William B. Norton
Console, Chief Scientist

Cloud Interconnections

William B. Norton
Console, Chief Scientist

3131 Jay Street
Santa Clara, CA
wbn@console.to

ABSTRACT

This paper presents a comparison of today's popular cloud interconnection models. For each cloud platform studied (Amazon Web Services, Google Cloud Platform, and Microsoft Azure) we describe the components of their interconnection model using their lingua franca. It turns out that there are a lot of cloud-specific terms that only apply in the context of that cloud offering. For each cloud service we present, we also present a simplified business case for directly connecting to each using a direct (Internet-bypass) connection.

1. INTRODUCTION

All major cloud services offer an "Internet-bypass" solution for directly connecting to their customers, and for good reason. Today's Internet is fraught with security, performance, and reliability issues. Denial-of-Service (DoS) attacks lead to congestion artifacts such as latency, jitter, and packet loss for all traffic traversing the same routers and links used by the attackers. Further, on average there are 4.3 networks[1] in between any two destinations on the Internet. Each of these networks contains potentially many routers and links, any of which can be compromised. Internet traffic can be mirrored, redirected. Even encrypted VPN traffic is subject to off-line decryption. The Internet traffic path presents what the security experts call a "large attack surface."

At the same time, organizations are now dependent on cloud-based applications that require a stable and secure high-performance connection. These applications range from the general cloud-based storage services that team members use to share project files with one another, to revenue-generating ad-network bidding systems where network quality can increase revenue or drag revenue down.

These two forces (reliability of and dependence on the Internet) collide when the business experiences an Internet hiccup that impacts one of their business-critical workflows. To prevent a recurrence, or to proactively increase network reliability, cloud technologists employ Internet-bypass networks to protect and harden the network for these mission-critical applications.

How does an Internet-bypass solution work?

This paper presents the interconnection models used by today's largest cloud services, Amazon Web Services, Microsoft Azure, and Google Cloud Platform.

1.1 A Note About Terminology

The major cloud services have chosen different names and different semantics for each of their cloud services, and the Internet bypass solutions are no exception:

- Amazon Web Services has "Direct Connect,"
- Microsoft Azure encourages all enterprises to connect directly using "ExpressRoute Circuits," and
- Google Cloud Platform interconnects with their customers over a "Google Cloud Interconnect (GCI)."

Each cloud uses their cloud-specific lingua franca when documenting, discussing, and assisting with troubleshooting their services. From a practical perspective, help is often found searching for phrases in user forums, so learning the cloud-specific terminology eases the path towards finding assistance. In this paper we will highlight only the cloud terminology required to understand the models and workflows a cloud service user will experience. We will now explore each cloud service in turn.

2. Amazon Web Services (AWS)

From a market perception perspective, AWS owns the corporate cloud mindshare. According to Gartner, AWS is 14 times larger than its next 10 competitors combined[2]. As the leader in the sector, AWS also pioneered the Internet-bypass solution market for business-critical applications or those with high-performance network requirements. The **AWS Direct Connect** interconnection model was released in 2011[3] in response to these customers' requirements.

2.1 The AWS Direct Connect Model

The AWS interconnect model consists of three parts: the AWS Cloud, the enterprise data center (office or colocation center), and a dedicated network connection in between (see Figure 1).

The customer's AWS resources are contained within a **Virtual Private Cloud (VPC)** and externalized back to the enterprise over an **Amazon Partner Network (APN)**[4]. Once the "**Direct Connect Connection**" is established, the corporate resource owners and users access their cloud resources directly over **Virtual Local Area Networks (VLANs)**.

Beyond the cloud-specific language, each cloud provider also has a collection of downloadable icons to describe workflows utilizing their services. AWS and their users are pretty consistent about using the AWS icons across all presentations and fora. This and the excellent documentation further smooths the path to cloud adoption. Let's follow the path from the AWS cloud back to the corporate data center using the **AWS Simple Icons**[5] to describe the AWS configuration.

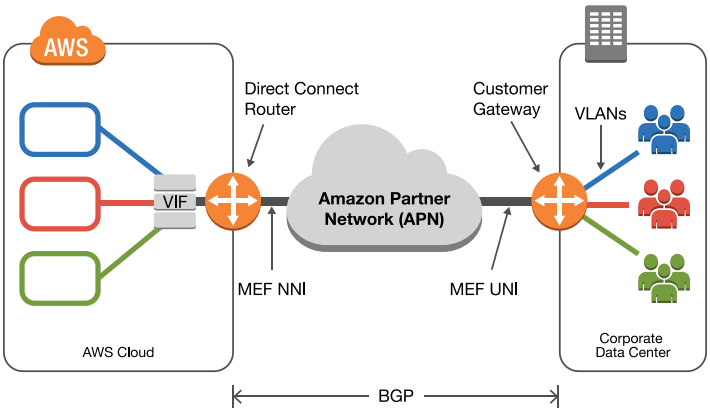

Figure 1 - The AWS Direct Connect interconnection model.

1 Source: RIPE NCC "Update on AS Path Lengths Over Time," https://labs.ripe.net/Members/mirjam/update-on-as-path-lengths-over-time

2 Gartner Report https://aws.amazon.com/resources/gartner-2015-mq-learn-more/

3 https://aws.amazon.com/releasenotes/AWS-Direct-Connect/7982464862957817

4 List of APN Partners: https://aws.amazon.com/partners/

5 AWS Icons: https://aws.amazon.com/architecture/icons/

Comments to the author welcome: wbn@console.net

The rounded rectangles here reflect our abstraction to of the enterprise's resources hosted within AWS, color-coded to match the colors of the enterprise resource owners and users back at the enterprise.

The VPCs contain the enterprise's **"Elastic Cloud Computing (EC2)"** resources, such as **EC2 Instances** (aka "Virtual Machines"), routing tables, storage, security groups, etc.). The VPC contains the enterprise resources that will be externalized back to the enterprise data center.

There are three steps to configure Direct Connect:

1) The enterprise orders a **Direct Connect Connection** from an **APN Partner Network**. For our examples, we will assume Console[6] is the provider, so the port, bandwidth, and region are selected from pull down menus on the Console portal[7]. Once the Direct Connect Connection is provisioned, Console signals the AWS portal that the customer Direct Connect Connection is ready.
2) The user is prompted to add **AWS Virtual Interface(s) (VIFs)** to their direct connect connection. Each VIF can be thought of as an AWS plug, one that is directly attached to the VLAN back at the enterprise data center.
3) Each VPC is provisioned with a **Virtual Gateway (VGW)** connected (routed) to the appropriate VIF. The VIF is configured with ASN, CIDR prefixes, etc. and a downloadable set of router configuration snippets can be downloaded to finish the peering configuration on the enterprise **Customer Gateway**.

After these three steps, the enterprise has in-building dedicated and secure access to their AWS resources, internally tagged as Virtual Local Area Networks (VLANs) routed to the appropriate internal networks.

In Figure 2 we expand our example into a *high-availability diverse-path cloud interconnect model.* This high-availability configuration is sometimes accompanied with a VPN over the Internet as the tertiary failover path.

Enterprises also employ this high-availability configuration across geographically distributed locations.

2.2 Regions and Availability Zones[8]

All AWS resources are physically hosted in geographically distributed **AWS Regions**. Each AWS Region may be spread across one or more non-interdependent data centers, making up separate **AWS Availability Zones**. The region code is articulated by appending zone letters (a,b,c, etc.) to the region name as shown in Table 1.

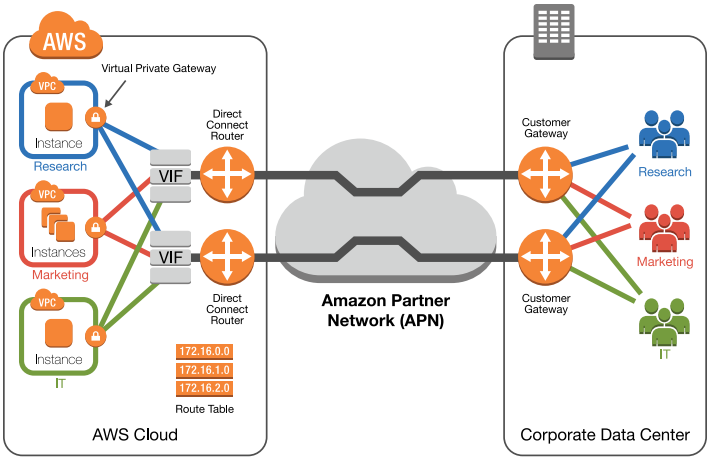

Figure 2 - AWS Direct Connect detailed view.

Table 1 - AWS Regions

Region Name and Location	Region Code (Append Availability Zones)
US East (N. Virginia)	us-east-1(a,b,d,e)
US West (N. California)	us-west-1(a,c)
US West (Oregon)	es-west-2(a,b,c)
EU (Ireland)	eu-west-1(a,b,c)
EU (Frankfurt)	eu-central-1(a,b)
Asia Pacific (Tokyo)	ap-northeast-1(a,c)
Asia Pacific (Seoul)	ap-northeast-2(a,c)
Asia Pacific (Singapore)	ap-southeast-1(a,b)
Asia Pacific (Sydney)	ap-southeast-2(a,b,c)
Asia Pacific (Mumbai)	ap-south-1(a,b)
South America (São Paulo)	sa-east-1(a,c)

When configuring cloud resources, one specifies (or allows to default) the **AWS Region** and **AWS Availability Zones** for their deployment. Next we explore some Direct Connect options.

2.3 Transport – Direct Connect Bandwidth

The Amazon Partner Network (APN) organizations provide connectivity from the customer location to the AWS cloud. AWS can directly accept 1G and 10G connections on their routers, but smaller denominations of interconnect capacity require going through an APN partner as shown in Table 2.

Even though the smallest port size for AWS direct connect is 50Mbps, most partners can deliver any bandwidth desired to connect into these ports. For example, an organization could order a 10Mbps Direct Connect into an AWS 50 Mbps port.

Even with the Direct Connect Connection, the customer still has to pay for the traffic that egresses the AWS cloud. The good news is that the data egress fees are

6 Full disclosure – the writer is employed by Console, Inc.
7 http://console.to
8 http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Concepts.RegionsAndAvailabilityZones.html

substantially lower for Direct Connect Connections than for traffic sent over the public Internet. We will discuss these data transfer fees next as part of an abbreviated "Business Case for Direct Connect[9]."

Table 2 - AWS Direct Connect Bandwidth Denominations

Direct Connect Capacity	AWS Direct	AWS Partner Network (APN)
50 Mbps		●
100 Mbps		●
200 Mbps		●
300 Mbps		●
400 Mbps		●
500 Mbps		●
1 Gbps	●	●
10 Gbps	●	●

2.4 The Business Case for Direct Connect

Most cloud companies charge for traffic on the egress, with all ingress traffic being free. They incent customers to connect directly by discounting the cost of egress traffic sent over the provisioned Direct Connect Connection.

So what is the cost difference between sending traffic over the public Internet versus over an AWS Direct Connect?

2.4.1 Traffic Sent Over The Internet

The traffic that traverses the public Internet is delivered using the **AWS Edge Network**, priced as the **AWS CloudFront**[10] service as shown in Table 3. The pricing for egress traffic is volumetric and in tiers – the more traffic you send, the lower the unit cost. The pricing varies widely by region, with the US and Europe egress fees being almost half the costs of sending the same amount of traffic out of an AWS Asia location.

Table 3 – AWS Internet Data Transfer Fees

	US GB/mo	EU GB/mo	HK+ GB/mo	Japan GB/mo
1st 10TB	$0.085	$0.085	$0.140	$0.140
Next 40TB	$0.080	$0.080	$0.135	$0.135
Next 100TB	$0.060	$0.060	$0.120	$0.120
Next 350TB	$0.040	$0.040	$0.100	$0.100
Next 524TB	$0.030	$0.030	$0.080	$0.080
Next 4PB	$0.025	$0.025	$0.070	$0.070
Over 5PB	$0.020	$0.020	$0.060	$0.060

2.4.2 Traffic Sent Over Direct Connect

When connecting over a Direct Connect Connection, customers pay an hourly[11] port fee (see table 4), a transport fee to the APN partner, and in return they get a lower egress data transfer fee for that traffic.

Table 4 – AWS Direct Connect Port Rental[12]

Direct Connect Port Speed	Port-Hour Rate	Port-Hour Rate Japan
50 Mbps	$0.03	$0.029
100 Mbps	$0.06	$0.057
200 Mbps	$0.12	$0.114
300 Mbps	$0.18	$0.171
400 Mbps	$0.24	$0.228
500 Mbps	$0.30	$0.285
1 Gbps	$0.30	$0.285
10 Gbps	$2.25	$2.142

The egress transfer fee for Direct Connect is about $0.02-$0.03/GB in the U.S. and Europe, $0.045/GB and $0.11/GB in South America. For our estimates we will assume the higher egress data transfer fee of $0.03 per GB per month.

To calculate the cost for the AWS Direct Connect solution, one simply sums the port fees, the APN partner fees, and the volumetric measure applied to the metered data transfer fee. Let's demonstrate this with an example.

2.4.3 AWS Comparison: Internet vs. Direct Connect

Traffic Delivered Over the Internet.

To compare exchanging data over the Internet against the cost of sending that traffic over the AWS Direct Connect, let's make a simplifying assumption that we have a sustained bidirectional 50Mbps of traffic to exchange with AWS.

Let's further assume that the ISP charges $2/Mbps for Internet traffic, so our ISP will accept this 50Mbps of traffic for $100 per month. But we also need to add in the AWS data egress transfer fees.

It turns out that 50Mbps sustained will generate 16,200 GB per month[13]. This traffic spans two pricing tiers (see Table 3), so we add our first 10TB of traffic pricing tier to the second tier pricing.

(10,000GB*$0.085)+(6,200GB*$0.08)

=$1346 per month

9 White paper also available from the author <wbn@console.to>.
10 https://aws.amazon.com/cloudfront/pricing/

11 Note that all Direct Connect providers have a monthly or yearly term. In my opinion there is not much utility in an hourly charge model here.
12 https://aws.amazon.com/directconnect/pricing/ as of the time of this writing
13 Calculation: (50,000,000 bits/sec*60 seconds/minute*60 minutes/hour*24 hours/day *30 days/month) / 8bits/byte

Adding the transit fee to the data transfer fee we see a total cost of **$1446 per month** when sending the data over the Internet.

Total cost for sending traffic sent over the Internet:

$1446 per month

Traffic Delivered Over the Direct Connect.

The cost of sending that same traffic over the AWS Direct Connect service can be calculated by summing the Direct Connect port fees, the APN Partner Network fee, and then applying the lower data transfer fee to our sustained 50Mbps of traffic. We will assume that we will want a 100Mbps port to cleanly handle our 50Mbps of traffic. (This is done to prevent peaks from congesting our circuit.)

Port Fee= $0.06/hour*24 hours/day * 30 days/month

=**$43.20/month**

APN Partner Network fees: 50Mbps = **$500/month**

Data Transfer Fee= 16,200GB*$0.03

=**$486 per month**

Total cost for sending traffic sent over the Direct Connect:

$1029 per month

From this analysis (your mileage will vary of course) we see that all costs of direct connect are completely covered by the cost savings from a lower data transfer fee. It is left as an exercise for the reader to adjust the model with different assumptions.

Table 5 – Summary AWS Internet vs. Direct Connect Costs

	Internet	Direct Connect
50Mbps	$1446 per month	$1029 per month

As stated earlier, enterprises deploy direct connect primarily for greater security, better performance and reliability. Table 5 highlights that the cost of direct connection may be less than, or about the same as, the cost of sending that same data over the Internet.

3. Google Cloud Platform (GCP)

Where Amazon dominates the mind share for corporate customers, **Google Cloud Platform (GCP)** seems particularly well suited to the software development community.

Cloud resources in GCP parlance are stored in a **Project**.

The Google direct connection method is called **Google Cloud Interconnect (GCI)** and it is delivered like an Internet Peering proxy.

3.1 The Google Cloud Interconnect Model

Conceptually, the Google model is the simplest: the enterprise routers "peer" with the Google routers to gain dedicated access to their corporate GCP resources[14] hosted in GCP as well as all on-line Google services (Gmail, maps, etc.). This is a relatively new service, having been launched in late 2014.

To illustrate, in Figure 3 we once again see enterprise departmental resources shown as colored rounded rectangles, owned and used by teams back at an enterprise data center. Notice that there are no VLANs here to segregate networks; everyone gets network access to Google resources or they don't. Users have other mechanisms to control access.

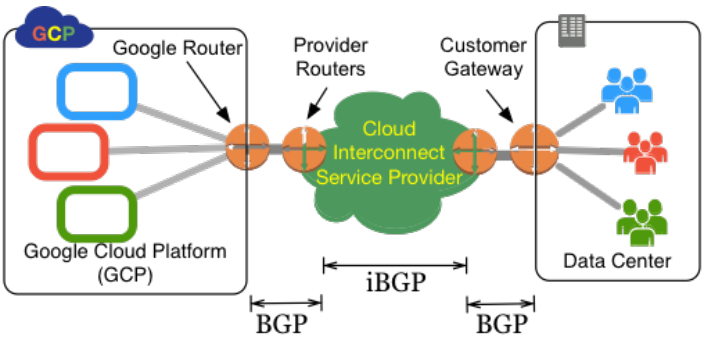

Figure 3 - The Google Cloud Interconnect (GCI) Model

In the GCI model, the customer orders connectivity from a **Cloud Interconnect Provider** and "peers" with the **Provider Router**. The Cloud Interconnect Provider also peers with Google and propagates those Google routes back to the customer, and the customer routes to Google. This interconnection is at layer 3, but over a private dedicated network distinct from the Internet.

Contrast this model with the AWS layer 2 connection which provides dedicated network paths at layer 2, with VLAN tags enabling dedicated path multiplexing and de-multiplexing.

At the core of this GCI interconnection model is the provider's Virtual Router Forwarder (VRF), a network tool used by the interconnection provider. The VRF is conceptually a completely separate routing table operated within the Cloud Interconnect Provider network, but dedicated to the users of that table (Google and the customer in our case). This VRF is not connected to the Internet; it effectively propagates traffic and routing announcements across to BGP speakers in the VRF. After this configuration is set up, Google and the customer are directly connected over an layer 3 Internet-bypass solution.

14 Source: Google Cloud Interconnect: https://cloud.google.com/interconnect/docs

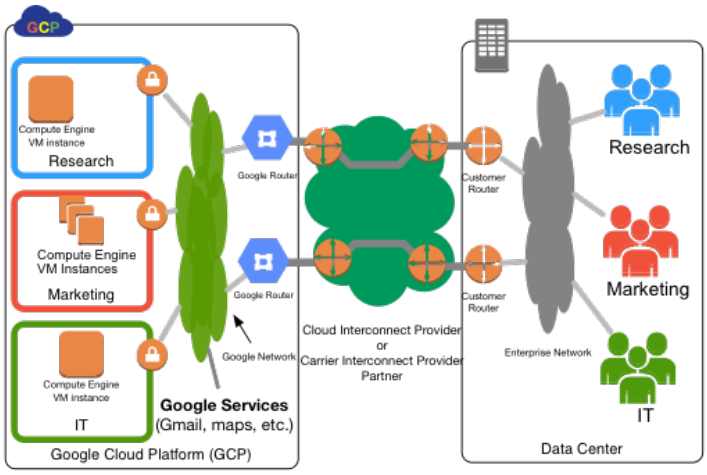

Figure 4 - The Redundant Google Cloud Interconnect Model

3.2 GCP Regions and Zones[15]

Google follows the Amazon model of geographically distributed regions, each with zones of non-interdependent data centers. The names of the regions are articulated by appending zone letters (a,b,c, etc.) to the region name as shown in Table 6.

Table 6 - Google Cloud Platform Regions and Zones[16]

GCI Region and Location	Zone Names
Eastern US The Dalles, Oregon	us-east1-(b,c,d)
Central US Council Bluffs, Iowa	us-central1-(a,b,c,f)
Western US Berkeley County, South Carolina	us-west1-(a,b)
Western Europe St. Ghislain, Belgium	europe-west1-(b,c,d)
Eastern Asia Changhua County, Taiwan	asia-east1-(a,b,c)

3.3 GCP Transport – Google Cloud Interconnect (GCI) Bandwidth

Google provides interconnection at a variety of port speeds.

These connections can be made over point-to-point circuits, across multipoint services, cloud exchanges, and cross connects within a common colocation center.

GCI Capacity	Cloud Interconnect Service Providers
50 Mbps	●
100 Mbps	●
200 Mbps	●
500 Mbps	●
1 Gbps	●
10 Gbps	●

Figure 5 - Google Cloud Interconnect Bandwidth

3.4 GCP Business Case for Cloud Interconnect

Like AWS, Google provides an economic incentive to exchange traffic over a GCI connection instead of over the public Internet. Ingress traffic is free, but all egress traffic incurs a metered data transfer fee.

Let's compare the cost of traffic sent over the Internet versus traffic exchanged over the GCI infrastructure.

3.4.1 Traffic sent over the Internet[17]

Like AWS, the data transfer fee for GCP is split into volumetric tiers (see Table 7).

Table 7 - GCP Internet Egress Data Transfer Fee

	Worldwide (excluding China and Australia) GB/month	China GB/month	Australia GB/month
1st 1TB	$0.12	$0.23	$0.19
Next 10TB	$0.11	$0.22	$0.18
10+TB	$0.08	$0.20	$0.15

Here again we see pricing varying widely across region, with traffic egressing an Australia GCP data centers costing almost double the cost of sending that traffic out of US or European data centers. Once again, in this model, we apply the 16,200 GB to the data egress transfer fee and pay the ISP for Internet transit to determine the cost for traffic exchange.

3.4.2 Traffic sent over Google Cloud Interconnect[18]

Customers pay a lower egress data transfer fee for traffic sent over their GCI connections. For North American GCI traffic for example, the data transfer fee of $0.04 per GB is about one-third the cost of sending that same traffic over the Internet.

15 Source: https://cloud.google.com/compute/docs/regions-zones/regions-zones
16 Source: Google Cloud Platform Regions and Zones: https://cloud.google.com/compute/docs/regions-zones/regions-zones
17 Source: https://cloud.google.com/compute/pricing#network
18 Google Cloud Interconnect Pricing: https://cloud.google.com/interconnect/docs#pricing

Comments to the author welcome: wbn@console.net

Table 8 – GCP GCI Egress Data Transfer Fees

	North America	Europe	APAC
GCI Egress	$0.04	$0.05	$0.06

Let's apply the costs of the GCI interconnection model using the same traffic assumptions as we did for the AWS business case.

3.4.3 Example: Internet vs. Direct Connect

Traffic Delivered Over the Internet.

To compare the cost of exchanging data over the Internet against the cost of sending that traffic over a direct connect service, let's again assume that we have a sustained 50Mbps of traffic to exchange with GCP.

The Internet transit fee paid to the ISP is $2/Mbps, so this 50Mbps of traffic costs $100 per month to send over the Internet. But we also need to add in the GCP data transfer fees (see Table 7).

The 16.2 TB of GCP traffic will span all three pricing tiers as shown in the equation below.

(1,000GB*$0.12)+(10,000*$0.11)+(5200GB*$0.08)

=$1636 per month

Adding the $100 per month transit fee to the data transfer fee we see a total cost of $1736 per month to send this data over the Internet.

Total cost for sending traffic sent over the Internet:

$1736 per month

Traffic Delivered Over the Google Cloud Interconnect.

There are no port fees with the GCI model, so the cost of GCI interconnection is the GCI Service Provider transport plus the GCP data transfer fees.

GCI Service Provider fees: 50Mbps = **$500/month**

Data Transfer Fee= 16,200GB*$0.04=**$648 per month**

Total cost for sending the traffic over the GCI service:

$1148 per month

From this analysis we see that all costs of GCI interconnection are covered by the cost savings from the lower data transfer fees.

Table 9 – Summary GCP Internet vs. GCI Costs

	Internet	GCI
50 Mbps	$1736 per month	$1148 per month

Here again we are pleasantly surprised that the cost of better connectivity is less than the next best alternative, sending that same data over the public Internet. The direct connection method more importantly provides higher security, better performance and better reliability.

4. Microsoft Azure (MAZ)

Where AWS has VPCs as containers, and Google has Projects, Microsoft has **Virtual Networks (VNETs)**[19] as their container object. Microsoft calls their virtual machines **Virtual Machines (VMs)**.

Microsoft strongly encourages all enterprises to connect to Azure over **ExpressRoute**[20] . ExpressRoute provides private network access to three collections of Microsoft network resources: **Azure Private Resources**, **Azure Public Resources**, as well as **Microsoft Software-as-a-Service Resources** such as Office 365 (Skype for Business, Exchange, SharePoint, etc.), and Dynamics CRM Online.

4.1 The Azure ExpressRoute Model

The three classes of Microsoft resources are delivered as an **ExpressRoute Circuit** provided by an **ExpressRoute Connectivity Provider.**

We will use the **Microsoft Azure Icon Set**[21] to show how the ExpressRoute service extends resources to the customer data center and sites.

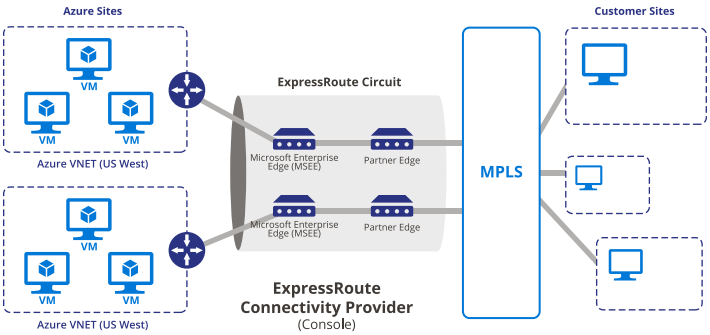

Figure 6 - The Azure ExpressRoute Interconnection Model shown using the Microsoft Azure Icon Set

The ExpressRoute interconnect is different from AWS and GCP in that Azure externalizes three distinct collections of resources back to the enterprise data center. Azure also requires redundant connections for its SLA to be in place.

19 Source: http://cloudacademy.com/blog/public-cloud-war-aws-vs-azure-vs-google/

20 https://azure.microsoft.com/en-us/documentation/articles/express-route-introduction/

21 Microsoft Azure, Cloud and Enterprise Symbol / Icon Set - Visio stencil, PowerPoint, PNG, SVG: https://www.microsoft.com/en-us/download/details.aspx?id=41937

The ExpressRoute Circuit can be conceptualized as a bundle containing both a primary and a secondary path bundle, each of which contains three conduits as shown in Figure 6.

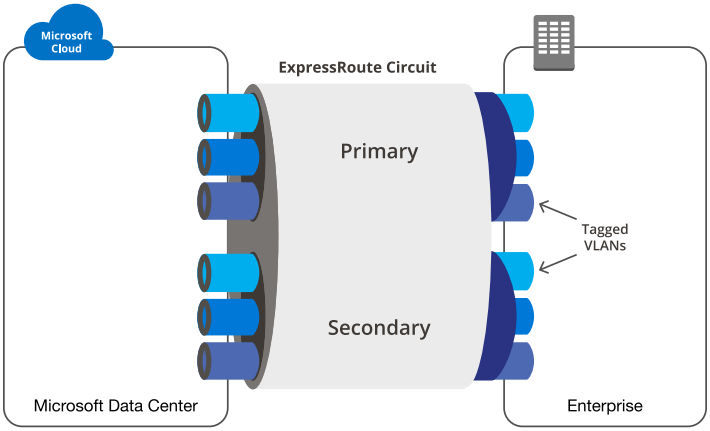

Figure 7 - The "ExpressRoute Circuit"

Let's apply our generalized model and walk through the path from Azure back to the enterprise data center (see Figure 8).

The colored rounded rectangles once again refer to enterprise resources stored within Azure, colored to match the group of owners and users back at the enterprise data center. The difference is that these corporate resources may be deployed across all three categories private, public and Microsoft peering domains.

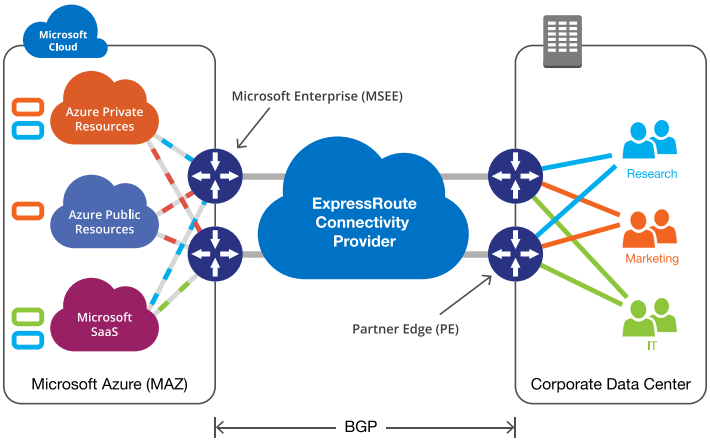

Figure 8 - Azure ExpressRoute Interconnection Model

4.1.1 Three Classes of Resources

The first class of Azure resources are "Private" in the sense that these resources are not reachable over the public Internet. These private resources can be externalized over an ExpressRoute Circuit as a pair of **Azure Private Peering** sessions.

The second class of Azure resources are resources accessible over the public Internet. These public resources are peered over an ExpressRoute Circuit as an **Azure Public Peering** session.

The final class of resources are the Microsoft Software-as-a-Service (SaaS) resources which are available over ExpressRoute Circuit as a **Microsoft Peering** session[22].

These three sets of resources are connected to the Azure side of the **Microsoft Enterprise Edge (MSEE)** routers. The ExpressRoute Connectivity Provider can either connect to the MSEE with a pair of **Network-to-Network-Interfaces (NNIs)** or via one of Microsoft's exchange provider partners. This provides the ability to extend enterprise department-specific resources as VLANs to specific departmental routers back at the enterprise data center.

The other end of the ExpressRoute circuit is attached to the **Partner Edge** router that delivers VLANs to the appropriate departments.

4.2 Azure Regions and Zones

Microsoft operates data centers for their services including Azure, hosted across geographically diverse locations called **Azure Regions**.

Unlike AWS and GCP, in MAZ parlance, a **Zone** is a term only relevant for determining egress data transfer and ExpressRoute port pricing. Further, when you connect to Azure, you are connecting to the Microsoft backbone, not a data center. The implications are that you get access to all regions within a geopolitical boundary (denoted by a zone), something that you would need to pay separately for when connecting to multiple AWS regions for example. As we will see again, egress traffic pricing varies depending on which zone traffic is sent from.

Traffic delivered out of Zone 3 is more expensive than traffic delivered out of Zone 1, for example.

There is also an **ExpressRoute Premium** option that provides global transit across geopolitical regions and loosens some of the Azure configuration limits.

It should also be noted that not all Microsoft services are available in all regions[23].

22 Office 365 and CRM Online require the ExpressRoute Premium service.
23 https://azure.microsoft.com/en-us/regions/#services

Table 10 - Azure Regions and Zones[24]

Azure Region	Regional Data Center Location[25]	Zone
Central-US	Iowa	1
East-US	Virginia	1
East-US-2	Virginia	1
North-Central-US	Illinois	1
South-Central-US	Texas	1
West-US	California	1
West-US-2	West US 2	1
West Central US	West Central US	1
North-Europe	Ireland	1
West-Europe	Netherlands	1
East-Asia	Hong Kong	2
Southeast-Asia	Singapore	2
Japan-East	Tokyo, Saitama	2
Japan-West	Osaka	2
Brazil-South	São Paulo State	3
Australia-East	New South Wales	2
Australia-Southeast	Victoria	2
Central-India	Pune	
South-India	Chennai	
West-India	Mumbai	
China-East	Shanghai	
China-North	Beijing	
Canada-Central	Toronto	1
Canada-East	Quebec City	1

4.3 Azure Transport – ExpressRoute Bandwidth

As with every cloud provider discussed so far, data transfer fees in Azure is charged volumetrically and only in the egress direction.

4.3.1 Traffic sent over the Internet

Traffic sent from Azure over the Internet incurs a data transfer fee, which, as typical, varies widely across zones. Consider for example that egress traffic sent from a U.S. Microsoft data center costs 5-8 cents per GB and egress traffic from Asia costs 16-18 cents per GB[26] as shown in Table 11.

On the other hand, traffic sent over the ExpressRoute service, data transfer fees can drop substantially, from $0.08/GB down to $0.025/GB.

Table 11 - Azure Egress Data Transfer Fees[27]

Outbound Data Transfer	Zone 1 (North America / Western Europe) per GB/month	Zone 2 (Asia) per GB/ month	Zone 3 (Asia2) per GB/ month
First 5GB	FREE	FREE	FREE
5GB-10TB	$0.087	$0.138	$0.181
Next 40TB	$0.083	$0.135	$0.175
Next 100TB	$0.07	$0.13	$0.17
Next 350TB	$0.05	$0.12	$0.16

4.3.2 Traffic sent over ExpressRoute

Azure provides two data pricing plans for ExpressRoute.

First, the **Metered Data** plan involves a monthly ExpressRoute port fee and a metered data transfer rate, based on which zone you are in (see Table 12). The ExpressRoute pricing varies by region connected. To illustrate, a 50 Mbps ExpressRoute dual-port service will cost $55 per month, and egress traffic sent out of the US would be charged $0.025 per GB per month[28].

Table 12 - ExpressRoute Metered and Unlimited Data Pricing (East US Region)

Express Route Port	Price per month (Dual ports) Metered Data Plan	Zone1 per GB	Zone 2 per GB	Zone 3 per GB	Price per month (Dual ports) Unlimited Data Plan
50 Mbps	$55	$0.025	$0.05	$0.14	$300
100 Mbps	$100	$0.025	$0.05	$0.14	$575
200 Mbps	$145	$0.025	$0.05	$0.14	$1150
500 Mbps	$290	$0.025	$0.05	$0.14	$2750
1 Gbps	$436	$0.025	$0.05	$0.14	$5700
2 Gbps	$872	$0.025	$0.05	$0.14	$11,400
5 Gbps	$2180	$0.025	$0.05	$0.14	$25,650
10 Gbps	$5000	$0.025	$0.05	$0.14	$51,300

The other pricing plan is the **Unlimited Data** plan, also shown in Table 12. An enterprise that orders the 50Mbps Unlimited Data plan would pay $300 per month and be able to send up to 50Mbps of egress traffic.

24 Source: https://azure.microsoft.com/en-us/regions/
25 Since one connects ExpressRoute to Microsoft network instead of a Microsoft data center, the specific data center location is less important in the Azure model.
26 https://azure.microsoft.com/en-us/pricing/details/data-transfers/
27 Source: https://azure.microsoft.com/en-us/pricing/details/data-transfers/
28 Microsoft says enterprise discounts may be applied here.

4.3.3 Example: Internet vs. ExpressRoute

Traffic Delivered Over the Internet. Using our same 50Mbps enterprise traffic assumptions, we see that sending 16,200GB will span the first three tiers of the egress traffic pricing:

Data Transfer Fee

= 5GB*$0.00+10,000GB*$0.087+6195*$0.083

=**$1384 per month**

When we add in the $100 (50Mbps @ $2/Mbps) Internet transit fees, we see that

the total cost for sending Azure traffic over the Internet:

$1484 per month

Traffic Delivered Over ExpressRoute – Metered Data Plan.

We will compare both the metered and unlimited plans and compare them against the Internet cost of egress traffic.

The dual 100Mbps ports will cost $100 per month. Why 100Mbps ports? We have an offered load of 50Mbps and we want to ensure we don't congest the pipe and cause packet loss. We will assume the Console price point will still be $500 per month for a redundant 50Mbps service delivered on those redundant 100Mbps ports. We will assume the traffic is delivered from North America (Zone 1) at $0.025 per GB per month.

Data Transfer Fee= 16,200GB*$0.025

=$405 per month

Metered Data Plan=$100+$500+$405

=$1005 per month

Traffic Delivered Over ExpressRoute – Unlimited Data Plan.

The Unlimited Data plan port will cost $575 per month for dual 100Mbps ports and includes all of our 50Mbps egress traffic. We will assume the transport provider will charge $500 per month for a redundant 50Mbps service delivered on a 100Mbps port.

Unlimited Data Plan = $575+$500

=$1075 per month

Under these assumptions, ExpressRoute metered may be a less expensive option than the Unlimited Data plan, and less expensive than sending that data over the Internet. The important part though is that we are delivering provably better security, performance, and reliability while not costing twice as much.

Total cost for sending the traffic over ExpressRoute:

$1005 or $1075 per month

Table 13 - Azure Internet vs. ExpressRoute

	Internet	ExpressRoute Metered	ExpressRoute Unlimited
50 Mbps	$1636/month	$1005/month	$1075/month

The Unlimited Data plan is a little more expensive under our assumptions, mostly because we are paying for unlimited at the 100Mbps level with only a 50Mbps load.

5. Conclusion

In ancient Egypt, a Rosetta Stone provided translation of text into three different scripts (hieroglyphic, demotic, and Greek) so priests, government officials, and rulers of Egypt could read what it said.

Figure 9 - Rosetta Stone

We have seen that all three major cloud services have different interconnection models and naming (see Table 14).

Amazon Web Services (AWS) offers its "Direct Connect" method for those that require high-availability or high-performance access to AWS resources. Google Cloud Platform (GCP) offers Google Cloud Interconnect (GCI) as the direct connection method. This interconnection model resembles traditional "Internet Peering." The Azure name for their direct connection method is "ExpressRoute," which utilizes a redundant collection of three peering sessions ("Azure Private Peering", "Azure Public Peering", and "Microsoft Peering") to connect to Azure Private resources, Azure Public resources, and Microsoft Software-as-a-Service (SaaS) offerings.

Comments to the author welcome: wbn@console.net

Table 14 - Cloud Rosetta Stone

Service	AWS	GCP	MAZ
Container	Virtual Private Cloud (VPC)	Project	Virtual Network (VNet)
Direct Connection Service	Direct Connect	Google Cloud Interconnect	ExpressRoute
Transport Providers	AWS Partner Network (APN)	Cloud Interconnect Provider	ExpressRoute Connectivity Provider
IaaS CPU	Elastic Cloud Computing (EC2) instances	Compute Engine VMs or instances	Virtual Machines (VM)

Most clouds charge on egress traffic volume, with directly exchanged traffic costing a fraction of the cost of traffic sent over the Internet. Price points vary, but the math in this white paper provides a starting point for calculating the cost of enhancing connectivity to cloud services.

6. Acknowledgements

Thanks to Blake Gillman, Morgan Snyder, Jay Turner, Prentice, Ted Linnenkamp, Rob Parker, Larry Lizotte, Joey Cappelletti, Tom Madej , Al Burgio, John Hill, Jane Yoon, Brian Sutterfield, Stephen Wilcox, Ruth Plater, Chandrasekar Kannan, Kailash Khailasi, and others who wish to remain anonymous like Larry Binetti.

7. About the Author

William B. Norton currently serves as the Chief Scientist for Console, a leading advanced interconnection provider. He is the only published author on the practice of interconnection, The Internet Peering Playbook: Connecting to the Core of the Internet. From 1998 to 2008, Mr. Norton served as Co-Founder and Chief Technical Liaison for Equinix. He has authored 20 white papers, spoken at over 200 conferences, toured over 100 data centers, and created and delivered peering workshops across Africa, Europe, and North America. From 1987 to 1998 he held several positions at Merit Network, including the Internet Engineering Group manager. He also created the business model for NANOG, and served as its first chairman.