOMMENDATION

Continuously monitor your external service provider networks. With increased visibility and awareness you can:

- Detect, triage and influence resolution of any issues that is impacting customer experience
- Keep your cloud providers accountable to service level agreements and operational responsiveness

UNDERSTANDING CLOUD CONNECTIVITY ARCHITECTURES

FINDING

Architectural and connectivity differences between the five cloud providers result in varied levels of Internet exposure.

EVIDENCE

Path visualizations indicate that traffic destined to AWS and Alibaba Cloud regions (data centers) enter their respective backbone closest to the target region. This is a marked difference from how GCP and Azure handle incoming traffic. Traffic enters the GCP and Azure backbone closest to the end-user, regardless of the destination region. IBM takes a hybrid approach to cloud connectivity, with some regions purely relying on the IBM backbone and others that primarily rely on Internet connectivity to transport user traffic to its hosting regions.

Cloud connectivity determines how users around the globe access resources deployed in the public cloud. For enterprises building their services on the public cloud, cloud connectivity architectures can directly impact the end-user experience. Deployments with an increased reliance on and exposure to the Internet are subject to greater operational challenges and risks. The Internet is a best effort medium, a constellation of networks that are vulnerable to security threats, DDoS attacks, congestion and infrastructure outages—so relying on the Internet increases unpredictability in performance, creates risk for cloud strategy and raises operational complexity.

Our analysis of network path data reveals important contrasts in cloud connectivity architectures between AWS, Azure, GCP, IBM Cloud and Alibaba Cloud, primarily around the level of Internet exposure in the end-to-end network paths. In this section, we'll take a comparative look at the most significant architectural differences and similarities between these five cloud providers.

ALIBABA CLOUD AND AWS

Network path data for Alibaba Cloud reveals a clear behavior of forcing traffic across the public Internet prior to absorbing the traffic into its backbone network. In Figure 20, you can see that traffic from a number of global locations, destined for Alibaba Cloud in Silicon Valley, CA, traverses the Internet over multiple hops before entering the provider's backbone in San Jose or San Francisco, CA—just prior to reaching the destination.

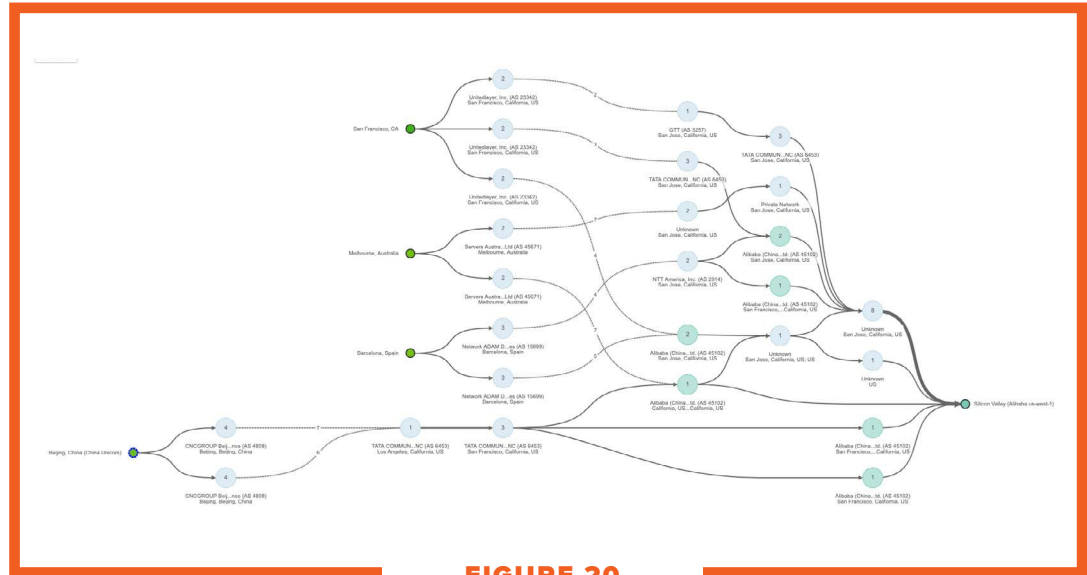


FIGURE 20

Irrespective of the user location, traffic from global locations destined to Alibaba Cloud's hosting region in Silicon Valley enters the Alibaba Cloud backbone in San Jose, CA or San Francisco, CA

We've seen a similar pattern of behavior from AWS in the past, and the same is true this year. In this example, shown in Figure 21, regardless of where users are located, the AWS network design forces traffic from the end user through the public Internet, only to enter the AWS backbone closest to the target region in Mumbai, India.

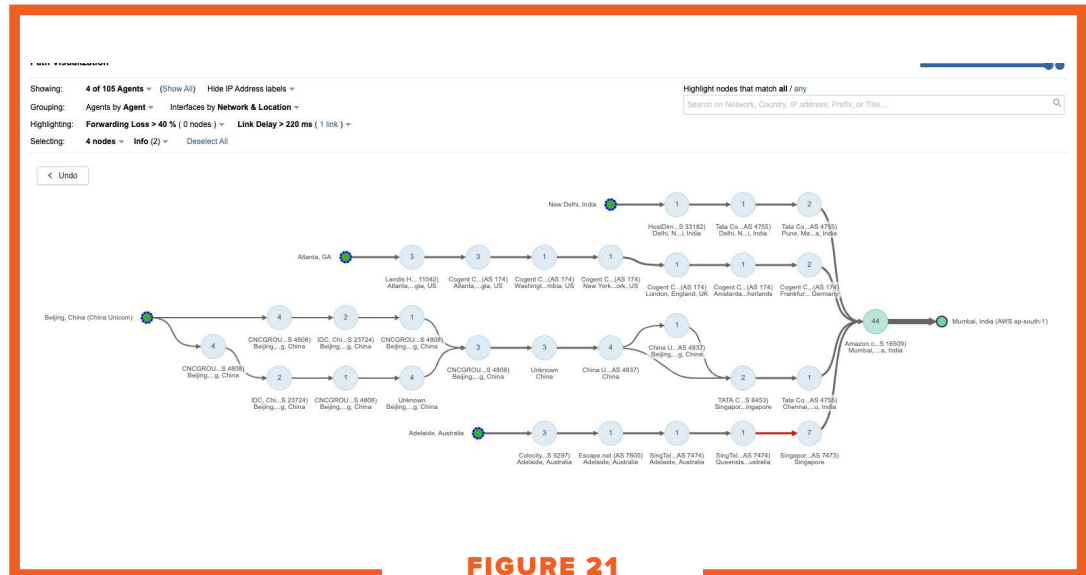


FIGURE 21

Irrespective of the user location, traffic from global locations destined to AWS' hosting region in Mumbai, India, enters the AWS backbone in Mumbai

Connectivity architectures vary between the cloud providers due to differences in global routing policies. As is the case for Amazon, AWS does not anycast public routes associated with each of their regions from global edge locations, resulting in traffic always flowing across the Internet to their regions. The resulting exposure to the Internet subjects its deployments to greater operational challenges and risks, especially in regions with less stable Internet performance, such as in Asia.

AZURE AND GCP

This behavior is in stark contrast to how Microsoft Azure and Google Cloud design their respective networks. As depicted in Figures 22 and 23, these providers absorb traffic from the end-user into their internal backbone network closest to the user, irrespective of geographical location, relying less on the Internet to move traffic between the two locations.

While normally relying on the provider's backbone results in lower latency values and thus better performance, it is sometimes the case that the absence of a direct path through the backbone results in circuitous routing and higher latency. To that end, our tests show that Google Cloud still has some significant global gaps that haven't been addressed since last year's report—notably that traffic from Europe and Africa takes 2.5-3x longer to get to India because it is routed through the GCP backbone in the US first. See the section “GCP's Cloud Backbone Continues to Impact Europe & India” for more details.

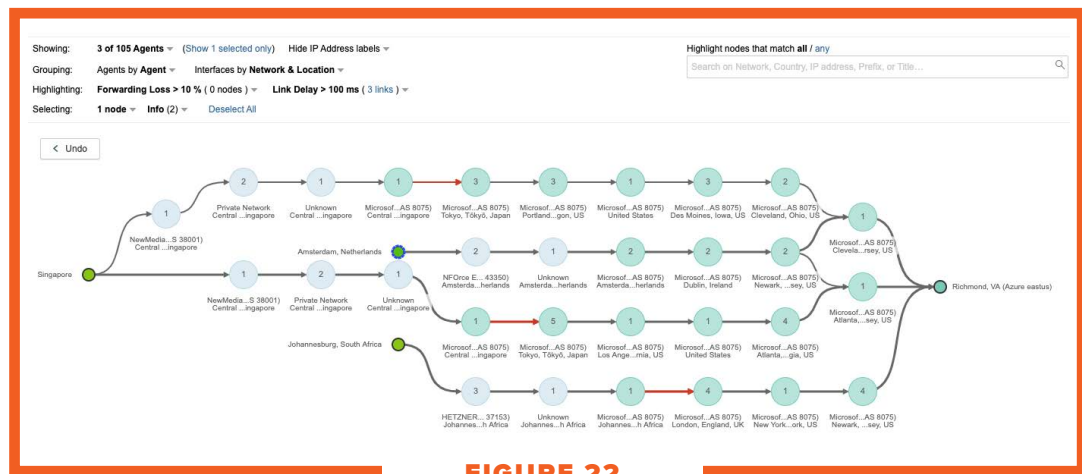


FIGURE 22

Traffic from global locations enter Azure's backbone closer to the end user

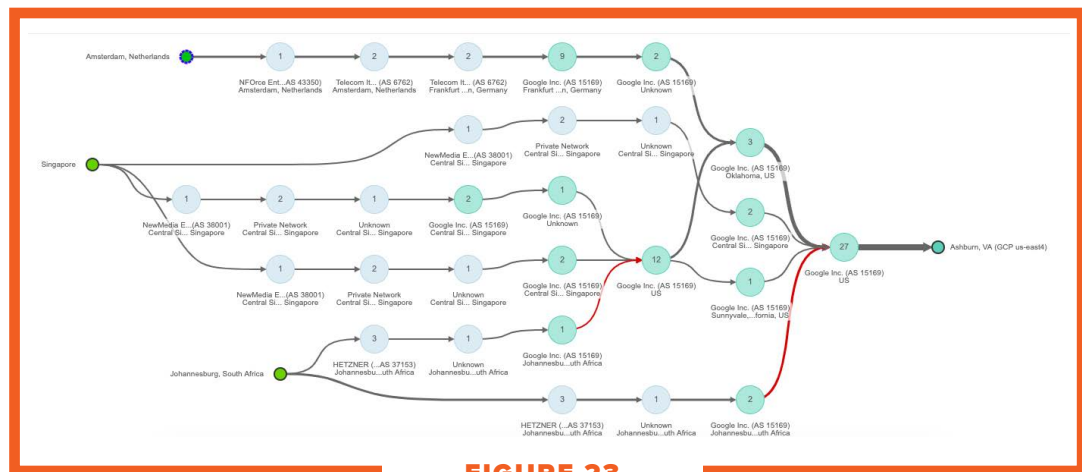


FIGURE 23

Traffic from global locations enter GCP's backbone closer to the end user

IBM CLOUD

IBM is the only cloud provider that takes a hybrid approach to cloud connectivity from users to hosting regions. Depending on the hosting region and the expanse of the Softlayer backbone, user traffic rides the Internet longer or enters the cloud provider's backbone closer to the end user. Let's look at this with an example. Figure 24 shows the network connectivity path from users in Atlanta, Singapore and Amsterdam accessing a workload in IBM's region in Chennai, India. Notice how traffic from these end locations traverse multiple ISPs and geographical regions before they enter the Softlayer network in Chennai, closest to the hosting region.

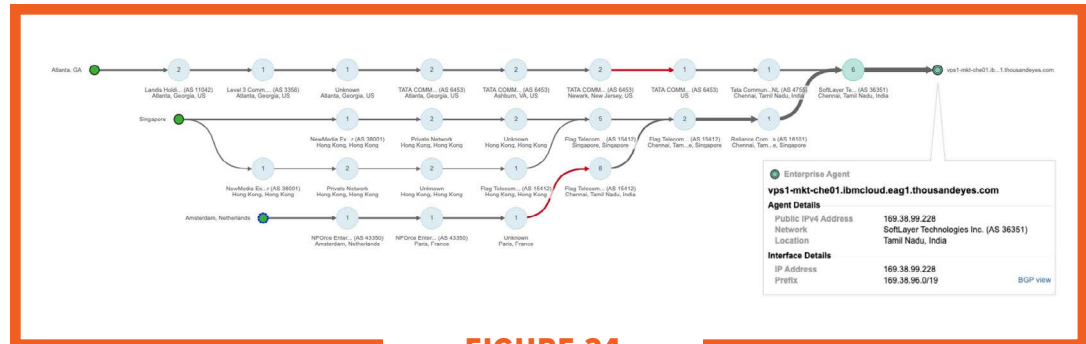


FIGURE 24

Traffic from global locations rely more on the Internet when accessing workloads in IBM's Chennai region

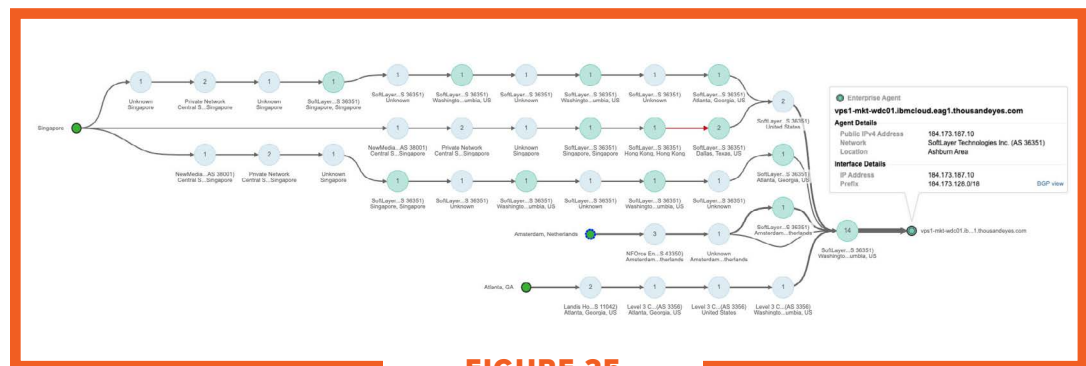


FIGURE 25

Traffic from global locations enter IBM's Softlayer backbone closer to the user location while accessing workloads in their Washington, DC region

When compared to the network path from the same locations to a hosting region in Washington, DC, as shown in Figure 25, we notice that traffic enters IBM's network closer to the end user with minimal reliance on the Internet.

THE IMPACT OF CLOUD CONNECTIVITY ON USER EXPERIENCE

Why AWS chooses to route its traffic through the Internet while the other big players use their internal backbone might have to do with how each of these service providers has evolved. Google and Microsoft have the historical advantage of building and maintaining a vast backbone network. AWS, the current market leader in public cloud offerings, focused initially on rapid delivery of services to the market, rather than building out a massive backbone network. Given their current position, increasing profitability and recent investments in undersea cables, it is likely that their connectivity architecture will change over time.

Enterprises considering a move to the public cloud should consider connectivity architectures to evaluate their appetite for risk while striking a balance with features and functionality. Enterprises should also be aware that even though public cloud backbones are each maintained by a single vendor, they are still multi-tenant service infrastructures that typically don't offer SLAs. Furthermore, public cloud connectivity architectures continuously evolve and can be subject to precipitous changes at the discretion of the provider.

While all public cloud providers rely on the public Internet to a certain extent, their level of dependence on the Internet varies greatly—and this can have downstream impacts on the enterprises they serve. Simply put, the less time spent riding the public Internet, the more reliable and stable of an experience enterprises can expect.

TAKEAWAY

AWS deployments rely on the Internet to a greater extent than Microsoft Azure or Google Cloud. For enterprises building their services on AWS, this translates into traffic spending relatively more time on the public Internet than the cloud provider's backbone.

RECOMMENDATION

Consider your organization's tolerance for exposure to the unpredictable nature of the Internet.

GLOBAL END USER NETWORK PERFORMANCE

FINDING

Despite generally consistent performance across the five cloud providers, we found important exceptions, particularly in Asia and LATAM.

EVIDENCE

The five cloud providers exhibited comparable, robust network performance across North America and Western Europe (UK, West EU), however performance exceptions surfaced in Asia and Latin America. For example:

- GCP exhibits 2.5-3x the network latency in comparison to AWS, Azure, Alibaba Cloud and 1.75x higher than IBM from Europe to regions in India
- Network latency from Rio to GCP's São Paulo hosting region is 6x compared to AWS, Azure and IBM Cloud due to a suboptimal reverse path

Choosing the right cloud provider and regions requires a data-driven approach—and that approach must take into account network performance as it relates to those who consume the service: the end users. In order to understand performance from this angle, we tested bi-directional network latency from global user locations to various geographical regions (data centers) of the five cloud providers.

While we found that all of the cloud providers exhibited generally comparable performance in the US and Europe, there are certainly wide regional discrepancies that exist. The graphs in Figure 26 and 27 represent the mean bidirectional latency from different continents to the hosting regions of the five cloud providers in Virginia, US, and the United Kingdom respectively. As evident from the data, latency is comparable across all five providers. The fluctuations in network latency, measured by standard deviation, is denoted by the vertical dark lines within each measurement. As one would expect, the parent continent where the data center is located will experience the minimum latency as traffic within the parent continent vantage points need to only traverse a short path to reach the data centers.

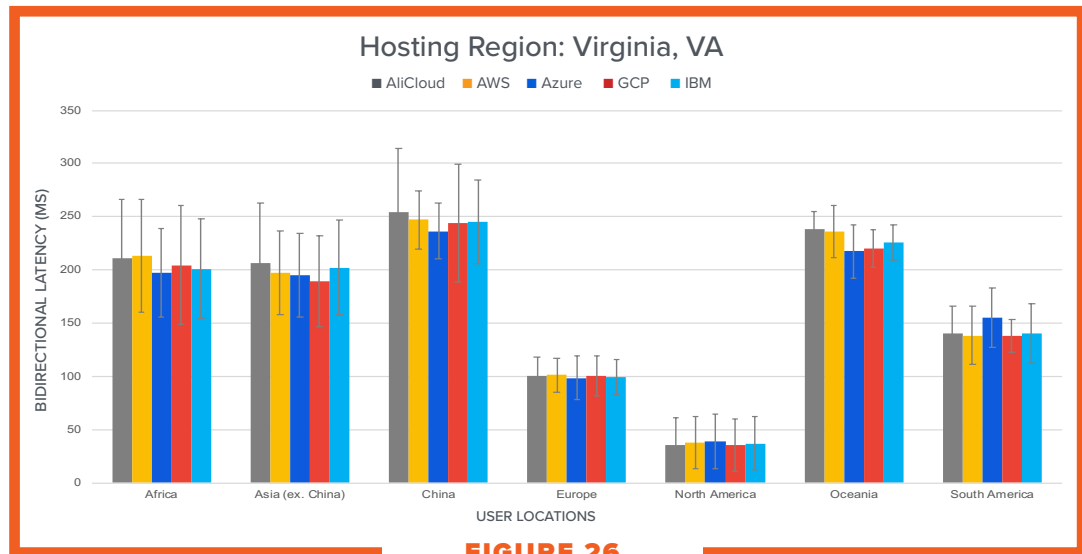


FIGURE 26

Bi-directional network latency between global user locations and the Virginia data centers (regions) of all five public cloud providers

When connecting to hosting regions in India from Europe, our tests revealed a distinct discrepancy for GCP users, in particular. Figure 27 represents the mean bi-directional latencies while connecting to data centers located in India from global regions of all five public cloud providers. Of note, it appears that Google Cloud exhibits 2.5x the network latency in comparison to AWS, Azure, Alibaba Cloud and 1.75x higher than IBM Cloud from Europe to regions in Mumbai, India and Chennai, India. Similarly, GCP users from Africa generally experience higher latency when connecting to its data center in India.

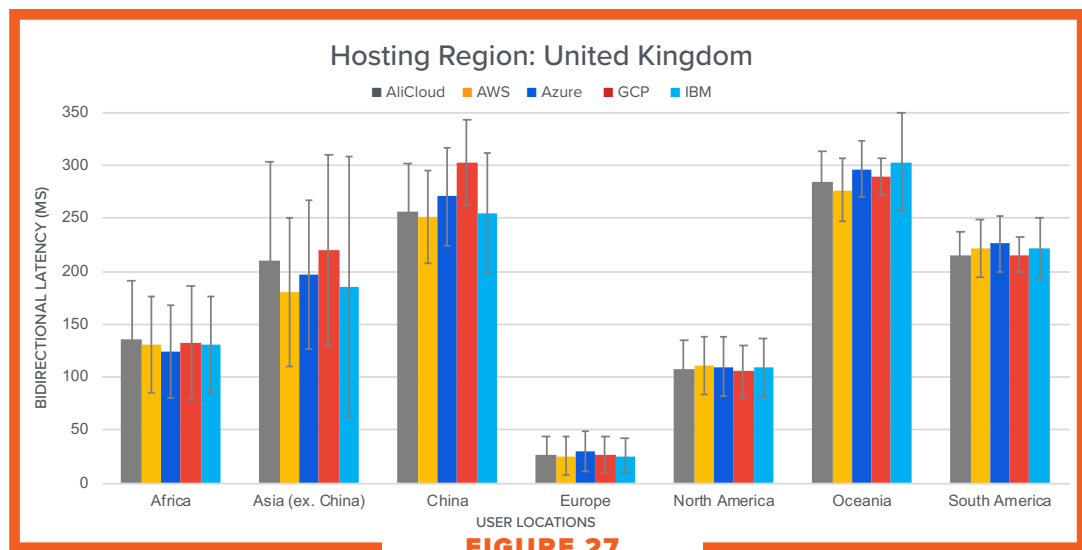


FIGURE 27

Bi-directional network latency between global user locations and the UK data centers (regions) of all five public cloud providers

Despite the generally consistent performance noted above, we found important exceptions in network latency, especially in geographical regions such as Asia and LATAM. These performance variations highlight the reality that public cloud vendors do not yet have uniform or consistent performance globally. Enterprises considering cloud deployments across multiple geographic regions can use these data points to inform their architectural decisions.

REGIONAL EXCEPTION 1: INDIA

When connecting to hosting regions in India from Europe, our tests revealed a distinct discrepancy for GCP users, in particular. Figure 28 represents the mean bi-directional latencies while connecting to the India data centers of all five cloud providers from global user vantage locations. Of note, it appears that Google Cloud exhibits 2.5-3x the network latency in comparison to AWS, Azure, Alibaba Cloud and 1.75x higher than IBM Cloud from Europe to regions in Mumbai, India and Chennai, India. Similarly, GCP users from Africa generally experience higher latency when connecting to their data center in India.

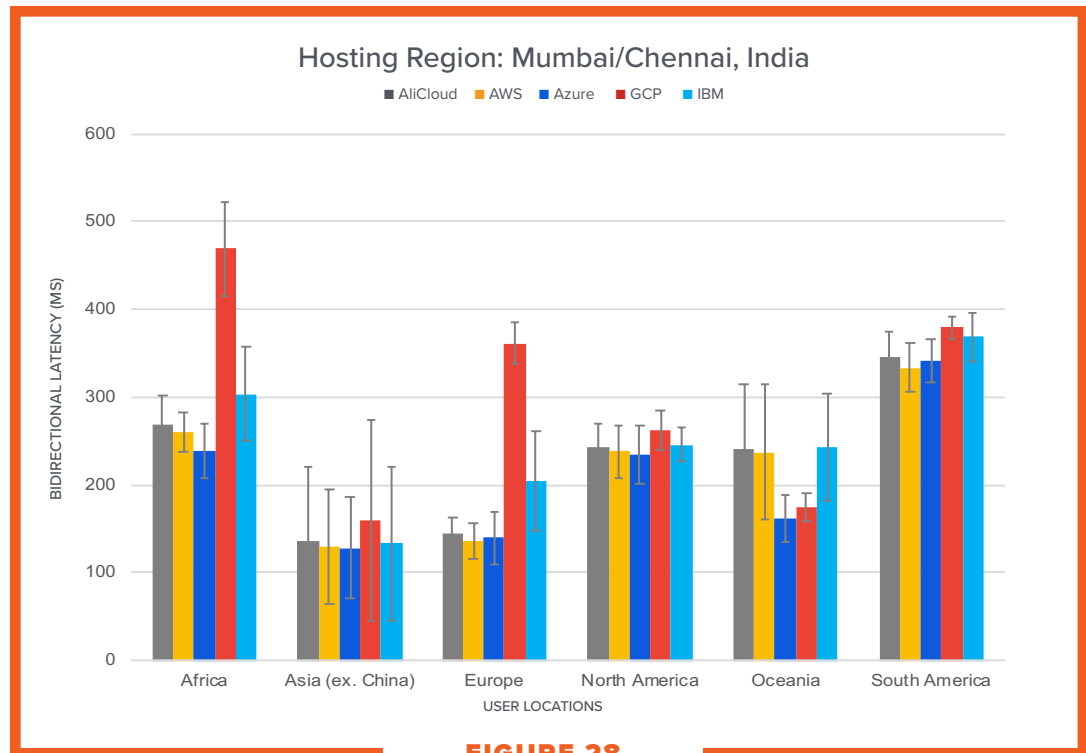


FIGURE 28

Bi-directional network latency between global user locations and the India data centers (regions) of all five public cloud providers

By examining the network path, Figure 29, of traffic originating in Spain, England and South Africa that is destined for India, we can begin to understand why GCP exhibits higher latency between these regions. As you can see, users originating in these three locations transit through Google's backbone in the US before reaching the destination in Mumbai. Of course, the circuitous route between these locations is not the most direct, and it results in much higher network latency with the potential to affect users connecting to workloads in GCP's hosting region in Mumbai, India.

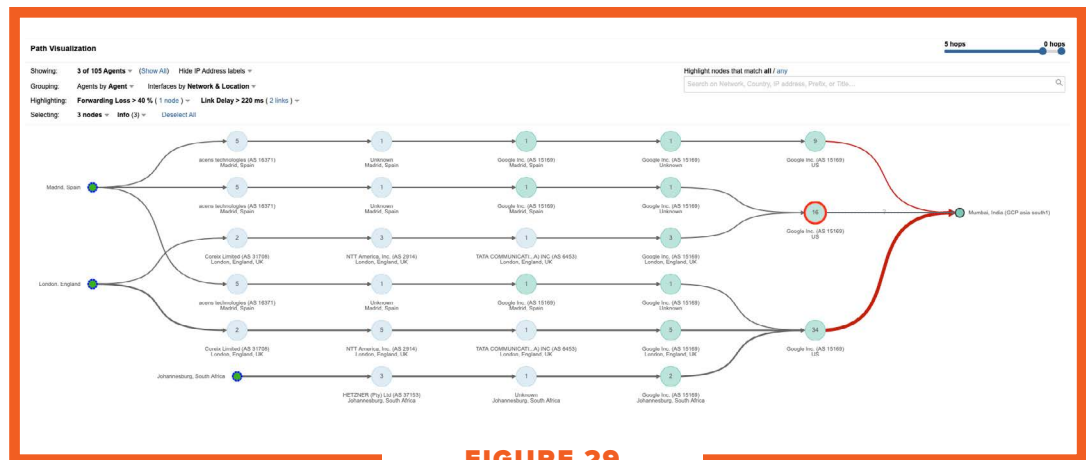


FIGURE 29

Traffic from Europe and Africa traverse a circuitous route through the GCP backbone via the United States to reach India

In contrast, Microsoft Azure users connecting from Europe and Africa to hosting regions in India follow a more optimal route. In Figure 30, you can see that users originating in Spain, England and South Africa destined for India enter the Microsoft backbone much closer to the end user before transiting to its destination. This results in much lower end-to-end latency for users accessing workloads hosted in Microsoft Azure's hosting region in Mumbai from Europe.

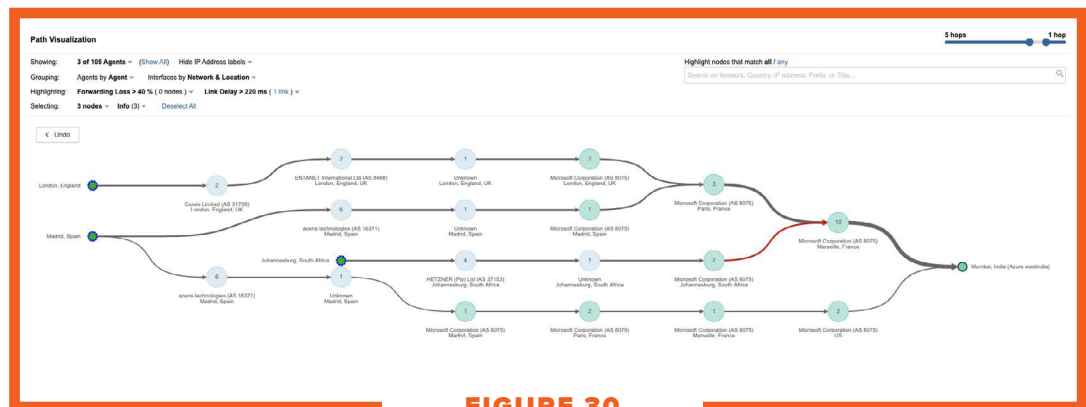


FIGURE 30

Users on Azure experience a more optimal network path from Europe and Africa to India

REGIONAL EXCEPTION 2: SINGAPORE

Global user performance to the Singapore data centers of all five cloud providers is more of a mixed bag but exhibits interesting differences in performance. For example, when connecting to services in GCP's Singapore data center, users from Africa, Asia and Europe experience higher latencies when compared to the other four cloud providers. It appears traffic from our vantage points was taking a circuitous route to reach Singapore from Africa and Europe. However, if you are serving customers in China from a Singapore hosting region, Alibaba Cloud shows the best latency while IBM is almost 3x slower in the same geography.

BIDIRECTIONAL LATENCY (MS)	CLOUD PROVIDERS				
	ALIBABA	AWS	AZURE	GCP	IBM
AFRICA	381.78	297.20	286.96	409.36	290.14
ASIA EX. CHINA	90.76	96.59	102.29	140.49	108.57
CHINA	65.69	111.96	97.63	129.63	188.10
EUROPE	250.76	232.17	196.68	301.58	196.87
NORTH AMERICA	225.50	226.40	222.01	203.58	223.23
OCEANIA	164.76	133.08	101.45	114.17	102.63
SOUTH AMERICA	354.05	358.33	366.88	320.78	357.08

Table 2: Customers in China from a Singapore hosting region, Alibaba Cloud shows the best latency while IBM is almost 3x slower in the same geography.

REGIONAL EXCEPTION 3: SOUTH AMERICA

While testing from global locations to cloud provider regions in São Paulo, we noticed that traffic from our vantage point in Rio experienced high latency while accessing workloads in GCP's hosting region in São Paulo (southamerica-east1). As seen in the graph in Figure 31 below, GCP experienced 6x latency compared to the other three providers. Note that Alibaba Cloud does not have a hosting region in São Paulo, so our comparisons in this region were between AWS, Azure, GCP and IBM.

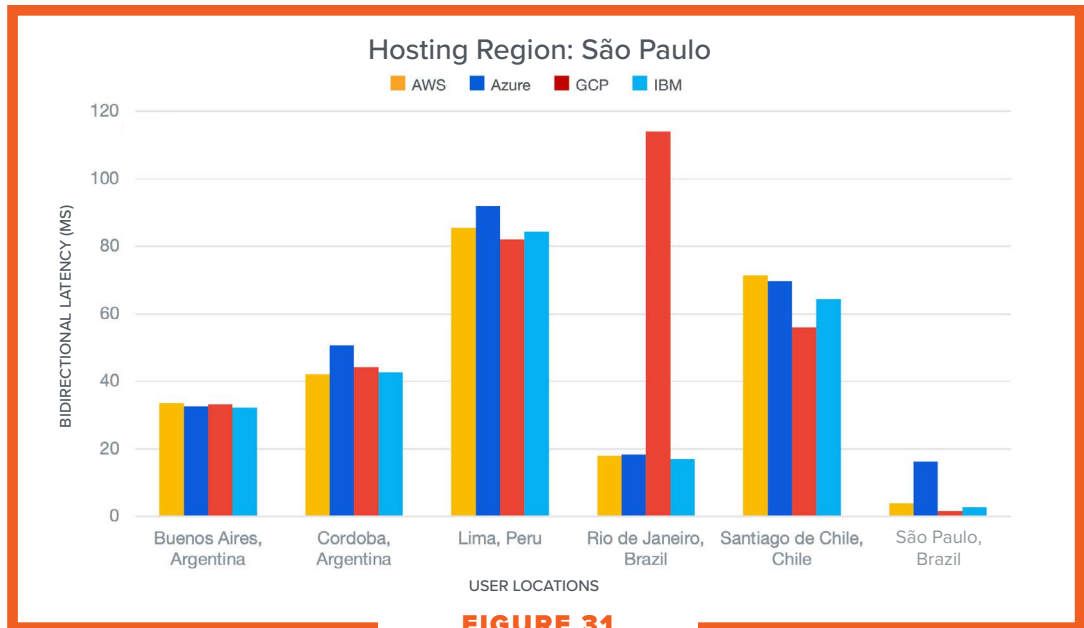


FIGURE 31

Network latency between multiple user locations in South America and the data centers (regions) of the public cloud providers with a presence in that region

In order to understand potential causes for this higher latency, we looked at Path Visualization that showed the network connectivity between Rio and GCP São Paulo. Analysis revealed a sub-optimal reverse path from GCP São Paulo back to our Rio vantage point that included a trans-oceanic route which appeared to add nearly 100ms of latency. We notice that the forward path from our vantage point in Rio exchanges traffic to GCP's network at an IXP in Rio. However, on the reverse path, traffic from GCP São Paulo is carried by Telefonica through Miami, and exchanged at the same IXP in Rio to reach our vantage point.

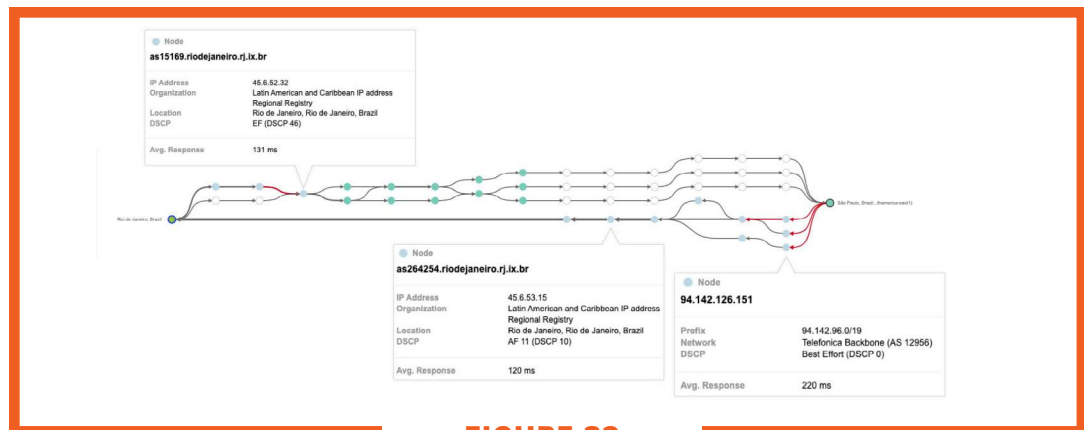


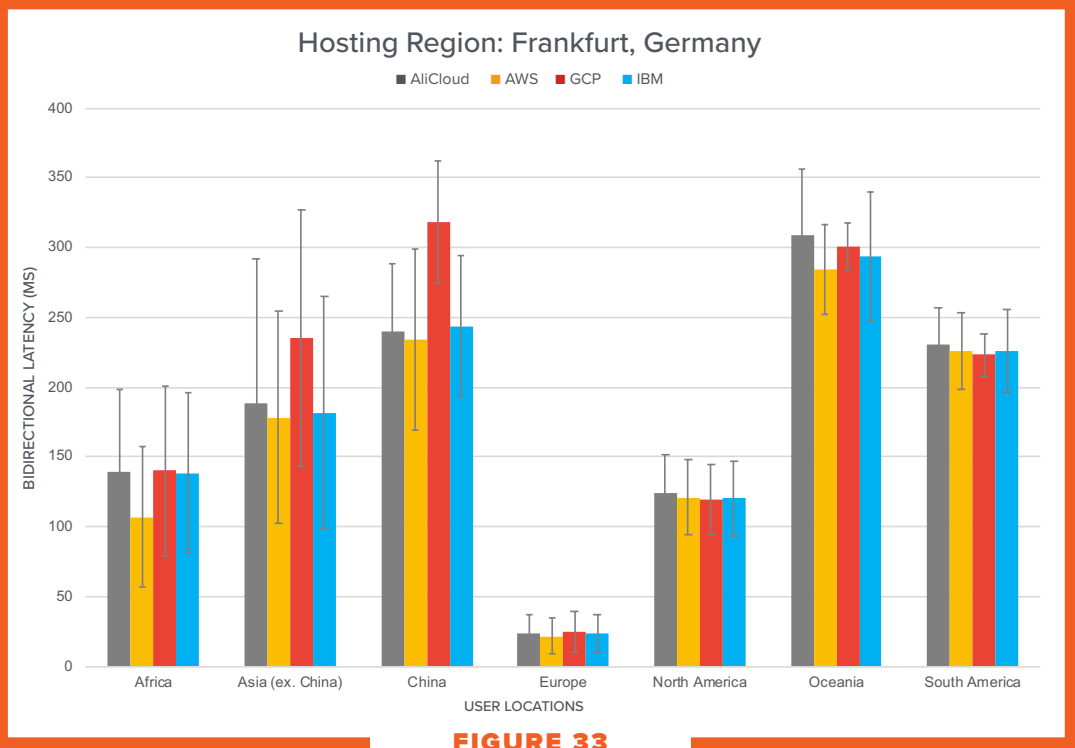
FIGURE 32

Sub-optimal reverse path from GCP São Paulo to Rio

As of April 20th, 2020 the sub-optimal reverse path seems to have been resolved, as manifested in the bidirectional latencies (please see the Appendix for more information). However, enterprises need to remain cognizant of the fact that even for the largest and most competent providers, routing anomalies can occur and create performance-impacting ripple effects.

REGIONAL EXCEPTION 4: FRANKFURT

Our vantage points in Asia were also impacted when connecting to services hosted in GCP's Frankfurt region. Our measurements from Asia and China showed relatively high network latency, ~20% more with GCP (Figure 33), when compared to other providers.



Bi-directional network latency between global user locations and the Frankfurt data centers (regions) of the public cloud providers with a presence in that region.

NETWORK PERFORMANCE IN CHINA

FINDING

The Great Firewall imposes a performance toll on all cloud provider traffic entering and exiting China.

Enterprises hesitant to choose a China-based hosting region with a cloud provider have other viable options that offer reasonable latency.

EVIDENCE

Traffic to and from China, irrespective of which cloud hosting region it is destined to, or originating from, is subject to high packet loss. On the contrary, traffic that is contained within China does not experience packet loss.

Data-driven decisions enable enterprises to pick the optimal cloud provider and hosting regions to serve customers in China. Singapore and Hong Kong are two viable hosting regions with optimal network latency from China.

- Alibaba Cloud has the best network latency between China and Singapore. Alibaba Cloud outperforms Azure and AWS by 32% and 41% respectively.
- Alibaba Cloud has the best network latency between China and Hong Kong, outperforming both Azure and IBM.

Enterprises expanding their global presence in the AsiaPac market are challenged with varying and unpredictable performance. Within Asia, China definitely holds a special position when it comes to network performance and Internet behavior. Heavy and opaque sovereign controls over Internet behavior have long contributed to characteristically unstable Internet performance throughout China. Sitting in between Chinese citizens and the global Internet is the Great Firewall of China, a sophisticated content filtering machine. Employing a multitude of censorship tools—such as IP blocking, DNS tampering and hijacking, deep packet inspection, and keyword filtering—the Great Firewall is designed to ensure that online content aligns with the government party line.

Privacy and ethics concerns aside, one of the drawbacks to this system is a vast reduction in performance. Our testing confirmed that any traffic that passes the Great Firewall is subject to heavy packet loss, a characteristic that was not common across any other political or geographical boundaries. For instance, Figure 34 below, represents the packet loss experienced by users in global continents while connecting to two different hosting centers of all five cloud providers—one in Virginia and the other in India. Irrespective of the cloud provider, packet loss between China and the hosting region is consistently high, showing that no one is exempt from paying the “China Performance Toll.”

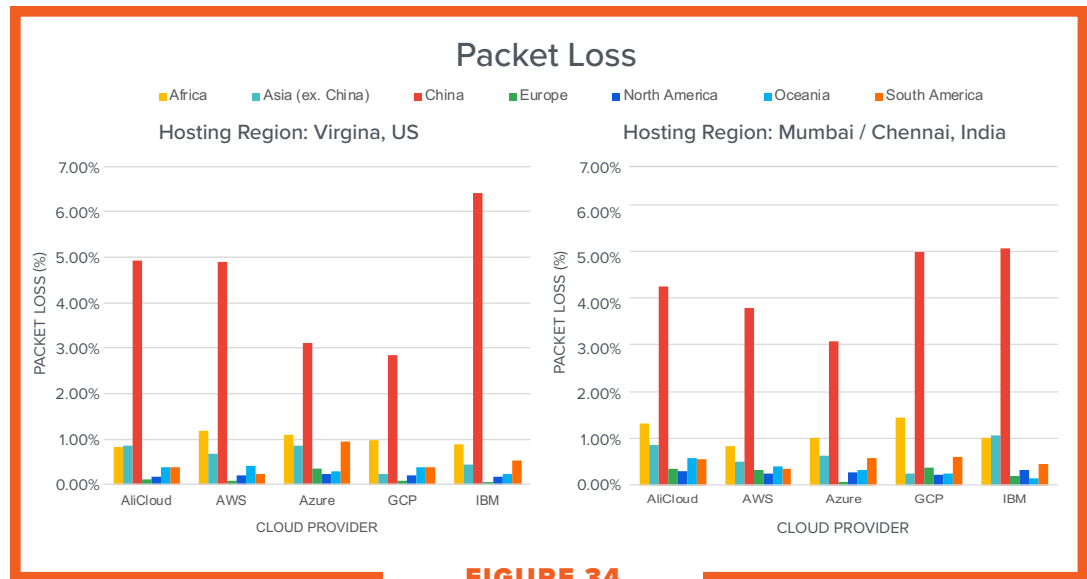


FIGURE 34

Irrespective of the cloud provider, traffic entering and exiting China pays a performance toll through high packet loss

As this behavior is an artifact of the Great Firewall, any traffic that does not cross the firewall is subject to very minimal loss. We observed this pattern, as seen in Figure 35 below, while testing from vantage points within China to Alibaba Cloud regions in China. Due to legal restrictions in using other cloud providers' regions in China, we limited these tests to only Alibaba Cloud regions in China.

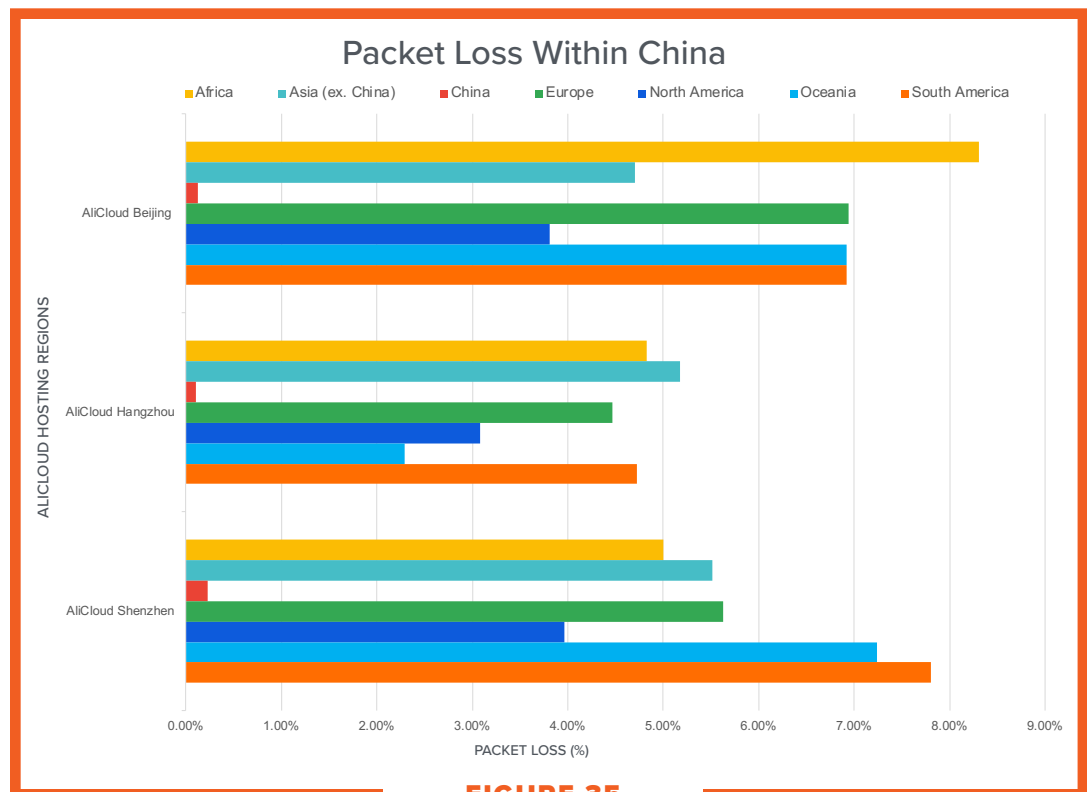


FIGURE 35

Traffic within China is not subject to the Great Firewall penalties

The data above shows that there is obvious benefits to picking an Alibaba Cloud region in China to host workloads and services. However, enterprises that are still cautious and hesitant to pick a cloud provider with origins in China or a hosting region in China do have options. We looked at bidirectional network latency and bidirectional packet loss for a few regions closest to China (Singapore, Hong Kong and India) and compared the cloud providers with presence in these regions (Figures 36 and 37).

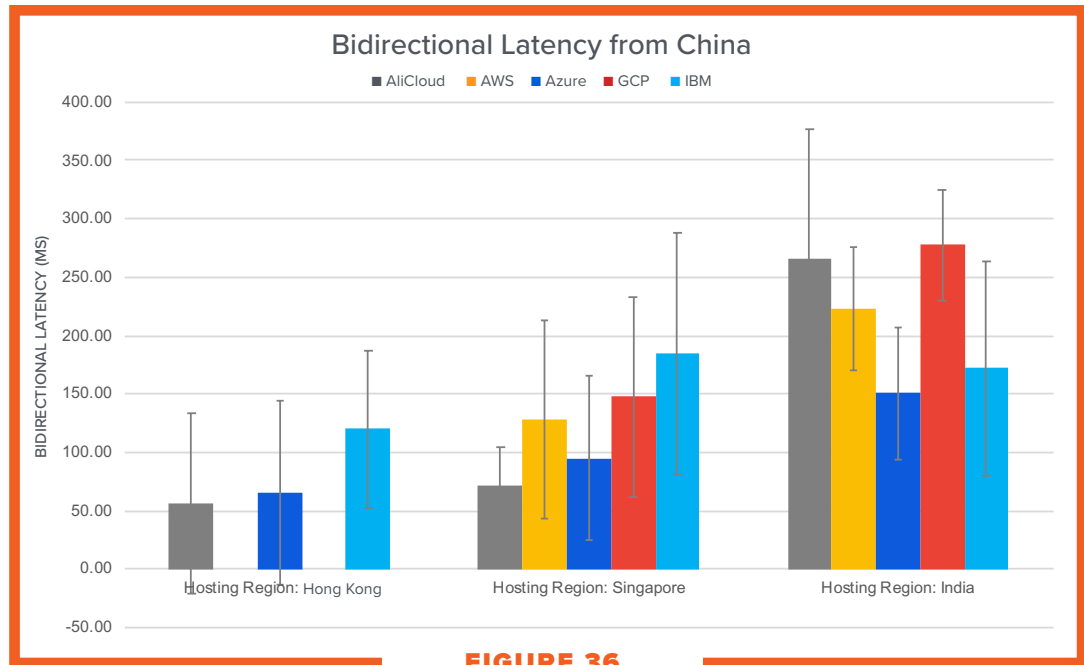


FIGURE 36

Bidirectional latency from China to hosting regions in Asia

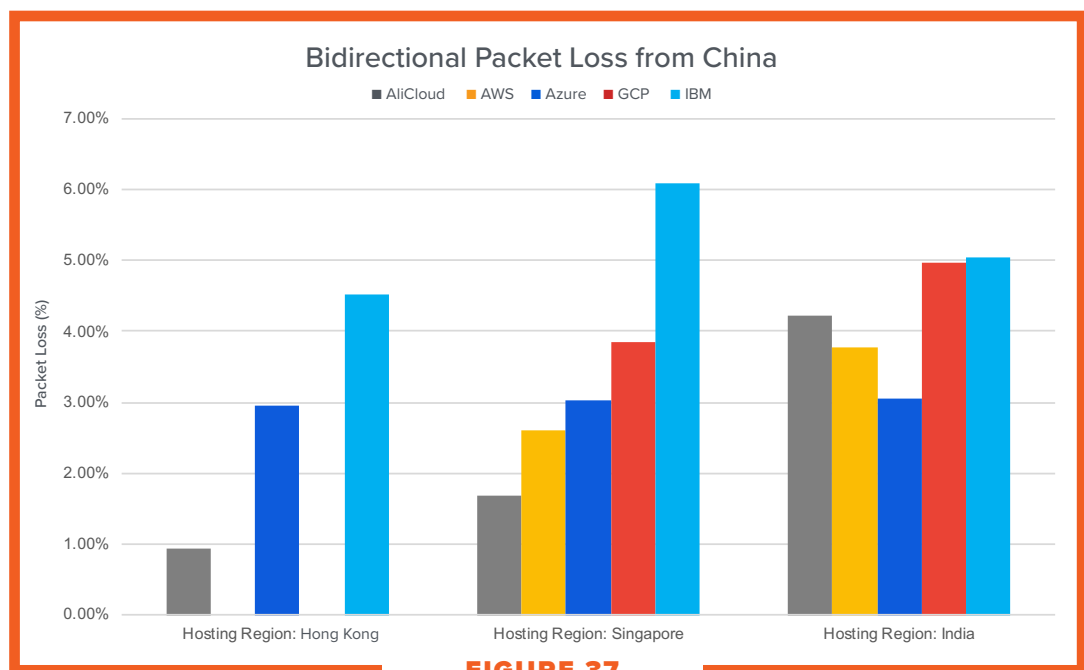


FIGURE 37

Bidirectional packet loss from China to hosting regions in Asia

Alibaba Cloud performs the best in both Singapore and Hong Kong for users connecting from China, from the perspective of network latency, predictability (black vertical lines) and packet loss, but not for India. If, for some reason, Alibaba Cloud is not your first choice, Azure performs equally well across all three hosting regions, not compromising on speed, predictability or packet loss.

TAKEAWAY

All cloud providers, including Alibaba, pay a performance toll when crossing the Great Firewall of China.

RECOMMENDATION

Use data from the report to evaluate hosting options to serve users in China.

BROADBAND ISP PERFORMANCE

FINDING

North American broadband provider connectivity to the cloud providers is generally robust, but performance anomalies exist even in this mature connectivity market.

EVIDENCE

Traffic from Verizon-connected sites located in Seattle, San Jose and Los Angeles that are accessing GCP's us-west2 region is routed to enter the Google backbone in New Jersey before being routed back to the hosting region, which is located in Los Angeles. This sub-optimal routing results in up to 10x the expected network latency.

The research compared bidirectional network latency and the predictability of latency, measured through standard deviation, from six broadband ISPs to cloud provider regions in North America. Overall, all providers did well, but we were able to characterize the broadband ISPs that recorded the lowest network latency (Table 3) and best latency predictability (Table 4) based on the testing city and cloud provider location. Measurements from these six cities showed that CenturyLink delivered the lowest latency more often than other providers, while Comcast delivered better performance predictability more often than other providers. These results are representative, but given the sheer number of potential branch office locations and local connectivity complexities, enterprises contemplating broadband-based hybrid or SD-WAN connectivity to the cloud are advised to take ISP audit measurements from their own locations and peerings.

BEST PERFORMERS: LATENCY																	
	ALIBABA CLOUD		AWS				AZURE				GCP				IBM		
CITY	US West	US East	US West	US Central	US East	Canada	US West	US Central	US East	Canada	US West	US Central	US East	Canada	US West	US East	Canada
ASHBURN, VA	verizon	COX	CenturyLink	Charter	Charter	CenturyLink	verizon	verizon	COX	COX	COMCAST	verizon	COX	COX	verizon	COX	COX
CHICAGO, IL	COMCAST	COMCAST	CenturyLink	CenturyLink	COMCAST	CenturyLink	CenturyLink	COX	COMCAST	AT&T	CenturyLink	COX	COMCAST	COMCAST	CenturyLink	COMCAST	CenturyLink
DALLAS, TX	CenturyLink	Charter	AT&T	COX	Charter	CenturyLink	AT&T	COX	COX	AT&T	AT&T	COX	COX	COX	COX	COX	CenturyLink
LOS ANGELES, CA	COX	verizon	CenturyLink	CenturyLink	Charter	CenturyLink	Charter	CenturyLink	Charter	AT&T	COX	Charter	CenturyLink	verizon	Charter	AT&T	verizon
SAN JOSE, CA	COX	COMCAST	CenturyLink	CenturyLink	COMCAST	CenturyLink	COX	CenturyLink	COMCAST	AT&T	COX	CenturyLink	COMCAST	COMCAST	COX	COMCAST	CenturyLink
SEATTLE, WA	COMCAST	CenturyLink	CenturyLink	CenturyLink	AT&T	AT&T	COMCAST	CenturyLink	verizon	AT&T	CenturyLink	CenturyLink	verizon	verizon	COMCAST	verizon	verizon

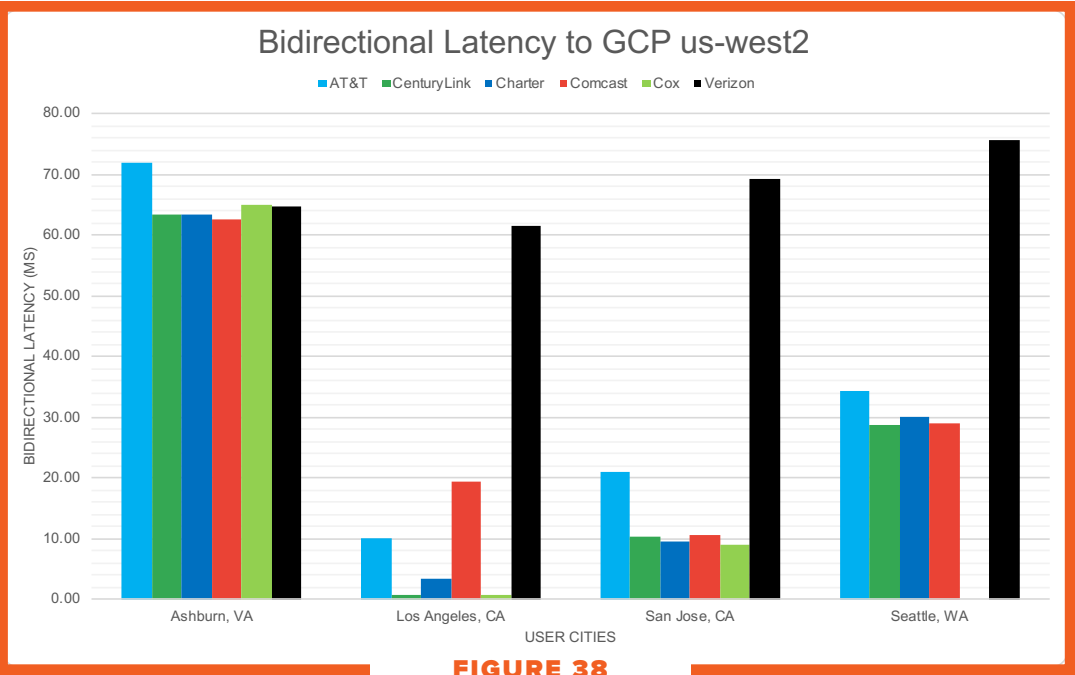
Table 3: Matrix of the Broadband ISP providers with the lowest network latency per city and cloud provider

BEST PERFORMERS: PREDICTABILITY																	
	ALIBABA CLOUD		AWS				AZURE				GCP				IBM		
CITY	US West	US East	US West	US Central	US East	Canada	US West	US Central	US East	Canada	US West	US Central	US East	Canada	US West	US East	Canada
ASHBURN, VA	COMCAST	Charter	COMCAST	Charter	CenturyLink	CenturyLink	COMCAST	CenturyLink	COX	COX	CenturyLink	verizon	COX	COX	COMCAST	CenturyLink	COX
CHICAGO, IL	CenturyLink	COMCAST	CenturyLink	CenturyLink	COMCAST	COMCAST	CenturyLink	CenturyLink	COMCAST	CenturyLink	CenturyLink	COMCAST	COMCAST	COMCAST	CenturyLink	COX	COMCAST
DALLAS, TX	COMCAST	CenturyLink	CenturyLink	CenturyLink	CenturyLink	COMCAST	CenturyLink	CenturyLink	COX	CenturyLink	CenturyLink	COX	CenturyLink	COMCAST	COMCAST	CenturyLink	COMCAST
LOS ANGELES, CA	COMCAST	COX	COX	Charter	Charter	Charter	Charter	COX	Charter	COMCAST	COX	Charter	COX	COX	Charter	Charter	COX
SAN JOSE, CA	Charter	COX	COMCAST	CenturyLink	COMCAST	COMCAST	Charter	CenturyLink	COMCAST	COMCAST	COX	Charter	COMCAST	COMCAST	Charter	COMCAST	Charter
SEATTLE, WA	COMCAST	CenturyLink	COMCAST	CenturyLink	COMCAST	COMCAST	Charter	CenturyLink	COMCAST	COMCAST	COMCAST	CenturyLink	COMCAST	COMCAST	COMCAST	COMCAST	COMCAST

Table 4: Matrix of the Broadband ISP providers with the best performance predictability (measured through standard deviation) per city and cloud provider



Our tests revealed an interesting peering relationship between Verizon-connected vantage point agents in the U.S. west coast accessing GCP's hosting region in Los Angeles, CA (us-west2). As seen from the graph in Figure 38, Verizon agents consistently experience high latencies, in the range of 60ms, from cities in the western United States. Notice that it takes the same time (~60ms) to go back and forth between Ashburn, VA and GCP us-west2.



Verizon-based broadband agents have extraordinarily high latencies when connecting to GCP's region in Los Angeles (us-west2)

A deeper dive into the network connectivity path held the answer to this anomaly. As you can see in Figure 39 below, during our collection period, Verizon agents in Seattle, Los Angeles and San Jose were handing off traffic into GCP's network in New Jersey in the eastern United States, with the traffic then turning around and coming back to GCP's region in Los Angeles (us-west2).

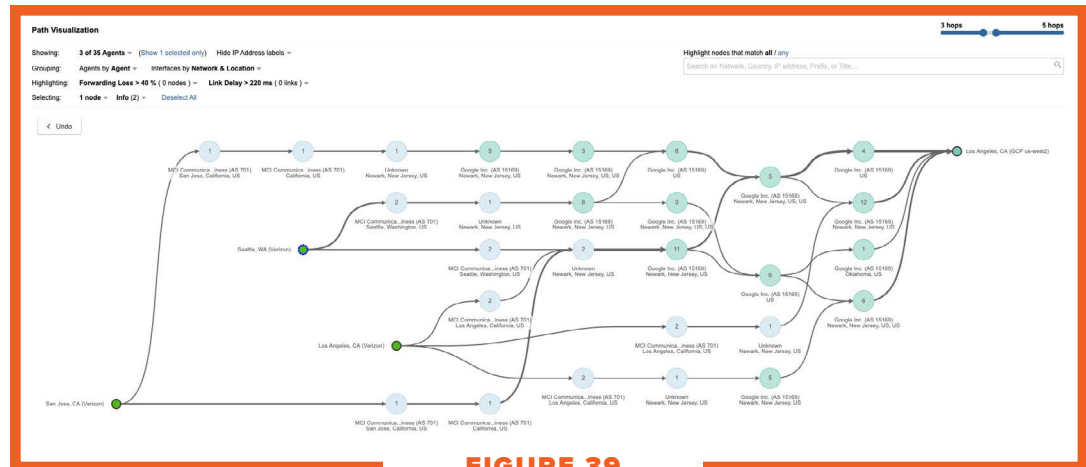


FIGURE 39

Sub-optimal peering between Verizon agents in the west to GCP's hosting region in Los Angeles (us-west2)

NOTE

As of November 8th, 2019 the suboptimal routing seen below has been resolved. Please look at the Appendix for more information.

Note that we did not see this trend for user locations hosted in other broadband ISP providers that we tested. For enterprises considering a hybrid WAN approach or choosing upstream ISPs for branch offices, understanding how your ISP networks with cloud providers is critical for sound pre-deployment baselining and performance audits, and for ongoing operational awareness and optimization.

TAKEAWAY

US broadband ISPs generally perform well in accessing the cloud, but routing anomalies can impact operational performance.

RECOMMENDATION

Ensure sound Internet visibility and performance measurements as part of hybrid WAN planning and deployment so that you can detect anomalies and work collaboratively with your ISP or cloud provider to resolve them.

INTER-REGION PERFORMANCE

FINDING

Alibaba Cloud exhibits distinct behavior with inter-region communication paths when compared to the other four public cloud providers.

Inter-region network latency is aligned with geographical expectations for the most part, with some regional exceptions.

EVIDENCE

Inter-region traffic between compute services stays within the cloud providers' internal network for AWS, Azure, GCP and IBM. However, inter-region traffic between Alibaba Cloud regions often traverses the public Internet.

GCP performed below Internet baselines between their Mumbai (asia-south1) and European regions, due to the lack of strong infrastructure between these geographical hubs.

Inter-region performance is critical for enterprises adopting a tiered multi-region architecture. A common practice across cloud architectures is to distribute compute or workloads in global regions but centralize common functions or services such as storage or databases in a single region. Tiered architectures with geographical expanse can incrementally affect network latencies, impacting end-user experience. In this report, we kept the number of inter-region combinations to 15 regions for AWS, Azure and GCP, 14 regions for IBM Cloud and 13 regions for Alibaba Cloud.

INTER-REGION CONNECTIVITY

Path analysis of inter-region connectivity shows that inter-region communication between compute services stays within the cloud providers' internal network for AWS, Azure, GCP and IBM. Alibaba Cloud, however, is the only provider where inter-region communication is not contained within its own internal backbone and involves Internet paths between two geographical regions. This behavior is illustrated in Figure 40 and Figure 41 below, which show inter-region paths within AWS and Alibaba Cloud respectively.

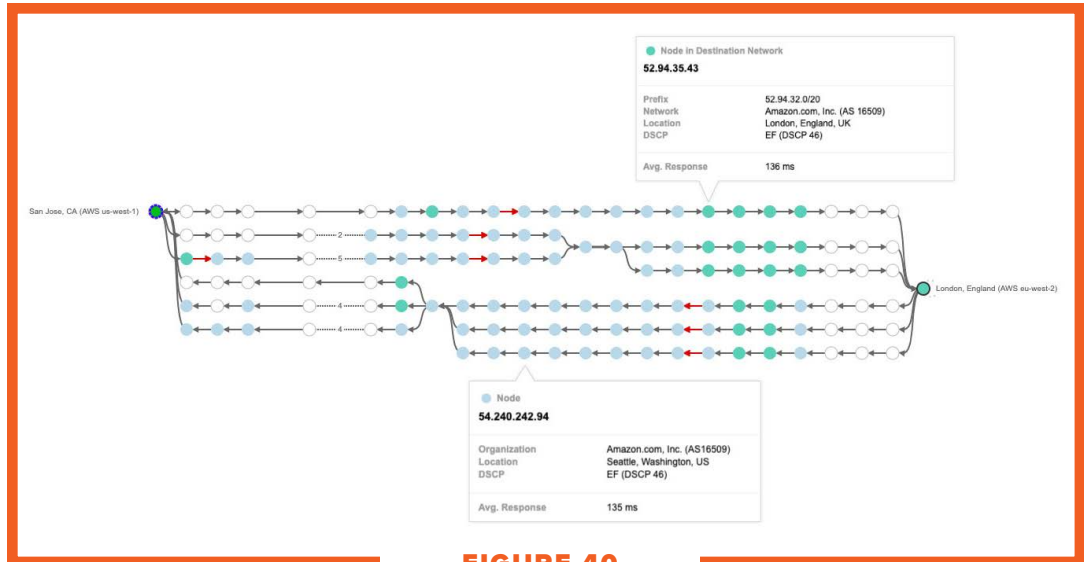


FIGURE 40

Inter-region network path between AWS regions

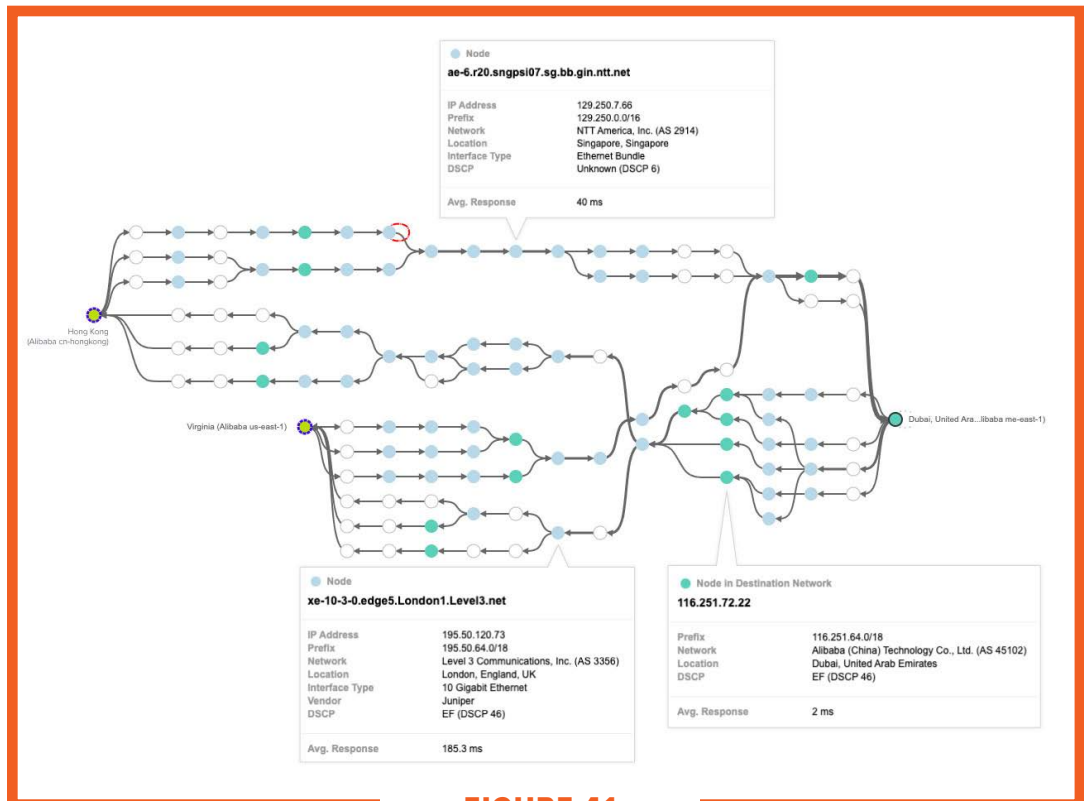


FIGURE 41

Inter-region network paths between Alibaba Cloud regions often cross the public Internet

INTER-REGION PERFORMANCE

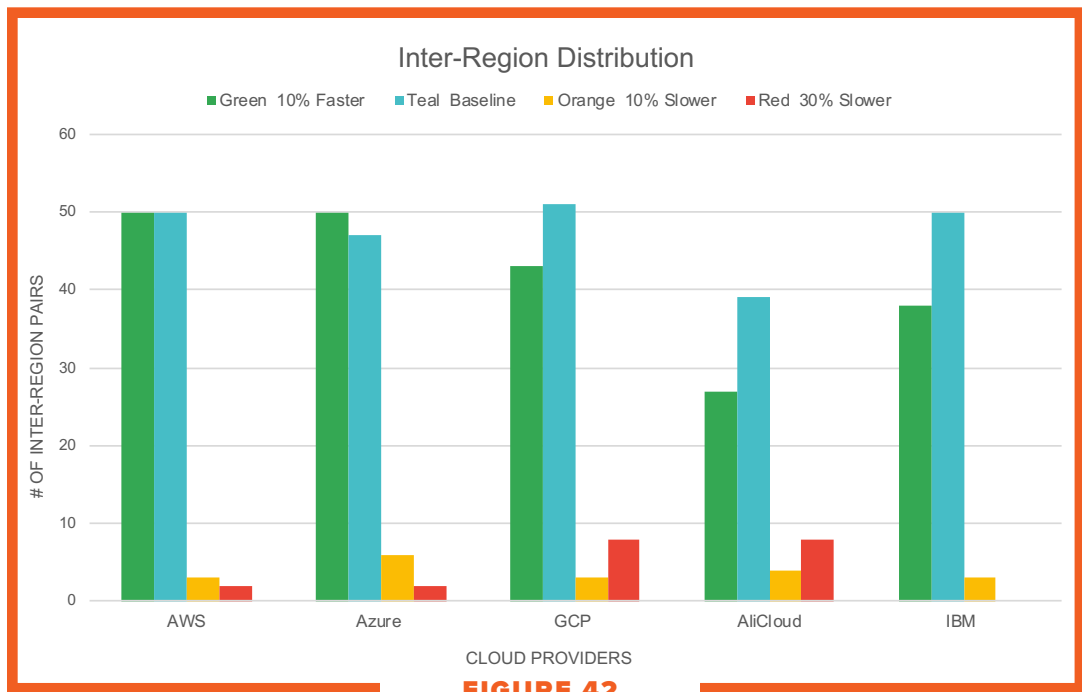
Inter-region latencies for each public cloud provider are compared against baseline measurements between geographically proximate locations over a typical Internet path. With a well-connected and robust internal backbone, we expect latency measurements within a provider to be quantitatively better than the baseline measurements. Use the following key to read through Tables 5-9 and Table 13 in the Appendix.

Inter-region performance across all five providers is better than, or the same as, Internet baselines for more than 85% of their regions (Figure 42). Providers that use their own backbone (AWS, Azure, GCP and IBM) have the least number of region pairs with below baseline performance. However, one exception exists with Google Cloud.

As seen in Table 8 Traffic between GCP's asia-south1 (Mumbai) location and regions in Europe (Frankfurt, London, Belgium and Netherlands) exhibit 30% slow network latency relative to the Internet path. This correlates with similar patterns for user traffic from Europe as noted in GCP's Cloud Backbone Continues to Impact Europe & India section of the report.

UNDERSTANDING INTER-REGION MEASUREMENTS

- 10% faster than baseline
- Same as baseline
- 10% slower than baseline
- 30% slower than baseline



Distribution of above baseline and below baseline regions across all five cloud providers in 2019

As seen above, Alibaba Cloud has the largest percentage ~15.38% of regions performing below baseline measurements. A contributing factor for this behavior could be its reliance on the public Internet for cross-region communication.

Note: The above comparison is from data gathered in the collection period defined in the Methodology section. As of March 2020, we re-tested Inter-Region measurements for Microsoft Azure alone and saw improved performance. Please refer to the Appendix for an updated Inter-Region performance table (Table 13) for Microsoft Azure.

TABLE 5. INTER-REGION NETWORK LATENCY MEASUREMENTS IN ALIBABA CLOUD (2019)

AVERAGE BI-DIRECTIONAL LATENCY (MS)	VIRGINIA	SILICON VALLEY	FRANKFURT, GERMANY	LONDON, ENGLAND	DUBAI, UNITED ARAB EMIRATES	TOKYO, JAPAN	SINGAPORE	KUALA LUMPUR, MALAYSIA	MUMBAI, INDIA	SYDNEY, AUSTRALIA	JAKARTA, INDONESIA	BEIJING, CHINA	SHANGHAI, CHINA
VIRGINIA		72.02	94.41	75.94	202.02	169.93	245.48	244.74	232.41	251.30	245.64	247.96	225.35
SILICON VALLEY			154.77	144.73	258.13	106.95	181.10	193.31	250.10	148.90	188.14	183.15	174.79
FRANKFURT, GERMANY				20.16	179.59	260.58	251.99	192.57	117.85	304.03	178.78	170.14	196.15
LONDON, ENGLAND					134.52	223.72	250.15	173.14	121.68	320.41	178.88	183.08	279.15
DUBAI, UNITED ARAB EMIRATES						158.86	256.50	96.31	36.45	290.32	101.04	259.42	192.25
TOKYO, JAPAN							78.08	85.18	130.71	148.55	89.77	63.01	49.18
SINGAPORE								61.69	57.97	198.76	31.95	81.46	74.21
KUALA LUMPUR, MALAYSIA									89.87	274.71	50.65	201.46	182.72
MUMBAI, INDIA										151.92	72.60	269.28	266.70
SYDNEY, AUSTRALIA											122.19	311.61	265.28
JAKARTA, INDONESIA												189.20	175.60
BEIJING, CHINA													24.65
SHANGHAI, CHINA													

10% faster than baseline Same as baseline 10% slower than baseline 30% slower than baseline

TABLE 6. INTER-REGION NETWORK LATENCY MEASUREMENTS IN AWS (2019)



AVERAGE BI-DIRECTIONAL LATENCY (MS)	CA-CENTRAL-1 MONTREAL, CANADA	SA-EAST-1 SÃO PAULO, BRAZIL	US-EAST-1 ASHBURN, VA	US-EAST-2 COLUMBUS, OH	US-WEST-1 SAN JOSE, CA	US-WEST-2 THE DALLES, OR	EU-CENTRAL-1 FRANKFURT, GERMANY	EU-WEST-1 DUBLIN, UK	EU-WEST-2 LONDON, UK	EU-WEST-3 PARIS, FRANCE	AP-NORTHEAST-1 TOKYO, JAPAN	AP-NORTHEAST-2 SEOUL, SOUTH KOREA	AP-SOUTH-1 MUMBAI, INDIA	AP-SOUTHEAST-1 SINGAPORE	AP-SOUTHEAST-2 SYDNEY, AUSTRALIA
CA-CENTRAL-1		124.30	14.85	24.78	80.62	69.27	100.82	75.96	87.16	93.75	157.31	184.70	196.04	224.01	202.04
SA-EAST-1			123.35	131.61	196.06	186.49	206.70	182.93	193.73	199.25	272.30	299.21	301.85	338.80	314.20
US-EAST-1				11.42	61.02	74.34	85.80	70.51	75.43	78.65	153.73	182.63	181.13	227.68	197.70
US-EAST-2					51.45	69.75	96.53	93.75	85.94	89.32	155.73	184.22	191.70	222.90	194.63
US-WEST-1						21.21	145.41	142.15	136.58	138.20	112.20	135.86	233.26	174.29	149.83
US-WEST-2							161.24	131.67	138.23	154.06	100.76	124.85	220.10	162.35	138.57
EU-CENTRAL-1								23.30	14.15	8.71	243.15	270.38	117.39	176.10	283.70
EU-WEST-1									10.80	16.38	209.15	238.62	119.25	177.27	259.47
EU-WEST-2										7.48	214.97	244.91	110.36	167.84	266.80
EU-WEST-3											240.85	269.30	105.75	162.06	276.44
AP-NORTHEAST-1												32.76	125.76	70.14	106.12
AP-NORTHEAST-2													153.84	95.51	146.63
AP-SOUTH-1														60.45	237.44
AP-SOUTHEAST-1															178.13
AP-SOUTHEAST-2															

■ 10% faster than baseline ■ Same as baseline ■ 10% slower than baseline ■ 30% slower than baseline

TABLE 7. INTER-REGION NETWORK LATENCY MEASUREMENTS IN MICROSOFT AZURE (2019) [Updated version on page 68](#)

AVERAGE BI-DIRECTIONAL LATENCY (MS)	CANADA CENTRALTORONTO, CANADA	BRAZIL SOUTH SÃO PAULO, BRAZIL	EAST US RICHMOND, VA	CENTRAL US DES MOINES, IA	WEST US SANTA CLARA, CA	WEST US 2 QUINCY, WA	NORTH EUROPE DUBLIN, IRELAND	UK WEST CARDIFF, UK	FRANCE CENTRAL PARIS, FRANCE	WEST EUROPE AMSTERDAM, NETHERLANDS	JAPAN EAST TOKYO, JAPAN	KOREA CENTRAL SEOUL, SOUTH KOREA	WEST INDIA MUMBAI, INDIA	SOUTHEAST ASIA SINGAPORE	AUSTRALIA EAST SYDNEY, AUSTRALIA
CANADA CENTRALTORONTO		141.09	38.44	24.26	72.89	66.24	88.97	99.07	102.67	106.12	163.13	187.23	214.67	224.98	225.35
BRAZIL SOUTH SÃO PAULO, BRAZIL			121.78	138.38	170.69	180.89	172.39	183.28	186.99	189.46	260.58	301.44	299.93	336.21	310.69
EAST US – RICHMOND, VA				23.38	59.17	69.29	70.74	80.73	80.80	87.45	163.76	176.33	192.49	225.58	199.74
CENTRAL US – DES MOINES, IA					53.90	47.14	85.77	97.25	100.21	102.50	140.03	164.96	212.93	204.40	203.24
WEST US – SANTA CLARA, CA						22.99	128.56	146.35	139.49	145.92	105.45	129.53	229.20	167.62	159.74
WEST US 2 – QUINCY, WA							131.46	139.64	147.06	147.45	96.84	121.57	221.54	159.79	178.37
NORTH EUROPE – DUBLIN, IRELAND								12.85	16.43	17.34	225.16	249.53	128.66	194.32	268.90
UK WEST – CARDIFF, UK									10.82	13.38	235.55	250.99	123.30	188.60	279.77
FRANCE CENTRAL – PARIS, FRANCE										11.71	239.32	243.76	114.37	180.26	279.92
WEST EUROPE – AMSTERDAM, NETHERLANDS											242.20	251.66	122.63	187.87	285.98
JAPAN EAST – TOKYO, JAPAN												30.13	128.76	67.54	107.31
KOREA CENTRAL – SEOUL, SOUTH KOREA													124.71	63.51	135.32
WEST INDIA – MUMBAI, INDIA														63.93	160.31
SOUTHEAST ASIA – SINGAPORE															98.47
AUSTRALIA EAST – SYDNEY, AUSTRALIA															

■ 10% faster than baseline ■ Same as baseline ■ 10% slower than baseline ■ 30% slower than baseline

TABLE 8. INTER-REGION NETWORK LATENCY MEASUREMENTS IN GCP (2019)

AVERAGE BI-DIRECTIONAL LATENCY (MS)	NORTHAMERICA-NORTHEAST1 MONTREAL, CANADA	SOUTHAMERICA-EAST1 SAO PAULO, BRAZIL	US-EAST4 ASHBURN, VA	US-CENTRAL1 COUNCIL BLUFFS, IA	US-EAST1 MONCKS CORNER, SC	US-WEST1THE DALLES, OR	EUROPE-WEST3 FRANFURT, GERMANY	EUROPE-WEST2 LONDON, UK	EUROPE-WEST1 ST. GHISLAIN, BELGIUM	EUROPE-WEST4 EEMSHAVEN, NETHERLANDS	ASIA-NORTHEAST1 TOKYO, JAPAN	ASIA-EAST1 CHANGHUA, TAIWAN	ASIA-SOUTH1 MUMBAI, INDIA	ASIA-SOUTHEAST1 SINGAPORE	AUSTRALIA-SOUTHEAST1 SYDNEY, AUSTRALIA
NORTHAMERICA-NORTHEAST1		143.03	14.69	31.88	25.84	65.41	88.65	77.23	82.72	86.97	154.00	183.46	276.19	216.87	204.20
SOUTHAMERICA-EAST1 SAO PAULO, BRAZIL			129.82	140.26	117.82	172.22	215.39	204.07	209.83	214.33	261.47	290.42	383.05	323.70	304.38
US-EAST4 – ASHBURN, VA				25.44	12.55	58.86	87.01	75.66	81.24	85.84	147.75	176.94	269.65	210.41	197.88
US-CENTRAL1 – COUNCIL BLUFFS, IA					33.63	34.94	106.20	94.82	100.31	104.53	123.76	152.93	245.44	186.28	173.82
US-EAST1 – MONCKS CORNER, SC						66.77	98.21	86.85	92.34	96.96	156.09	184.91	277.62	218.31	198.77
US-WEST1 – THE DALLES, OR							139.58	128.20	134.15	138.20	90.10	118.93	211.60	152.16	162.53
EUROPE-WEST3 – FRANFURT, GERMANY								13.44	7.68	7.75	228.19	257.48	350.53	291.02	278.57
EUROPE-WEST2 – LONDON, UK									7.11	11.80	216.62	246.09	338.81	279.42	267.11
EUROPE-WEST1 – ST. GHISLAIN, BELGIUM										7.60	222.63	251.89	344.33	285.04	272.62
EUROPE-WEST4 – EEMSHAVEN, NETHERLANDS											226.41	255.90	348.65	289.37	276.97
ASIA-NORTHEAST1 – TOKYO, JAPAN												34.28	127.26	67.84	115.33
ASIA-EAST1 – CHANGHUA, TAIWAN													106.56	47.13	138.42
ASIA-SOUTH1 – MUMBAI, INDIA														60.00	151.49
ASIA-SOUTHEAST1 – SINGAPORE															91.93
AUSTRALIA-SOUTHEAST1 – SYDNEY, AUSTRALIA															

■ 10% faster than baseline ■ Same as baseline ■ 10% slower than baseline ■ 30% slower than baseline

TABLE 9. INTER-REGION NETWORK LATENCY MEASUREMENTS IN IBM CLOUD (2019)

AVERAGE BI-DIRECTIONAL LATENCY (MS)	MONTREAL, CANADA	SÃO PAULO, BRAZIL	ASHBURN, VA, US	DALLAS, TEXAS, US	SAN JOSE, CALIFORNIA, US	FRANKFURT AM MAIN, GERMANY	LONDON, ENGLAND, UK	PARIS, FRANCE	AMSTERDAM, NETHERLANDS	TOKYO, JAPAN	SEOUL, SOUTH KOREA	CHENNAI, TAMIL NADU, INDIA	CENTRAL SINGAPORE, SINGAPORE	SYDNEY, AUSTRALIA
MONTREAL, CANADA		120.03	15.99	41.73	68.78	87.92	77.94	85.63	86.77	152.35	196.29	218.17	250.84	211.80
SÃO PAULO, BRAZIL			117.26	144.28	181.94	189.08	179.20	186.86	188.03	277.84	319.77	319.19	355.57	316.41
ASHBURN, VA, US				31.39	68.47	84.85	75.08	89.63	80.25	152.02	195.19	221.89	244.15	205.23
DALLAS, TEXAS, US					39.94	121.66	111.76	119.49	109.57	136.14	177.74	247.09	213.54	174.42
SAN JOSE, CALIFORNIA, US						148.82	138.92	146.56	145.69	99.35	141.60	212.17	178.73	154.63
FRANKFURT AM MAIN, GERMANY							12.12	10.45	6.80	226.59	216.24	143.18	150.36	241.34
LONDON, ENGLAND, UK								8.52	9.90	237.87	227.58	140.88	161.68	252.66
PARIS, FRANCE									11.56	234.12	223.77	133.53	157.88	248.86
AMSTERDAM, NETHERLANDS										244.37	233.96	143.87	168.08	259.12
TOKYO, JAPAN											40.84	111.47	77.92	121.23
SEOUL, SOUTH KOREA												101.14	67.58	150.55
CHENNAI, TAMIL NADU, INDIA													35.19	126.20
CENTRAL SINGAPORE, SINGAPORE														92.73
SYDNEY, AUSTRALIA														

■ 10% faster than baseline ■ Same as baseline ■ 10% slower than baseline ■ 30% slower than baseline

TAKEAWAY

Inter-region network paths vary across the providers, including geographical and Internet vs backbone variations.

RECOMMENDATION

While building a multi-region, cloud architecture, be aware of regional anomalies within your cloud provider. Use the data in this report and your own baselining to inform the choice of inter-region pairs.



INTER-AZ PERFORMANCE

FINDING

Inter-AZ bidirectional network latency is comparable across all five providers and within publicly advertised limits.

A few provider regions exhibit outliers with very low bidirectional inter-AZ latency.

EVIDENCE

Inter-AZ latency averages across multiple regions of all five providers are between 0.5ms-2.5ms. Global averages place GCP in the lead followed by Azure, AWS, Alibaba Cloud and IBM Cloud.

Alibaba Cloud inter-AZ latencies in eu-west-1 averaged at 0.27ms, while GCP's southamerica-east1 region demonstrated a low average of 0.37ms.

Availability Zones (AZ) or Zones are independent, fault-tolerant failure zones within a single physical region of a public cloud provider. AZs hold relevance only within a single geographical region and provide an extra layer of resiliency that is commonly used for redundancy, load-balancing and disaster recovery. All zones within a region are connected to each other through low-latency links that deliver fast network connectivity. Enterprises building a fault tolerant application usually spread resources across multiple zones in a region to protect against unexpected failures. [AWS Availability Zones](#) are physically separated within a typical metropolitan region and have discrete uninterruptible power supply and onsite backup generation facilities. However, [GCP zones](#) don't always correspond to a single physical building, yet GCP maintains that they are single failure domains within a region. Enterprises adopting an inter-AZ architecture should not assume consistency with regards to zone definition across all public cloud providers and gather relevant details to understand how inter-AZ redundancy and fault-tolerance works for the specific providers under consideration.

While all five of the cloud providers evaluated in this report use AZs as the core building block of their data centers, it is a relatively new offering from Microsoft Azure, starting out in North America and Europe. For this reason, we limited the number of Azure regions that were tested for this study. At the time of writing this report, the globally available regions and AZs for each of the five public cloud provider are listed below (Figure 44).

Data gathered through a four-week period, indicates that all providers have comparable and strong inter-AZ performance. The cloud providers exhibit consistent and predictable inter-AZ latencies between 0.5-2.5 ms (Figure 43).

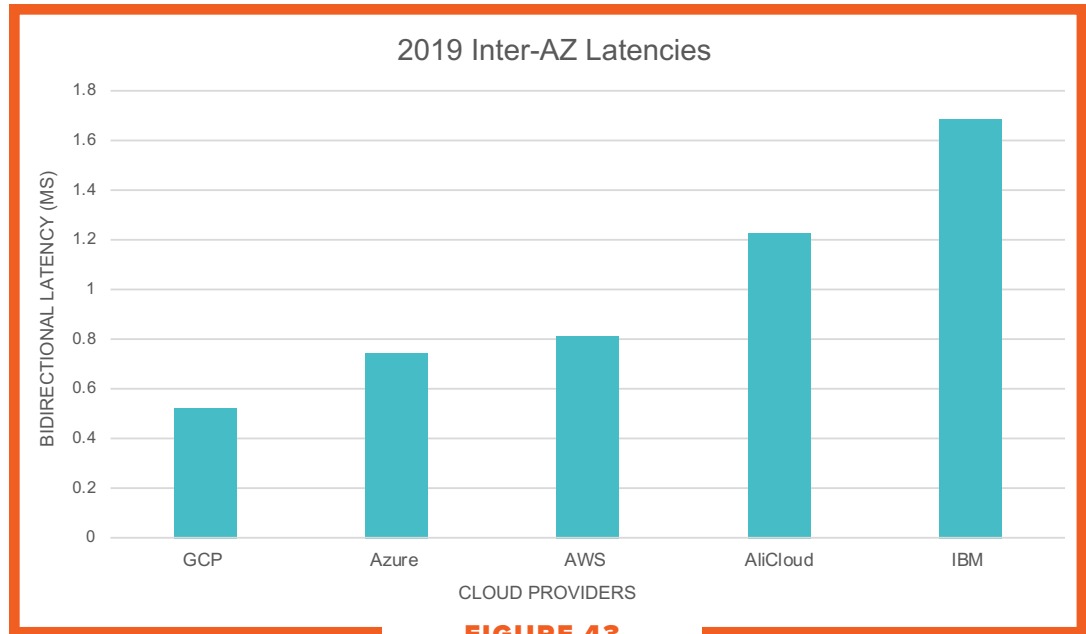


FIGURE 43

2019 global averages of inter-AZ latencies across the top five public cloud providers

Figure 44 below shows granular measurements across all the five cloud providers and their regions where bidirectional inter-AZ latencies were gathered. Low latencies within a region facilitate redundant, dynamically orchestrated overflow architectures that are becoming increasingly common in cloud application deployments.

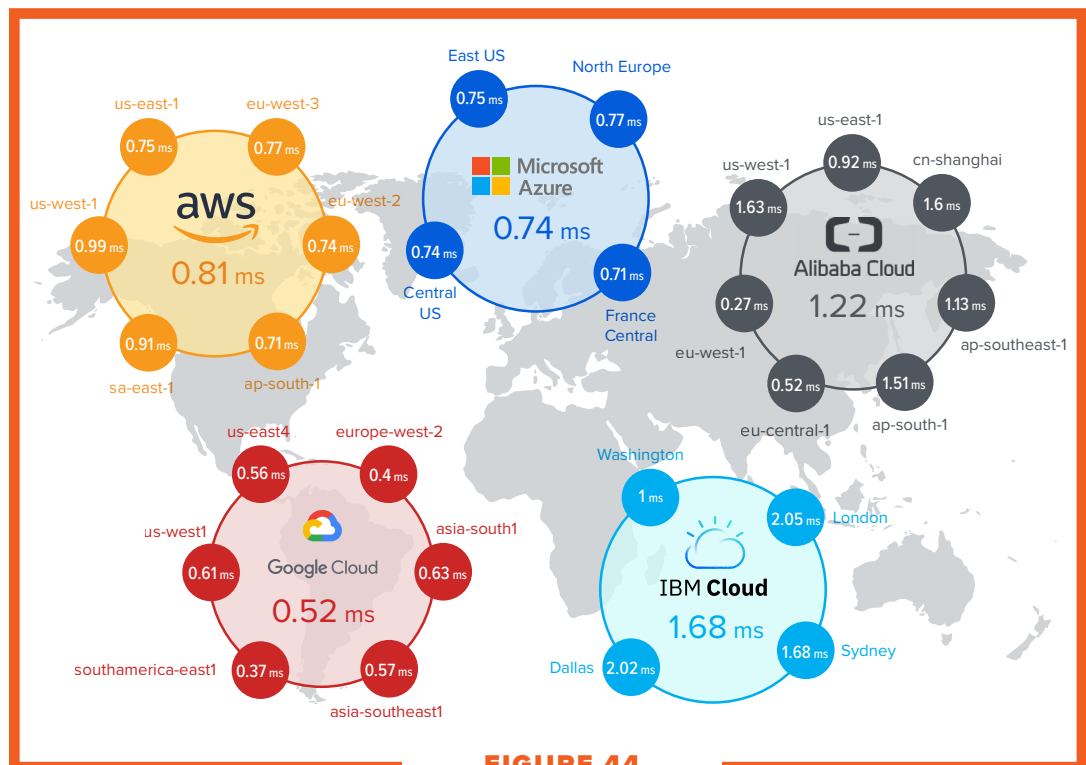


FIGURE 44

Average inter-AZ latency measurements for Azure, AWS, GCP, IBM and Alibaba Cloud over a four-week period

TAKEAWAY

Multi-AZ architectures are ready for production deployments that require robust resiliency.

Very low bidirectional inter-AZ latency measurements in some provider regions indicate relatively small distances between the fault-recovery zones.

RECOMMENDATION

Use inter-AZ and inter-region hosting architectures to build more resilient cloud deployments.

Enterprises adopting an inter-AZ architecture should not assume consistency with regard to zone definitions across all public cloud providers and gather relevant details to understand how inter-AZ redundancy and fault-tolerance works across all providers.



AWS GLOBAL ACCELERATOR

FINDING

AWS Global Accelerator is not a universal solution for performance improvements and consistency for AWS deployments.

EVIDENCE

While the Global Accelerator definitely uses an optimized route through AWS' densely connected backbone, performance improvements were not uniform across the globe. In many cases, the Global Accelerator trumps the Internet connectivity path in performance, but there are also examples of negligible performance improvements and even cases of worse performance when compared to default AWS connectivity.

As discussed earlier in the [Understanding Cloud Connectivity Architectures](#) section, AWS relies heavily on the Internet to move traffic from global user locations to their regions where workloads are hosted. The reason AWS connectivity works this way is because AWS does not anycast public routes associated with each of their regions from global edge locations, resulting in traffic flowing through the Internet longer. The impact of an "Internet-intensive" connectivity architecture can be seen in lower performance predictability as previously outlined in Figure 11.

The AWS Global Accelerator is a commercially available service that enterprises can pay for to leverage the benefits of AWS densely-connected backbone network. Instead of using the Internet to carry user traffic, AWS Global Accelerator directs traffic to optimal endpoints on the AWS edge network by anycasting static IP addresses designated for your service. This results in traffic entering the AWS network closest to the user and making its way to the destination service region through the AWS private backbone, as seen in Figure 45.

Our tests clearly indicated the difference in network connectivity paths seen below. Notice how in Figure 45, traffic from a user location in Seoul enters AWS backbone in Seoul with the Global Accelerator to reach an EC2 workload hosted in Ashburn. Without the Global Accelerator, traffic from Seoul hops through the Internet, through multiple ISPs and only enters AWS' backbone in the United States (Figure 46).

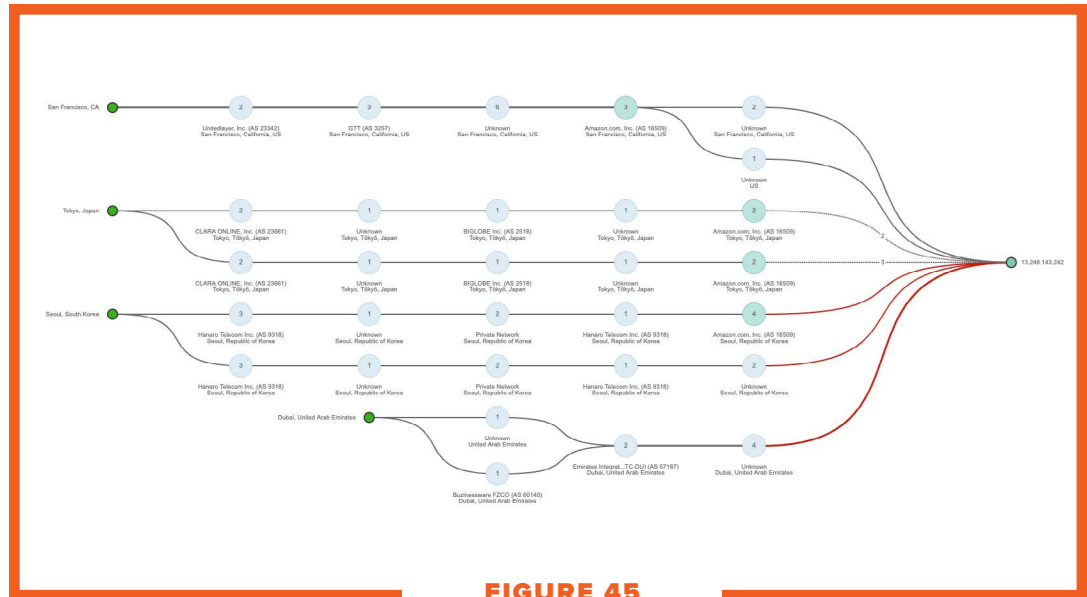


FIGURE 45

With the AWS Global Accelerator, user traffic takes a more optimized path and uses the AWS backbone heavily

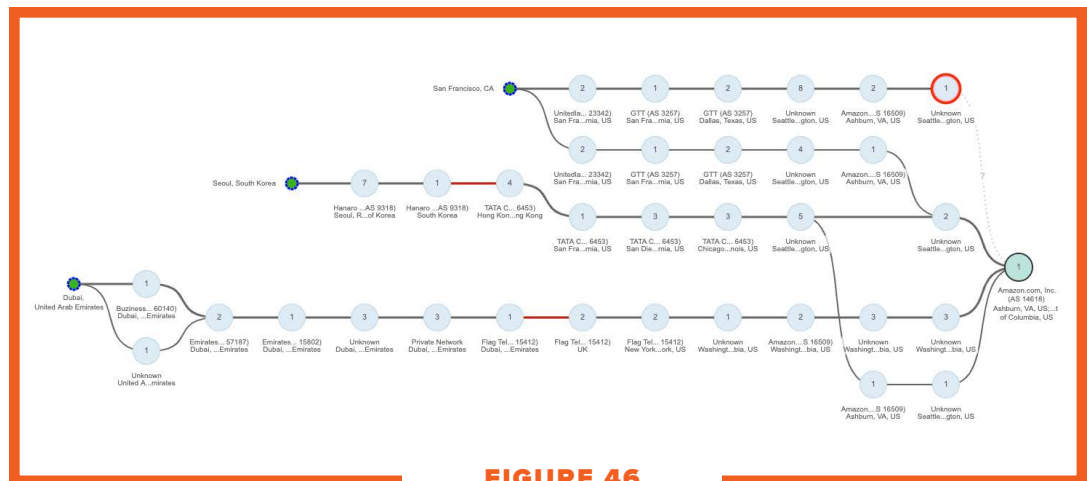


FIGURE 46

By default, user traffic destined to AWS regions heavily relies on the Internet

GLOBAL ACCELERATOR PERFORMANCE COMPARISON

We evaluated the AWS Global Accelerator (AGA) network performance (round-trip network latency, jitter and variation of latency) over a four-week period from 38 global vantage points and compared it to the default Internet-intensive AWS connectivity from the same vantage points. A detailed, comparative view of the data can be found in the Appendix. In this section, we will highlight our key takeaway with a couple of examples. The following table (Table 10) compares round trip latency, jitter and the variation from latency with and without Global Accelerator to AWS us-east-1 (Ashburn) from three global locations—Seoul, San Francisco and Bangalore.

USER VANTAGE POINTS	LATENCY				JITTER			
	DEFAULT INTERNET CONNECTION	GLOBAL ACCELERATOR CONNECTION	LATENCY IMPROVEMENT	% IMPROVEMENT	DEFAULT INTERNET CONNECTION	GLOBAL ACCELERATOR CONNECTION	JITTER IMPROVEMENT	% IMPROVEMENT
SEOUL, SOUTH KOREA	267.87	179.29	88.58	33.07%	14.48	0.25	14.23	98.27%
SAN FRANCISCO, CA	69.88	62.37	7.51	10.75%	0.08	0.07	0.01	12.50%
BANGALORE, INDIA (RELIANCE)	219.74	323.27	-103.53	-47.11%	0.72	1.17	-0.45	-62.50%

Table 10: Comparison of AWS Global Accelerator performance versus default Internet connectivity for the us-east-1 region in Ashburn, VA during the data collection period between 10/5/2019 - 11/01/2019

As evident from Table 10 above, Seoul showed a massive improvement in performance in all three domains. On Global Accelerator, Seoul improved its latency by 88ms while lowering jitter from 14ms to <1ms, almost a 98% improvement. The predictability or consistency of latency also improved by 92%. However, San Francisco did not see that drastic an improvement in performance. Network latency, jitter and variation in latency (standard deviation) is comparable with and without Global Accelerator. Bangalore on the other hand saw a performance drop with Global Accelerator. Latency from the Bangalore vantage point got worse by 100ms and predictability decreased by 38% from 17ms (without AGA) to 85ms (with the AGA).

While there is undeniable improvements to performance from many locations (see the table in Appendix for metrics from all 38 vantage points), it is not to be assumed that this performance uplift will be consistent across global locations and AWS regions. One should also remember that performance can be affected by various factors. For example, the network in which the user or vantage point is located, AWS' peering relationships with global ISPs and whether the closest edge location supports AGA or not. Given these various contributing factors to performance, enterprises should always evaluate the readiness of a new deployment from vantage points that are representative of their customers for accuracy.

Also, as seen earlier in the report, cloud providers are continuously optimizing their networks and connectivity, so one can expect AWS to make improvements to AGA over time. However, remember that there is no steady state in the cloud and that continuous monitoring along with having the evidence to escalate to your cloud provider is critical for successful operations.

Note: As of March 31st, 2020 we noticed multiple instances of improvements to the AWS Global Accelerator performance. Please refer to the Appendix for more information. As per our methodology, we retested over a 4-week period, from March 1st - March 31st, 2020. User vantage points from Mumbai (Reliance), Frankfurt (Cogent) and Berlin (Telia) were not used in the second data collection period, bringing the total vantage points we tested the AWS Global Accelerator to 35.

TAKEAWAY

AWS Global Accelerator doesn't always outperform the Internet.³

RECOMMENDATION

Enterprises considering investing in AGA should conduct a performance audit to evaluate relative performance benefits and determine if adopting AGA will deliver a sufficient ROI. In addition, ongoing performance measurement should be maintained to ensure that baseline improvements are evident over time.

MULTI-CLOUD CONNECTIVITY

FINDING

Multi-cloud connectivity is not consistent across providers and geographical boundaries.

EVIDENCE

The Big 3 (AWS, Azure and GCP) peer directly with each other. However, IBM and Alibaba Cloud don't have fully established, direct peerings with other providers, and rely in many cases on ISPs to connect their clouds to other providers.

Multi-cloud strategies and initiatives are on the rise as enterprises look to reduce vendor lock-in and access best-of-breed services from different cloud providers. Network performance has not been a traditional metric to consider in the formulation of a multi-cloud strategy; however, global performance variations presented above reinforce the case for multi-cloud.

Given the complex matrix of multi-cloud region pairs to test, we only evaluated connectivity patterns from a handful of regions across the providers. For example, we looked at network connectivity from 10 global IBM regions to 4 regions of AWS, Azure and GCP. For Alibaba Cloud, we evaluated connectivity patterns from 6 Alibaba regions to 3 AWS, Azure, GCP and IBM regions.

AWS, Azure and GCP peer directly with each other in a full mesh of connectivity, eliminating the dependence on third-party ISPs for multi-cloud communication. These three cloud providers have vast networks and are well connected across multiple popular colocation facilities.

Based on the combinations we tested, IBM showed a strong peering relationship with GCP and Azure, but had spotty peering with AWS and Alibaba Cloud. With Alibaba Cloud, we noticed strong peering with Azure and GCP in well-connected geographical regions such as US East, West and London, but not in Asia. Alibaba Cloud and AWS for the most part did not have any direct peerings, irrespective of geography.

TAKEAWAY



AWS, Azure and GCP remain the best inter-connected clouds. Alibaba Cloud and IBM present a mixed architectural picture, with greater reliance on the Internet to connect with other cloud providers.

RECOMMENDATION

Factor in cloud provider peering and direct versus Internet-based connectivity when selecting hosting regions for multi-cloud application architectures.

CONCLUSION AND RECOMMENDATIONS

Network performance is strong across all five public cloud providers, yet important variances exist. Cloud providers are continuously optimizing their networks to improve performance, resulting in shifting (and mostly improving) performance baselines. While the overall performance data is encouraging, it's important to remember that cloud provider network connectivity is still generally a “best effort” service. Furthermore, even cloud giants aren’t exempt from architectural anomalies and events that can negatively impact performance. Continuous monitoring, data-driven planning and a well-defined operational process supported by thorough visibility are essential to success.

FINDING	EXAMPLE	RECOMMENDATION	FINDING	EXAMPLE	RECOMMENDATION
<p>Some clouds rely heavily on the public Internet to transport traffic while others do not.</p> 	Azure and GCP extensively use their backbones to carry user to hosting region traffic. AWS and Alibaba heavily rely on the Internet for user traffic transport, while IBM takes a hybrid approach.	Plan your public cloud connectivity in consideration of your organization's tolerance for exposure to the unpredictable nature of the Internet.	<p>US broadband ISP choice makes a difference in cloud performance. ²</p> 	Broadband performance is relatively consistent across providers, but anomalies do occur, even in the mature US market.	Ensure sound Internet visibility measures as part of hybrid WAN planning and deployment.
<p>Significant cloud performance anomalies exist depending on provider, hosting region, and user locations. ¹</p> 	The GCP network demonstrates higher latency between users in Europe and data center regions in Asia as well as South America.	Include user to hosting region performance data in your public cloud region selection criteria.	<p>All cloud providers, including Alibaba, pay a performance toll when crossing the Great Firewall of China.</p> 	All cloud providers experience heavy packet loss when crossing through China's Great Firewall; however, traffic that does not cross the firewall is subject to very minimal loss.	Use data from the report to evaluate hosting options to serve users in China.
<p>AWS Global Accelerator doesn't always out-perform the Internet.</p> 	AWS Global Accelerator can offer significant performance improvements, but in some cases the paid service performs worse than the public Internet.	Evaluate the performance against your return on investment goals.	<p>Performance Toll for Crossing the Great Firewall</p>		

1. GCP's infrastructure map indicates an update to their connectivity (observed as of November 12th, 2019) between Europe and India. However the results from our end-user and GCP hosting region vantage points across Europe, and Africa did not reflect a change of routes or performance metrics that would correspond with those changes.

2. As of November 8th, the suboptimal peering between GCP and Verizon has been resolved. Refer Appendix for more details.

3. AWS has made significant optimizations to their peering relationship and added multiple new PoP's improving the performance of the Global Accelerator. Refer to Appendix for more details

APPENDIX

UPDATES

GLOBAL END-USER NETWORK PERFORMANCE

HOSTING REGIONS	United States, East
	United States, west
	São Paulo, Brazil
	Canada (Montreal/Quebec City)
	Mumbai/Chennai, India
	London/Cardiff, UK

AWS GLOBAL ACCELERATOR

HOSTING REGIONS	Ashburn, VA
	San Jose, CA
	Frankfurt, Germany
	Sydney, Australia
	Mumbai, India

UPDATES

The results presented in this report have been gathered during the timeframe mentioned in the methodology section and do not reflect any changes made by cloud providers after the data collection period. However, given that there is no steady state in the cloud, it is only fair that we highlight any major changes and optimizations to performance such that readers have an accurate picture of the state of the cloud. In this section, we will call out a few such optimizations.

QUARTERLY UPDATE 1: DECEMBER 15TH, 2019

GOOGLE AND VERIZON PEERING RESOLVED

As of November 8th, 2019, the suboptimal peering exhibited between GCP and Verizon-connected vantage points (as seen in Figure 39 in the Broadband ISP Performance section) in the U.S west coast has been resolved.

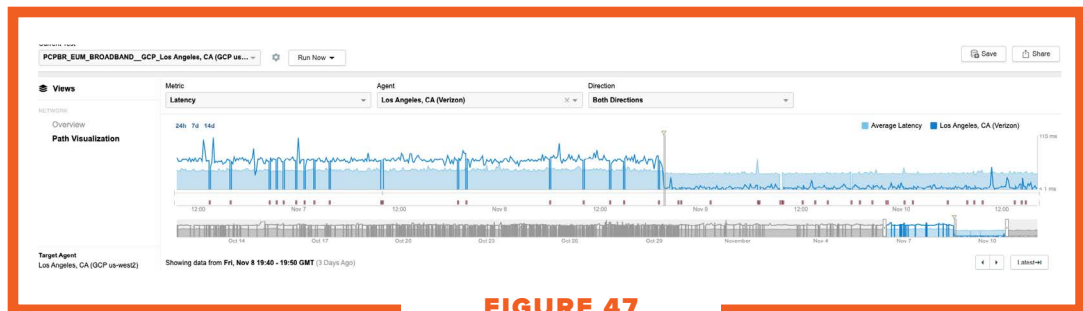


FIGURE 47

Bidirectional latency from Verizon-connected vantage location in Los Angeles to GCP's us-west2 region in Los Angeles reduced from 60ms to 4ms

Verizon-connected agents are no longer handing off traffic into GCP's network in New Jersey (as shown in Figure 39). The new path, seen below in Figure 48, is more optimized with a handoff to GCP closest to the vantage point location, in California, resulting in improvements to latency.

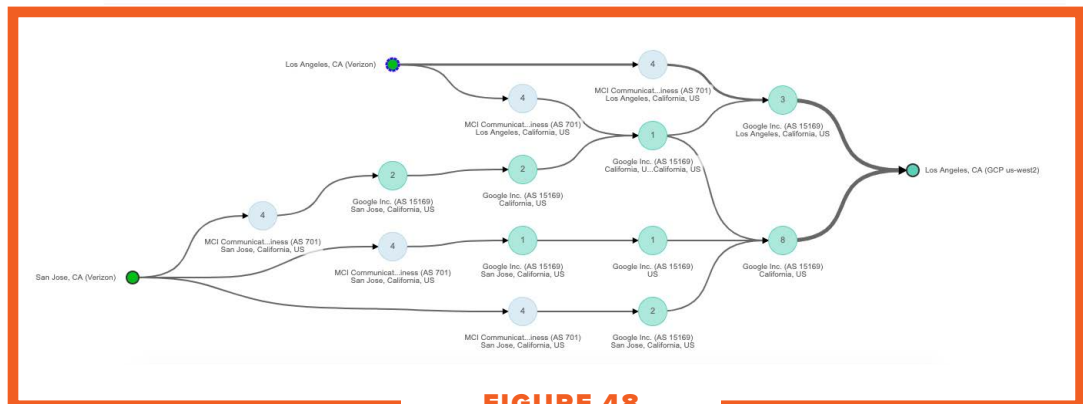


FIGURE 48

As of November 8th, 2019, we see a more optimized route between GCP and Verizon-connected vantage points in the U.S West

QUARTERLY UPDATE 2: APRIL 30TH, 2020

AWS GLOBAL ACCELERATOR PERFORMANCE IMPROVES

Over the last 6 months, the AWS GA team has been making improvements to their network, like improving routing for optimized paths to AWS' edge network, supporting AWS GA PoPs in newer cities or introducing performance enhancing features such as TCP termination. In March 2020, we re-ran the AWS GA tests to understand the impacts of these optimizations. Later in the Appendix section, detailed AWS Global Accelerator measurements including network latency, jitter and standard deviation have been updated to reflect the most recent data gathered between March 1st, 2020 - March 31st, 2020.

We noticed significant improvements to network performance and network connectivity for a few locations. Highlighted below in Table 1 is a comparison of round-trip network latency from user vantage points to a service in us-east-1 (Ashburn). Let's look at Los Angeles for instance. In October 2019, a user connecting from LA to Ashburn on the AWS GA would have noticed very negligible improvements to network latency, from 74.92ms to 74.45ms. However, latest results show that the same user connecting from Los Angeles to AWS us-east-1 will see an improvement in network latency by 15ms, ~ a 20% improvement from the past state.

Note: Measurements taken from other locations or other ISP connections in similar locations may yield different results.

LATENCY (ms)				
USER VANTAGE POINTS	DEFAULT INTERNET CONNECTIVITY	GLOBAL ACCELERATOR (OCT 2019)	GLOBAL ACCELERATOR (MAR 2020)	6-MONTH IMPROVEMENT
BANGALORE, INDIA (RELIANCE)	219.74	323.27	241.1	25.42%
LOS ANGELES, CA	74.92	74.45	59.54	20.03%
MADRID, SPAIN	93.98	106.11	94.89	10.57%
MUNICH, GERMANY	97.64	91.66	91.29	0.40%

Table 11: Network latency measurements with and without the Global Accelerator show steady improvement over time.

Despite using the AWS GA, traffic from Los Angeles enters AWS backbone in Washington in 2019 (Figure 49). However, March 2020 (Figure 50) tells a different story, traffic from Los Angeles enters the AWS backbone at a peering location earlier in the path and closer to the user.

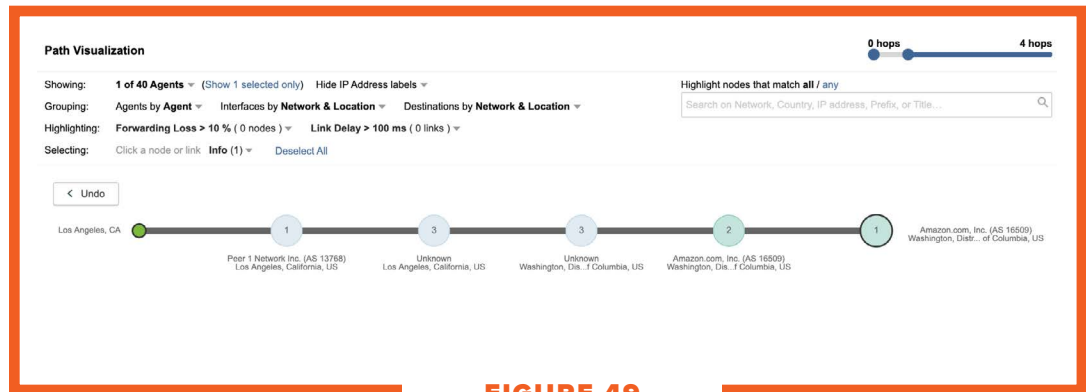


FIGURE 49

October 2019: With AWS GA traffic from Los Angeles enters AWS' network in Washington DC while accessing a service in us-east-1

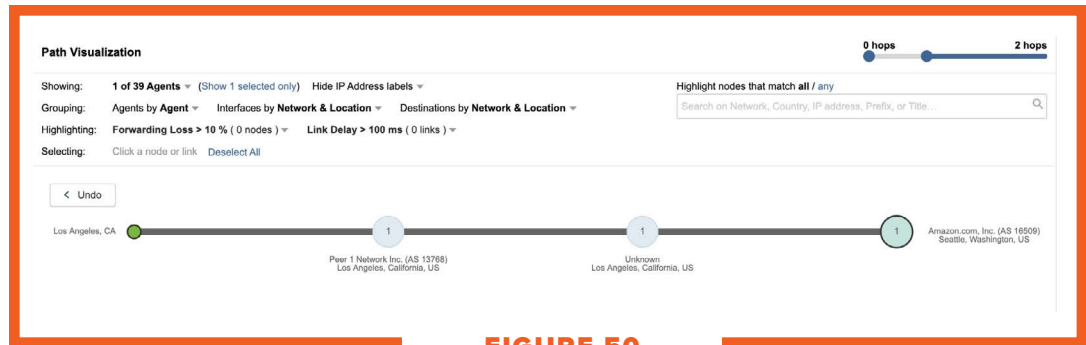


FIGURE 50

March 2020: With AWS GA traffic from Los Angeles enters AWS' network in LA

IMPROVEMENTS IN MICROSOFT AZURE'S INTER-REGION PERFORMANCE

As with any major cloud provider, network optimizations are always underway. Microsoft Azure's continuous improvements on adding new capacity, optimizing routing path selection and optical topology optimizations improved inter-region network latencies as shown in Table 7 in the Inter-Region section. The following region pairs showed noticeable improvements:

	BIDIRECTIONAL LATENCY (MEASURED OCT 2019)	BIDIRECTIONAL LATENCY (MEASURED MAR 2020)
FRANCE CENTRAL - JAPAN EAST	239.32	235.45
FRANCE CENTRAL - SOUTHEAST ASIA	180.26	151.67
CANADA EAST - WEST EUROPE	106.12	103.72
UK WEST - WEST EUROPE	13.38	12.32

Table 12: *Inter-region network latencies improve over the last 6 months*

Comparing the inter-region dispersion across the comparative scale used in this report, the optimizations done by Microsoft Azure showed overall improvements. Between October 2019 (initial set of measurements) and March 2020 (after optimizations), the number of inter-region pairs doing better than the Internet (10% Faster + Baseline) and worse than the Internet (10% slower + 30% slower) increased from 93.33% to 99.05% and decreased from 6.67% to 0.95% respectively.

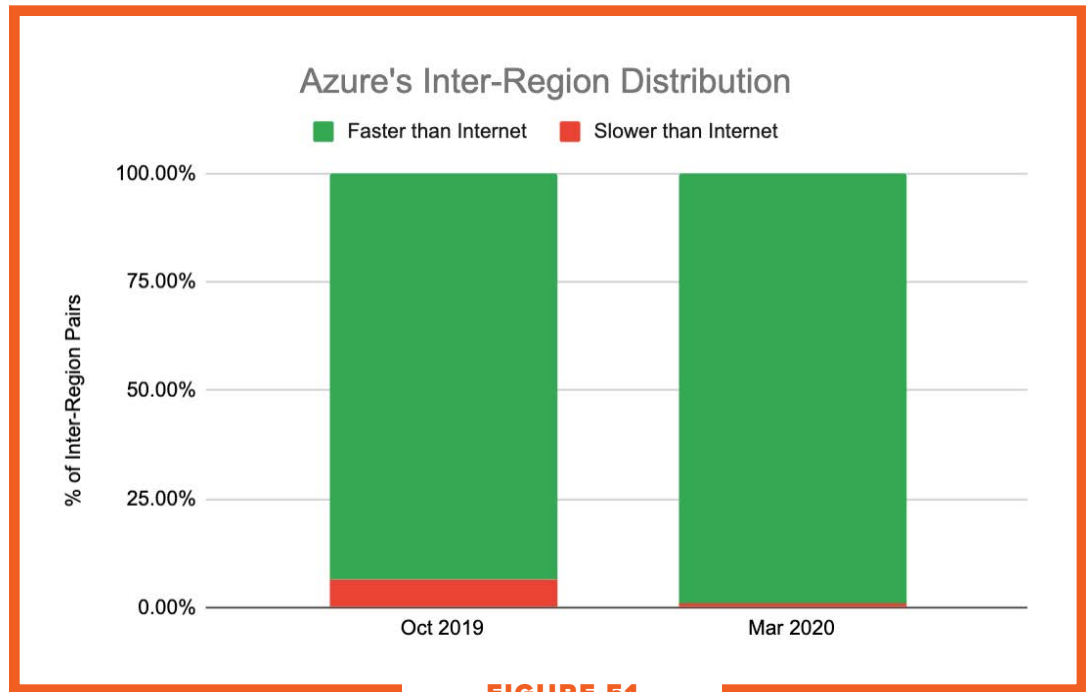


FIGURE 51

Distribution of Azure's inter-region pairs shows improvement

TABLE 13. INTER-REGION NETWORK LATENCY MEASUREMENTS IN MICROSOFT AZURE (UPDATED 2020)

2019 AVERAGE LATENCY MEASUREMENTS	CANADA CENTRALTORONTO, CANADA	BRAZIL SOUTH SÃO PAULO, BRAZIL	EAST US RICHMOND, VA	CENTRAL US DES MOINES, IA	WEST US SANTA CLARA, CA	WEST US 2 QUINCY, WA	NORTH EUROPE DUBLIN, IRELAND	UK WEST CARDIFF, UK	FRANCE CENTRAL PARIS, FRANCE	WEST EUROPE AMSTERDAM, NETHERLANDS	JAPAN EAST TOKYO, JAPAN	KOREA CENTRAL SEOUL, SOUTH KOREA	WEST INDIA MUMBAI, INDIA	SOUTHEAST ASIA SINGAPORE	AUSTRALIA EAST SYDNEY, AUSTRALIA
CANADA CENTRALTORONTO		130.09	24.00	21.97	59.17	57.31	79.20	88.97	92.90	95.98	153.31	178.64	207.51	223.38	194.35
BRAZIL SOUTH SÃO PAULO, BRAZIL			113.30	138.26	171.07	181.17	172.25	183.19	186.88	187.44	264.17	300.06	297.62	329.25	307.66
EAST US – RICHMOND, VA				25.09	59.02	68.89	69.97	80.14	80.92	86.06	164.09	177.89	193.29	227.86	195.69
CENTRAL US – DES MOINES, IA					51.88	43.68	87.86	96.76	100.80	102.33	147.30	165.03	211.93	202.51	188.13
WEST US – SANTA CLARA, CA						23.16*	132.25	139.19	139.91	144.40	105.35	129.64	220.44	167.93	138.37
WEST US 2 – QUINCY, WA							132.03	137.62	147.10	146.77	112.98	122.04	212.60	160.17	158.66
NORTH EUROPE – DUBLIN, IRELAND								11.74	16.37	16.30	220.50	228.38	127.27	166.90	259.37
UK WEST – CARDIFF, UK									9.57	12.32	230.97	222.52	121.50	161.12	258.74
FRANCE CENTRAL – PARIS, FRANCE										11.47	218.52	213.59	112.23	151.67	253.47
WEST EUROPE – AMSTERDAM, NETHERLANDS											235.45	222.25	120.76	160.64	259.64
JAPAN EAST – TOKYO, JAPAN												29.99	120.08	67.88	119.94
KOREA CENTRAL – SEOUL, SOUTH KOREA													115.17	62.89	149.08
WEST INDIA – MUMBAI, INDIA														52.71	140.41
SOUTHEAST ASIA – SINGAPORE															88.31
AUSTRALIA EAST – SYDNEY, AUSTRALIA															

10% faster than baseline Same as baseline 10% slower than baseline 30% slower than baseline

* The baseline locations for this region pair (Seattle - San Jose) adds a 20% overhead. Upon normalization, network latency for the region pair is equivalent to Internet baselines

SUBOPTIMAL REVERSE PATH IMPACTING WORKLOADS IN GCP SÃO PAULO RESOLVED

At the time of initial data collection, between September-October 2019 we noticed that traffic from our vantage point in Rio destined towards the GCP region in São Paulo (southamerica-east1) was experiencing very high network latency. The root cause of this was isolated to a sub-optimal reverse path from GCP São Paulo to our Rio vantage point that included a trans-oceanic route via Miami adding nearly 100ms of latency.

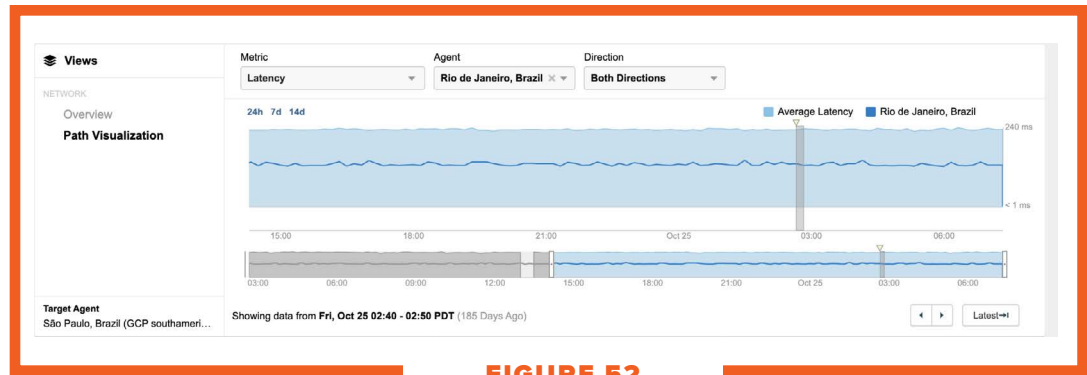


FIGURE 52

Bidirectional network latencies from our vantage point in Rio to GCP's region in São Paulo (southamerica-east1) is approximately 125ms in October 2019

As early as March 2020 we noticed that this sub-optimal routing has been resolved and latencies dropped back to expected normal ranges, ~ 12 ms from our vantage point in Rio to GCP's region in São Paulo (southamerica-east1).



FIGURE 53

Bidirectional network latencies from our vantage point in Rio to GCP's region in São Paulo (southamerica-east1) reduces to 12 ms in April 2020

GLOBAL END-USER NETWORK PERFORMANCE

HOSTING REGION: UNITED STATES, EAST

	LATENCY					JITTER					PACKET LOSS				
	CLOUD PROVIDERS					CLOUD PROVIDERS					CLOUD PROVIDERS				
Continent Name	AliCloud	AWS	Azure	GCP	IBM	AliCloud	AWS	Azure	GCP	IBM	AliCloud	AWS	Azure	GCP	IBM
Africa	211.02	213.40	197.01	204.64	201.13	0.29	0.31	0.41	0.33	0.31	0.82%	1.19%	1.09%	0.98%	0.87%
Asia (ex. China)	205.85	197.38	194.94	188.80	202.35	0.69	0.57	0.58	0.76	0.61	0.33%	0.53%	0.51%	0.23%	0.23%
China	254.25	247.02	236.38	244.19	245.33	1.24	1.12	1.03	1.36	0.99	6.30%	5.11%	3.81%	3.50%	5.14%
Europe	100.26	101.07	98.50	100.16	99.02	0.31	0.29	0.44	0.42	0.34	0.10%	0.07%	0.36%	0.09%	0.06%
North America	35.34	37.82	38.40	35.46	36.79	0.26	0.22	0.36	0.29	0.33	0.17%	0.20%	0.23%	0.20%	0.18%
Oceania	238.04	235.94	217.20	220.05	225.97	0.72	0.90	0.58	0.70	0.68	0.37%	0.40%	0.29%	0.38%	0.24%
South America	139.69	138.50	155.42	138.25	140.52	0.65	0.45	0.88	0.71	0.65	0.44%	0.28%	1.10%	0.40%	0.61%

HOSTING REGION: UNITED STATES, WEST

	LATENCY					JITTER					PACKET LOSS				
	CLOUD PROVIDERS					CLOUD PROVIDERS					CLOUD PROVIDERS				
Continent Name	AliCloud	AWS	Azure	GCP	IBM	AliCloud	AWS	Azure	GCP	IBM	AliCloud	AWS	Azure	GCP	IBM
Africa	273.91	279.28	264.03	259.02	265.63	0.48	0.32	0.42	0.34	0.36	1.41%	1.29%	1.09%	0.98%	1.59%
Asia (ex. China)	180.16	181.03	175.96	166.87	187.94	0.57	0.53	0.49	0.59	0.53	0.40%	0.23%	0.29%	0.16%	0.18%
China	188.96	186.00	183.77	190.54	184.87	0.68	0.75	0.79	1.00	0.65	4.20%	3.57%	4.21%	4.30%	3.18%
Europe	166.85	165.26	167.22	156.15	163.17	0.41	0.30	0.39	0.39	0.41	0.12%	0.07%	0.35%	0.35%	0.30%
North America	54.96	53.84	52.43	49.92	54.79	0.27	0.23	0.35	0.30	0.31	0.18%	0.21%	0.17%	0.20%	0.16%
Oceania	169.07	167.70	168.63	161.32	171.52	0.86	0.50	0.66	0.70	0.43	0.28%	0.30%	0.18%	0.33%	0.16%
South America	194.49	181.80	190.70	160.29	190.44	0.87	0.71	0.93	0.70	0.75	1.01%	1.36%	0.28%	0.40%	1.04%

HOSTING REGION: SÃO PAULO, BRAZIL

	LATENCY				JITTER				PACKET LOSS			
	CLOUD PROVIDERS				CLOUD PROVIDERS				CLOUD PROVIDERS			
Continent Name	AWS	Azure	GCP	IBM	AWS	Azure	GCP	IBM	AWS	Azure	GCP	IBM
Africa	222.35	218.99	206.89	205.24	0.26	0.42	0.28	0.35	0.01	0.01	0.01	0.01
Asia (ex. China)	204.10	206.37	198.59	207.42	0.60	0.62	0.62	0.64	0.00	0.00	0.00	0.00
China	250.70	248.08	249.52	255.19	0.88	1.02	1.17	0.61	0.04	0.05	0.04	0.04
Europe	106.78	121.46	103.38	103.88	0.32	0.42	0.40	0.37	0.00	0.00	0.00	0.00
North America	40.46	53.81	42.61	45.03	0.24	0.38	0.29	0.32	0.00	0.00	0.00	0.00
Oceania	238.28	234.29	227.57	230.55	1.14	0.73	1.16	0.74	0.00	0.00	0.00	0.00
South America	151.54	179.80	150.16	150.54	0.60	0.91	0.88	0.74	0.01	0.00	0.01	0.01

HOSTING REGION: CANADA (MONTREAL/QUEBEC CITY)

	LATENCY				JITTER				PACKET LOSS			
	CLOUD PROVIDERS				CLOUD PROVIDERS				CLOUD PROVIDERS			
Continent Name	AWS	Azure	GCP	IBM	AWS	Azure	GCP	IBM	AWS	Azure	GCP	IBM
Africa	222.35	218.99	206.89	205.24	0.26	0.42	0.28	0.35	0.01	0.01	0.01	0.01
Asia (ex. China)	204.10	206.37	198.59	207.42	0.60	0.62	0.62	0.64	0.00	0.00	0.00	0.00
China	250.70	248.08	249.52	255.19	0.88	1.02	1.17	0.61	0.04	0.05	0.04	0.04
Europe	106.78	121.46	103.38	103.88	0.32	0.42	0.40	0.37	0.00	0.00	0.00	0.00
North America	40.46	53.81	42.61	45.03	0.24	0.38	0.29	0.32	0.00	0.00	0.00	0.00
Oceania	238.28	234.29	227.57	230.55	1.14	0.73	1.16	0.74	0.00	0.00	0.00	0.00
South America	151.54	179.80	150.16	150.54	0.60	0.91	0.88	0.74	0.01	0.00	0.01	0.01

HOSTING REGION: MUMBAI/CHENNAI, INDIA

	LATENCY					JITTER					PACKET LOSS				
	CLOUD PROVIDERS					CLOUD PROVIDERS					CLOUD PROVIDERS				
Continent Name	AliCloud	AWS	Azure	GCP	IBM	AliCloud	AWS	Azure	GCP	IBM	AliCloud	AWS	Azure	GCP	IBM
Africa	268.92	259.90	238.80	468.60	303.85	0.27	0.27	0.29	0.30	0.37	0.01	0.01	0.01	0.01	0.01
Asia (ex. China)	136.24	129.31	127.75	159.36	132.96	0.52	0.48	0.52	0.51	0.54	0.01	0.00	0.00	0.00	0.01
China	297.84	233.83	152.01	276.26	174.34	0.97	0.95	0.73	1.00	0.81	0.05	0.05	0.04	0.06	0.06
Europe	144.01	136.16	139.38	361.52	204.79	0.39	0.32	0.37	0.53	0.47	0.00	0.00	0.00	0.00	0.00
North America	243.60	237.75	234.47	262.01	245.58	0.42	0.35	0.48	0.38	0.43	0.00	0.00	0.00	0.00	0.00
Oceania	241.88	237.57	162.25	174.12	243.85	1.18	0.89	0.74	0.98	0.58	0.01	0.00	0.00	0.00	0.00
South America	346.44	333.26	341.40	379.40	368.81	0.92	0.78	1.22	0.86	1.07	0.01	0.00	0.01	0.01	0.01

HOSTING REGION: LONDON/CARDIFF, UK

	LATENCY					JITTER					PACKET LOSS				
	CLOUD PROVIDERS					CLOUD PROVIDERS					CLOUD PROVIDERS				
Continent Name	AliCloud	AWS	Azure	GCP	IBM	AliCloud	AWS	Azure	GCP	IBM	AliCloud	AWS	Azure	GCP	IBM
Africa	134.94	131.20	124.23	132.11	130.86	0.26	0.29	0.38	0.31	0.35	0.01	0.01	0.01	0.01	0.01
Asia (ex. China)	210.35	179.61	196.66	219.11	184.83	0.54	0.76	0.56	0.59	0.64	0.00	0.00	0.00	0.00	0.01
China	256.38	251.31	270.44	301.92	254.97	0.96	1.12	1.03	1.07	0.98	0.05	0.03	0.05	0.04	0.04
Europe	26.83	25.44	29.67	26.52	25.49	0.28	0.30	0.41	0.31	0.34	0.00	0.01	0.00	0.00	0.00
North America	108.20	110.72	109.90	104.99	109.60	0.28	0.28	0.43	0.32	0.33	0.00	0.00	0.00	0.00	0.00
Oceania	284.50	276.52	296.45	289.02	302.95	1.20	1.25	0.87	0.51	1.14	0.00	0.00	0.00	0.00	0.00
South America	214.25	221.73	226.23	215.36	221.31	0.79	0.95	1.07	0.94	0.87	0.00	0.00	0.00	0.00	0.00

AWS GLOBAL ACCELERATOR

HOSTING REGION: ASHBURN, VA

	LATENCY				JITTER				STD DEV			
Sources	Default Internet Connection	Global Accelerator Connection	Latency Improvement	% Improvement	Default Internet Connection	Global Accelerator Connection	Jitter Improvement	% Improvement	Default Internet Connection	Global Accelerator Connection	Std Dev Improvement	% improvement
Amsterdam, Netherlands	89.04	81.05	7.99	8.97%	0.06	0.09	-0.03	-50.00%	2.54	0.64	1.9	74.80%
Atlanta, GA	13.19	13.44	-0.25	-1.90%	0.03	0.02	0.01	33.33%	0.75	1.63	-0.88	-117.33%
Bangalore, India (Reliance)	219.74	323.27	-103.53	-47.11%	0.72	1.17	-0.45	-62.50%	17.84	85.84	-68	-381.17%
Barcelona, Spain	109.77	106.59	3.18	2.90%	3.15	2.83	0.32	10.16%	56.29	47.49	8.8	15.63%
Beijing, China (China Mobile)	267.16	247.78	19.38	7.25%	15.57	31.02	-15.45	-99.23%	49.54	72.17	-22.63	-45.68%
Beijing, China (China Telecom)	240.05	281.03	-40.98	-17.07%	21.51	25.82	-4.31	-20.04%	71.94	101.28	-29.34	-40.78%
Boston, MA	11.95	12.8	-0.85	-7.11%	0.02	0.03	-0.01	-50.00%	0.22	2.24	-2.02	-918.18%
Brussels, Belgium	89.61	84.92	4.69	5.23%	1.06	0.41	0.65	61.32%	4.65	1.61	3.04	65.38%
Chengdu, China (China Mobile)	287.41	270	17.41	6.06%	15.04	1.92	13.12	87.23%	75.45	21.56	53.89	71.42%
Chicago, IL	32.78	25.49	7.29	22.24%	2.07	1.33	0.74	35.75%	5.91	5.34	0.57	9.64%
Columbus, OH	24.61	13.42	11.19	45.47%	0.47	0.42	0.05	10.64%	1.8	3.6	-1.8	-100.00%
Copenhagen, Denmark	93.09	96.46	-3.37	-3.62%	0.05	0.05	0	0.00%	2.89	2.29	0.6	20.76%
Dallas, TX	29.99	30.07	-0.08	-0.27%	0.07	0.07	0	0.00%	1.7	1.79	-0.09	-5.29%
Dubai, United Arab Emirates	193.79	186.8	6.99	3.61%	0.44	0.24	0.2	45.45%	13.85	4.14	9.71	70.11%
Dublin, Ireland	116.18	90.31	25.87	22.27%	1.8	1.74	0.06	3.33%	17.86	17.93	-0.07	-0.39%
Hyderabad, India	222.12	224.13	-2.01	-0.90%	9.5	13.58	-4.08	-42.95%	32.66	30.53	2.13	6.52%
Johannesburg, South Africa	262.63	233.46	29.17	11.11%	0.25	0.15	0.1	40.00%	10.33	1.77	8.56	82.87%
Lisbon, Portugal	94.31	109.51	-15.2	-16.12%	0.19	0.7	-0.51	-268.42%	1.63	2.59	-0.96	-58.90%
London, England	74.66	75.75	-1.09	-1.46%	0.09	0.09	0	0.00%	1.06	0.54	0.52	49.06%
Los Angeles, CA	74.92	74.45	0.47	0.63%	0.42	0.44	-0.02	-4.76%	9.51	9.53	-0.02	-0.21%
Madrid, Spain	93.98	106.11	-12.13	-12.91%	0.38	0.47	-0.09	-23.68%	2.93	10.12	-7.19	-245.39%
Montreal, Canada	17.84	17.33	0.51	2.86%	0.42	0.19	0.23	54.76%	2.46	2.8	-0.34	-13.82%
Munich, Germany	97.64	91.66	5.98	6.12%	0.01	0.05	-0.04	-400.00%	1.32	0.51	0.81	61.36%
New Delhi, India	224.81	216.84	7.97	3.55%	0.11	0.15	-0.04	-36.36%	7.56	46	-38.44	-508.47%
New York, NY (Cogent)	7.01	6.31	0.7	9.99%	0.01	0.08	-0.07	-700.00%	0.04	1.37	-1.33	-3325.00%
Paris, France	83.9	81.74	2.16	2.57%	0.38	0.35	0.03	7.89%	2.68	0.59	2.09	77.99%
Portland, OR	66.13	69.33	-3.2	-4.84%	0.18	0.16	0.02	11.11%	3.32	5.26	-1.94	-58.43%
Raleigh, NC	6.93	7	-0.07	-1.01%	0.05	0.01	0.04	80.00%	0.21	0.05	0.16	76.19%
San Francisco, CA	69.88	62.37	7.51	10.75%	0.08	0.07	0.01	12.50%	1.35	0.94	0.41	30.37%
Seattle, WA	74.54	66.47	8.07	10.83%	0.09	0.05	0.04	44.44%	6.89	4.86	2.03	29.46%
Seoul, South Korea	267.87	179.29	88.58	33.07%	14.48	0.25	14.23	98.27%	70.68	5.6	65.08	92.08%
Singapore	256.14	217.89	38.25	14.93%	0.5	0.14	0.36	72.00%	10.2	5.03	5.17	50.69%
Sydney, Australia	225.64	202.11	23.53	10.43%	0.87	0.23	0.64	73.56%	4.93	3.95	0.98	19.88%
Tokyo, Japan	156.83	147.4	9.43	6.01%	0.1	0.13	-0.03	-30.00%	2.65	4.69	-2.04	-76.98%
Toronto, Canada	35.96	36.08	-0.12	-0.33%	0.11	0.04	0.07	63.64%	5.57	5.83	-0.26	-4.67%

HOSTING REGION: SAN JOSE, CA

Sources	LATENCY				JITTER				STD DEV			
	Default Internet Connection	Global Accelerator Connection	Latency Improvement	% Improvement	Default Internet Connection	Global Accelerator Connection	Jitter Improvement	% Improvement	Default Internet Connection	Global Accelerator Connection	Std Dev Improvement	% improvement
Amsterdam, Netherlands	140.17	140.91	-0.74	-0.53%	0.08	0.07	0.01	12.50%	2.98	1.58	1.4	46.98%
Atlanta, GA	61.89	55.44	6.45	10.42%	0.12	0.01	0.11	91.67%	2.58	2.91	-0.33	-12.79%
Bangalore, India (Reliance)	234.91	277.8	-42.89	-18.26%	0.84	0.51	0.33	39.29%	9.03	84.73	-75.7	-838.32%
Barcelona, Spain	175.03	164.66	10.37	5.92%	2.47	3.39	-0.92	-37.25%	30.65	41.01	-10.36	-33.80%
Beijing, China (China Mobile)	196.48	233.86	-37.38	-19.02%	37.04	30.89	6.15	16.60%	66.39	105.05	-38.66	-58.23%
Beijing, China (China Telecom)	170.98	213.6	-42.62	-24.93%	9.59	15.81	-6.22	-64.86%	37	72.02	-35.02	-94.65%
Boston, MA	74.47	85.61	-11.14	-14.96%	0.04	0.06	-0.02	-50.00%	1.04	5.36	-4.32	-415.38%
Brussels, Belgium	145.82	144.82	1	0.69%	0.55	0.28	0.27	49.09%	3.51	1.76	1.75	49.86%
Chengdu, China (China Mobile)	211.93	225.35	-13.42	-6.33%	0.88	2.26	-1.38	-156.82%	9.93	19.18	-9.25	-93.15%
Chicago, IL	72.49	60.78	11.71	16.15%	2.59	1.58	1.01	39.00%	7.07	3.66	3.41	48.23%
Columbus, OH	69.45	73.54	-4.09	-5.89%	0.25	0.49	-0.24	-96.00%	0.94	1.62	-0.68	-72.34%
Copenhagen, Denmark	160.67	154.9	5.77	3.59%	0.07	0.06	0.01	14.29%	2.49	2.68	-0.19	-7.63%
Dallas, TX	49.01	38.28	10.73	21.89%	0.01	0.08	-0.07	-700.00%	0.18	0.41	-0.23	-127.78%
Dubai, United Arab Emirates	259.09	244.81	14.28	5.51%	0.77	0.41	0.36	46.75%	31.06	10.58	20.48	65.94%
Dublin, Ireland	166	150.67	15.33	9.23%	2.21	1.91	0.3	13.57%	37.41	18.18	19.23	51.40%
Hyderabad, India	255.85	241.77	14.08	5.50%	20.06	22.03	-1.97	-9.82%	40.46	39.08	1.38	3.41%
Johannesburg, South Africa	325.07	293.48	31.59	9.72%	0.5	0.15	0.35	70.00%	14.91	2.13	12.78	85.71%
Lisbon, Portugal	148.3	168.11	-19.81	-13.36%	0.29	0.71	-0.42	-144.83%	1.15	2.59	-1.44	-125.22%
London, England	138.33	135.8	2.53	1.83%	0.03	0.07	-0.04	-133.33%	3.61	1.3	2.31	63.99%
Los Angeles, CA	7.19	131.32	-124.13	-1726.43%	0.21	0.33	-0.12	-57.14%	3.15	23.22	-20.07	-637.14%
Madrid, Spain	179.12	166.02	13.1	7.31%	0.36	0.37	-0.01	-2.78%	9.08	4.47	4.61	50.77%
Montreal, Canada	78.93	87.73	-8.8	-11.15%	0.33	0.31	0.02	6.06%	2.13	3.98	-1.85	-86.85%
Munich, Germany	163.31	150.12	13.19	8.08%	0.13	0.03	0.1	76.92%	1.22	1.24	-0.02	-1.64%
New Delhi, India	262.84	254.1	8.74	3.33%	0.6	0.11	0.49	81.67%	10.29	30.98	-20.69	-201.07%
New York, NY (Cogent)	70.18	78.61	-8.43	-12.01%	0.03	0.06	-0.03	-100.00%	0.49	5.52	-5.03	-1026.53%
Paris, France	145.7	140.15	5.55	3.81%	0.36	0.38	-0.02	-5.56%	0.63	2.6	-1.97	-312.70%
Portland, OR	15.64	23.02	-7.38	-47.19%	0.08	0.11	-0.03	-37.50%	0.76	0.87	-0.11	-14.47%
Raleigh, NC	71.98	66.81	5.17	7.18%	0.14	0.02	0.12	85.71%	0.82	0.87	-0.05	-6.10%
San Francisco, CA	1.5	1.83	-0.33	-22.00%	0.02	0.09	-0.07	-350.00%	0.51	0.34	0.17	33.33%
Seattle, WA	25.31	21.05	4.26	16.83%	0.04	0.06	-0.02	-50.00%	1.73	2.14	-0.41	-23.70%
Seoul, South Korea	135.85	136.29	-0.44	-0.32%	2.28	0.26	2.02	88.60%	40.84	1.23	39.61	96.99%
Singapore	185.66	174.2	11.46	6.17%	0.32	0.24	0.08	25.00%	6.99	0.73	6.26	89.56%
Sydney, Australia	152.42	147.8	4.62	3.03%	0.69	0.14	0.55	79.71%	4.45	3.08	1.37	30.79%
Tokyo, Japan	99.51	111.35	-11.84	-11.90%	0.07	0.12	-0.05	-71.43%	2.51	2.01	0.5	19.92%
Toronto, Canada	74.49	69.39	5.1	6.85%	0.1	0.03	0.07	70.00%	0.84	1.18	-0.34	-40.48%



HOSTING REGION: FRANKFURT, GERMANY

Sources	LATENCY				JITTER				STD DEV			
	Default Internet Connection	Global Accelerator Connection	Latency Improvement	% Improvement	Default Internet Con- nection	Global Accelerator Connection	Jitter Improvement	% Improvement	Default Internet Con- nection	Global Accelerator Connection	Std Dev Improvement	% improvement
Amsterdam, Netherlands	8.11	16.66	-8.55	-105.43%	0.28	0.08	0.2	71.43%	11.64	1.05	10.59	90.98%
Atlanta, GA	99.73	83.71	16.02	16.06%	0.03	0.03	0	0.00%	1.57	3.42	-1.85	-117.83%
Bangalore, India (Reliance)	175.03	282.58	-107.55	-61.45%	0.97	0.98	-0.01	-1.03%	21.32	86	-64.68	-303.38%
Barcelona, Spain	37.05	44.51	-7.46	-20.13%	3.07	2.33	0.74	24.10%	74.29	49.76	24.53	33.02%
Beijing, China (China Mobile)	201.78	335.82	-134.04	-66.43%	33.88	43.67	-9.79	-28.90%	69.95	81.29	-11.34	-16.21%
Beijing, China (China Telecom)	168.87	310.98	-142.11	-84.15%	13.6	21.5	-7.9	-58.09%	57.77	72.51	-14.74	-25.51%
Boston, MA	86.12	70.67	15.45	17.94%	0.07	0.03	0.04	57.14%	1.95	3.3	-1.35	-69.23%
Brussels, Belgium	16.63	20.78	-4.15	-24.95%	0.96	0.92	0.04	4.17%	4.33	2.8	1.53	35.33%
Chengdu, China (China Mobile)	215.14	342.92	-127.78	-59.39%	0.94	14.65	-13.71	-1458.51%	35	48.6	-13.6	-38.86%
Chicago, IL	96.52	85.74	10.78	11.17%	2.95	2.28	0.67	22.71%	6.34	5.51	0.83	13.09%
Columbus, OH	107.13	90.19	16.94	15.81%	0.54	0.27	0.27	50.00%	2.35	3.59	-1.24	-52.77%
Copenhagen, Denmark	13.1	34.69	-21.59	-164.81%	0.11	0.21	-0.1	-90.91%	0.2	2.19	-1.99	-995.00%
Dallas, TX	127.98	105.42	22.56	17.63%	0.04	0.07	-0.03	-75.00%	2.68	2.61	0.07	2.61%
Dubai, United Arab Emirates	113.8	125.12	-11.32	-9.95%	0.18	0.66	-0.48	-266.67%	6.91	19.65	-12.74	-184.37%
Dublin, Ireland	28.13	26.52	1.61	5.72%	1.6	1.52	0.08	5.00%	16.16	18.02	-1.86	-11.51%
Hyderabad, India	160.78	163.5	-2.72	-1.69%	12.79	12.67	0.12	0.94%	40.25	34.93	5.32	13.22%
Johannesburg, South Africa	187.6	169.47	18.13	9.66%	0.14	0.22	-0.08	-57.14%	1.58	2.41	-0.83	-52.53%
Lisbon, Portugal	43.24	47.9	-4.66	-10.78%	0.09	0.62	-0.53	-588.89%	1.66	2.99	-1.33	-80.12%
London, England	13.2	11.57	1.63	12.35%	0.05	0.06	-0.01	-20.00%	1.33	0.65	0.68	51.13%
Los Angeles, CA	168.73	150.55	18.18	10.77%	0.05	0.52	-0.47	-940.00%	7.55	9.97	-2.42	-32.05%
Madrid, Spain	32.58	42.12	-9.54	-29.28%	0.54	0.66	-0.12	-22.22%	4.34	6.68	-2.34	-53.92%
Montreal, Canada	94.27	78.68	15.59	16.54%	0.36	1.09	-0.73	-202.78%	2.17	3.99	-1.82	-83.87%
Munich, Germany	9.27	28.55	-19.28	-207.98%	0.02	0.04	-0.02	-100.00%	1.24	1.12	0.12	9.68%
New Delhi, India	150.45	155.52	-5.07	-3.37%	0.42	0.16	0.26	61.90%	18.23	49.42	-31.19	-171.09%
New York, NY (Cogent)	83.25	66.96	16.29	19.57%	0.05	0.14	-0.09	-180.00%	1.12	2.31	-1.19	-106.25%
Paris, France	11.27	19.79	-8.52	-75.60%	0.27	0.29	-0.02	-7.41%	1.7	0.9	0.8	47.06%
Portland, OR	152.49	125.65	26.84	17.60%	0.62	0.11	0.51	82.26%	19.64	8.04	11.6	59.06%
Raleigh, NC	93.24	82.62	10.62	11.39%	0.02	0.03	-0.01	-50.00%	2.06	2.83	-0.77	-37.38%
San Francisco, CA	153.56	139.73	13.83	9.01%	0.1	0.09	0.01	10.00%	1.83	2.52	-0.69	-37.70%
Seattle, WA	152.57	126.83	25.74	16.87%	0.06	0.16	-0.1	-166.67%	1.05	8.07	-7.02	-668.57%
Seoul, South Korea	310.92	233.84	77.08	24.79%	2.33	0.43	1.9	81.55%	26.25	8.83	17.42	66.36%
Singapore	173.29	177.85	-4.56	-2.63%	7.59	0.27	7.32	96.44%	96.86	8.36	88.5	91.37%
Sydney, Australia	247.12	257.04	-9.92	-4.01%	0.61	0.25	0.36	59.02%	2.01	4.9	-2.89	-143.78%
Tokyo, Japan	209.39	202.48	6.91	3.30%	0.76	0.09	0.67	88.16%	16.12	6.92	9.2	57.07%
Toronto, Canada	113.47	93.85	19.62	17.29%	0.06	0.07	-0.01	-16.67%	2.24	4.16	-1.92	-85.71%

HOSTING REGION: SYDNEY, AUSTRALIA

Sources	LATENCY				JITTER				STD DEV			
	Default Internet Connection	Global Accelerator Connection	Latency Improvement	% Improvement	Default Internet Connection	Global Accelerator Connection	Jitter Improvement	% Improvement	Default Internet Connection	Global Accelerator Connection	Std Dev Improvement	% improvement
Amsterdam, Netherlands	291	271.67	19.33	6.64%	0.06	0.1	-0.04	-66.67%	8.49	3.19	5.3	62.43%
Atlanta, GA	248.22	186.76	61.46	24.76%	0.28	0.05	0.23	82.14%	14.37	8.53	5.84	40.64%
Bangalore, India (Reliance)	245.95	276.26	-30.31	-12.32%	0.47	0.55	-0.08	-17.02%	11.69	85.39	-73.7	-630.45%
Barcelona, Spain	298.87	304.82	-5.95	-1.99%	2.99	2.55	0.44	14.72%	59.81	81.81	-22	-36.78%
Beijing, China (China Mobile)	251.65	328.44	-76.79	-30.51%	27.73	27.87	-0.14	-0.50%	73.08	48.86	24.22	33.14%
Beijing, China (China Telecom)	254.4	389	-134.6	-52.91%	34.28	70.17	-35.89	-104.70%	107.48	144.11	-36.63	-34.08%
Boston, MA	222.8	203.59	19.21	8.62%	0.06	0.03	0.03	50.00%	6.13	4.49	1.64	26.75%
Brussels, Belgium	321.17	276.24	44.93	13.99%	1.53	1.34	0.19	12.42%	8.8	4.48	4.32	49.09%
Chengdu, China (China Mobile)	208.96	291.51	-82.55	-39.51%	2.18	3.68	-1.5	-68.81%	19.36	76.12	-56.76	-293.18%
Chicago, IL	213.59	181.25	32.34	15.14%	4.28	2.85	1.43	33.41%	15.55	11.83	3.72	23.92%
Columbus, OH	252.23	210.22	42.01	16.66%	0.64	0.2	0.44	68.75%	10.73	5.55	5.18	48.28%
Copenhagen, Denmark	308.54	294.08	14.46	4.69%	0.08	0.17	-0.09	-112.50%	6.24	2.31	3.93	62.98%
Dallas, TX	180.64	169.29	11.35	6.28%	0.05	0.01	0.04	80.00%	6.95	7.36	-0.41	-5.90%
Dubai, United Arab Emirates	333.11	261.1	72.01	21.62%	0.96	0.41	0.55	57.29%	34.16	25.18	8.98	26.29%
Dublin, Ireland	341.94	282.37	59.57	17.42%	2.43	2.05	0.38	15.64%	31.26	17.72	13.54	43.31%
Hyderabad, India	444.35	237.12	207.23	46.64%	17.22	14.57	2.65	15.39%	49.06	36.29	12.77	26.03%
Johannesburg, South Africa	495.18	423.68	71.5	14.44%	0.68	0.23	0.45	66.18%	11.4	3.74	7.66	6719%
Lisbon, Portugal	316.39	306.89	9.5	3.00%	0.18	0.74	-0.56	-311.11%	10.07	3.57	6.5	64.55%
London, England	284.82	265.68	19.14	6.72%	0.07	0.09	-0.02	-28.57%	6.5	3.23	3.27	50.31%
Los Angeles, CA	162.19	267.55	-105.36	-64.96%	0.06	0.39	-0.33	-550.00%	14.06	23.67	-9.61	-68.35%
Madrid, Spain	345.22	296.12	49.1	14.22%	0.68	0.55	0.13	19.12%	7.59	5	2.59	34.12%
Montreal, Canada	244.91	207.83	37.08	15.14%	0.76	0.27	0.49	64.47%	5.79	4.44	1.35	23.32%
Munich, Germany	340.55	289.34	51.21	15.04%	0.34	0.04	0.3	88.24%	9.6	3.21	6.39	66.56%
New Delhi, India	241.65	253.61	-11.96	-4.95%	0.23	0.13	0.1	43.48%	8.48	31.74	-23.26	-274.29%
New York, NY (Cogent)	246.83	197.14	49.69	20.13%	0.21	0.03	0.18	85.71%	11.69	4.32	7.37	63.05%
Paris, France	297.34	279.34	18	6.05%	0.12	0.41	-0.29	-241.67%	6.81	2.72	4.09	60.06%
Portland, OR	182.98	141.24	41.74	22.81%	0.16	0.06	0.1	62.50%	6.43	2.36	4.07	63.30%
Raleigh, NC	217.54	203.25	14.29	6.57%	0.02	0.08	-0.06	-300.00%	6.82	4.1	2.72	39.88%
San Francisco, CA	194.3	149.24	45.06	23.19%	0.07	0.12	-0.05	-71.43%	5.66	3.23	2.43	42.93%
Seattle, WA	182.77	139.55	43.22	23.65%	0.08	0.04	0.04	50.00%	15.57	3.8	11.77	75.59%
Seoul, South Korea	173.55	146.87	26.68	15.37%	2.7	0.45	2.25	83.33%	32.97	5.4	27.57	83.62%
Singapore	171.91	170.22	1.69	0.98%	0.08	0.2	-0.12	-150.00%	3.53	1.38	2.15	60.91%
Sydney, Australia	1.04	1.16	-0.12	-11.54%	0.11	0.08	0.03	27.27%	0.24	1.36	-1.12	-466.67%
Tokyo, Japan	102.62	105.69	-3.07	-2.99%	0.3	0.09	0.21	70.00%	2.69	1.84	0.85	31.60%
Toronto, Canada	239.38	188.39	50.99	21.30%	0.2	0.01	0.19	95.00%	8.68	3.76	4.92	56.68%



HOSTING REGION: MUMBAI, INDIA

Sources	LATENCY				JITTER				STD DEV			
	Default Internet Connection	Global Accelerator Connection	Latency Improvement	% Improvement	Default Internet Connection	Global Accelerator Connection	Jitter Improvement	% Improvement	Default Internet Connection	Global Accelerator Connection	Std Dev Improvement	% improvement
Amsterdam, Netherlands	166.26	124.34	41.92	25.21%	0.1	0.09	0.01	10.00%	20.59	2.91	17.68	85.87%
Atlanta, GA	206.81	193.77	13.04	6.31%	0.04	0.06	-0.02	-50.00%	2.75	1.5	1.25	45.45%
Bangalore, India (Reliance)	87	166.47	-79.47	-91.34%	0.25	0.4	-0.15	-60.00%	4.53	85.86	-81.33	-1795.36%
Barcelona, Spain	209.61	123.38	86.23	41.14%	3.91	1.89	2.02	51.66%	59.97	59.94	0.03	0.05%
Beijing, China (China Mobile)	164.58	507.72	-343.14	-208.49%	15.83	92.09	-76.26	-481.74%	41.77	139.37	-97.6	-233.66%
Beijing, China (China Telecom)	278.59	496.36	-217.77	-78.17%	22.92	111.4	-88.48	-386.04%	73.24	176.35	-103.11	-140.78%
Boston, MA	193.85	193.21	0.64	0.33%	0.04	0.04	0	0.00%	2.89	2.4	0.49	16.96%
Brussels, Belgium	122.68	128.96	-6.28	-5.12%	0.37	1.25	-0.88	-237.84%	3.3	4.82	-1.52	-46.06%
Chengdu, China (China Mobile)	256.86	344.16	-87.3	-33.99%	5.99	3.69	2.3	38.40%	44.69	106.31	-61.62	-137.88%
Chicago, IL	205.33	207.18	-1.85	-0.90%	4.37	2.6	1.77	40.50%	8.1	3.6	4.5	55.56%
Columbus, OH	262.12	194	68.12	25.99%	0.61	0.48	0.13	21.31%	25.53	4.02	21.51	84.25%
Copenhagen, Denmark	129.47	128.93	0.54	0.42%	0.05	0.07	-0.02	-40.00%	2.61	3.68	-1.07	-41.00%
Dallas, TX	255.87	210.7	45.17	17.65%	0.23	0.09	0.14	60.87%	7.14	2.4	4.74	66.39%
Dubai, United Arab Emirates	33.94	26.27	7.67	22.60%	0.86	0.24	0.62	72.09%	24.79	0.57	24.22	97.70%
Dublin, Ireland	139.7	126.15	13.55	9.70%	1.93	1.94	-0.01	-0.52%	18.83	17.92	0.91	4.83%
Hyderabad, India	30.29	41.07	-10.78	-35.59%	12.76	8.37	4.39	34.40%	32.56	29.72	2.84	8.72%
Johannesburg, South Africa	282.52	284.45	-1.93	-0.68%	0.25	0.22	0.03	12.00%	12.04	3.82	8.22	68.27%
Lisbon, Portugal	136.52	127.31	9.21	6.75%	0.08	0.68	-0.6	-750.00%	3.14	3.67	-0.53	-16.88%
London, England	122.8	110.35	12.45	10.14%	0.05	0.08	-0.03	-60.00%	3.03	1.68	1.35	44.55%
Los Angeles, CA	265.77	254.75	11.02	4.15%	0.36	0.56	-0.2	-55.56%	8.48	10.64	-2.16	-25.47%
Madrid, Spain	138.52	140.71	-2.19	-1.58%	0.56	0.46	0.1	17.86%	8.1	4.11	3.99	49.26%
Montreal, Canada	205.66	197.85	7.81	3.80%	0.5	0.58	-0.08	-16.00%	8.03	3.1	4.93	61.39%
Munich, Germany	121.75	123.84	-2.09	-1.72%	0.04	0.03	0.01	25.00%	2.56	2.95	-0.39	-15.23%
New Delhi, India	51.41	34.55	16.86	32.80%	0.19	0.08	0.11	57.89%	10.46	48.97	-38.51	-368.16%
New York, NY (Cogent)	189.58	186.92	2.66	1.40%	0.09	0.04	0.05	55.56%	2.92	1.72	1.2	41.10%
Paris, France	118.12	108.83	9.29	7.86%	0.38	0.33	0.05	13.16%	2.57	1.22	1.35	52.53%
Portland, OR	268.53	220.07	48.46	18.05%	0.04	0.09	-0.05	-125.00%	3.27	3.42	-0.15	-4.59%
Raleigh, NC	257.01	187.16	69.85	27.18%	0.13	0.08	0.05	38.46%	6.94	2	4.94	71.18%
San Francisco, CA	259.24	230.73	28.51	11.00%	0.39	0.07	0.32	82.05%	18.03	4.96	13.07	72.49%
Seattle, WA	256.46	220.47	35.99	14.03%	0.08	0.08	0	0.00%	3.28	3.64	-0.36	-10.98%
Seoul, South Korea	262.95	136.04	126.91	48.26%	0.57	0.29	0.28	49.12%	31.61	18.91	12.7	40.18%
Singapore	60.32	58.73	1.59	2.64%	0.19	0.18	0.01	5.26%	2.47	3.12	-0.65	-26.32%
Sydney, Australia	188	225.43	-37.43	-19.91%	0.69	0.37	0.32	46.38%	4.52	3.97	0.55	12.17%
Tokyo, Japan	129.33	125.88	3.45	2.67%	0.2	0.1	0.1	50.00%	3.64	3.14	0.5	13.74%
Toronto, Canada	248.41	215.54	32.87	13.23%	0.14	0.05	0.09	64.29%	9.26	2.12	7.14	77.11%





201 Mission Street, Suite 1700
San Francisco, CA 94105
(415) 231-5674

www.thousandeyes.com

About ThousandEyes

ThousandEyes delivers Network Intelligence—immediate visibility into experience for every user and application delivered over any network, so companies can deliver superior digital experiences, modernize their enterprise WAN, and successfully migrate to the cloud.

© 1992–2020 Cisco Systems, Inc. All rights reserved. All product and company names are trademarks™ or registered® trademarks of their respective holders. Use of them does not imply any affiliation with or endorsement by them.