ESnet

ENERGY SCIENCES NETWORK

AWS Pilot Report

M. O'Connor, Y. Hines

July 2016

Version 1.3

Ernest Orlando Lawrence Berkeley National Laboratory
1 Cyclotron Road, Berkeley, CA 94720-8148

This work was supported by the Director, Office of Science, Office of Advanced Scientific Computing Research of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Contents

Executive Summary

The ESnet Amazon Web Services (AWS) pilot objective is to determine specifically if the Amazon Web Services (AWS) “Direct Connect” or “DX” service provides advantages to ESnet customers above and beyond that of our standard Amazon connections at public Internet exchange points. DX services are supported on a direct physical link between AWS and the customer or their provider. Through the course of our investigation of DX, we reached the conclusion that Virtual Private Cloud (VPC) is a key technology that generally satisfies the network requirements of the Research & Education networking support model. VPC is essential to controlling the routed paths that support data movement within a distributed collaboration. DX is one of two available approaches to implementing VPC and for this reason DX service remains interesting. However, due to significant connection costs, lower available connection speeds and billing complexities, we recommend that as their first choice, our customers should consider the alternative IP Security (IPSec) tunneling approach for provisioning VPC with AWS instead of a DX hardware based solution.

Introduction

Through ESnet customer contact during this pilot as well as at ESnet Site Coordinator Conferences (ESCC) meetings it has become clear that commercial cloud computing services are in wide use today and have become an integral element in IT planning for the future. The ESnet AWS pilot objective is to determine specifically if the Amazon Web Services (AWS) “Direct Connect” or “DX” service provides advantages to ESnet customers above and beyond that of our conventional Amazon connections at public Internet exchange points.

Cloud computing services can be broken down along two principal axis, the first being Software as a Service (SaaS) vs. Infrastructure as a Service (IaaS) and the second, Public Cloud services vs Private Cloud services.

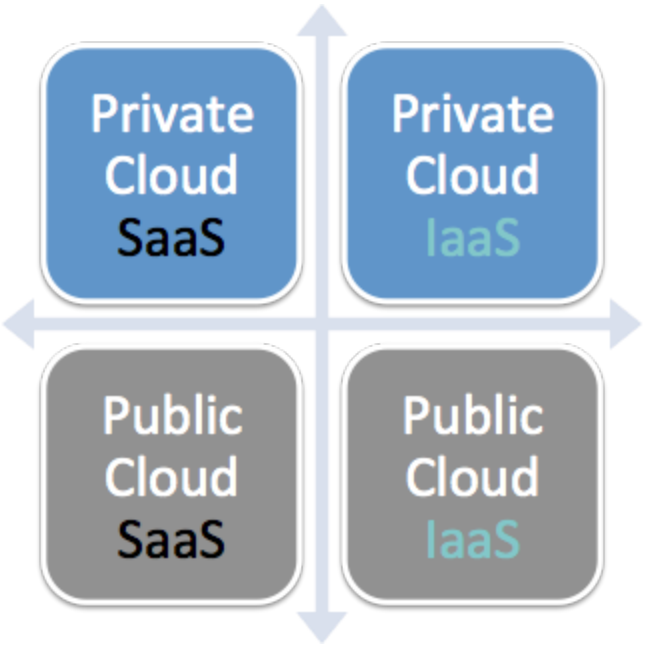


Fig. 1.0

Amazon Web Services offerings are commonly viewed as falling into the Infrastructure as a Service (IaaS) category and the scope of this pilot is specifically within AWS and their respective IaaS offerings.
The ability to perform work where and when resources are available has become a hallmark of globally distributed scientific collaboration. The Research and Education community has established an extensive history sharing computing infrastructure, from the early days of time sharing supercomputers through more recent sharing of locally managed resources within a collaborative grid computing framework. With great success sharing local resources already established, the science community is particularly well suited to incorporate commercial cloud IaaS into their collaborative computing workflows.

ESnet initiated a pilot with AWS to explore areas where our organization could facilitate science in the use of IaaS.

- Describing the practical aspects of supporting the physical layer networking required by DX
- Shedding light on the reality of global inter-region data transport and the role of high-performance R&E networks in this environment
- Contrasting how IaaS is used today and demonstrating why current practice will not scale to meet globally distributed collaborative computing requirements

This document describes the pilot, its results, recommendations and the increasingly critical role R&E networks play in linking heterogeneous and geographically distributed commercial data-centers in support of science.

1 The Pilot

The decision to investigate AWS was based on Amazon's position as a leader in the emerging IaaS sector as well as early interest within our DOE Lab community in 2015. With Collaborative IaaS as our primary focus of investigation, participation in the ESnet AWS pilot was limited to Labs willing to test DirectConnect (DX) and Virtual Private Cloud (VPC) services. The three participating Labs, BNL, FNAL and PNNL all had groups within their respective enterprises already using AWS in some capacity and were willing to execute the required network provisioning to connect their enterprise LANs to AWS over DX.

ESnet AWS fees during the ESnet pilot were covered by Amazon through AWS in Education Grants. Each of the three DOE Lab participants also received AWS in Education Grants.

1.1 Direct Connect (DX) Testing Architecture

The Direct Connect (DX) service requires a direct fiber optic connection between AWS and the network service provider serving the customer. For the ESnet Pilot, we chose to connect to AWS in Seattle at the Pacific Northwest Gigapop facility.
ESnet purchased two fiber cross connects between our existing router and the AWS POP. Each fiber cross connect supported 10Gbps of bandwidth between ESnet and AWS. Both cross connects were capable of carrying multiple customer connections, with bandwidth usage requirements determining resource sharing levels between customers. Each customer DX connection is composed of two VLANs, one public (AWS) and the other private (VPC) as depicted in the following diagram.

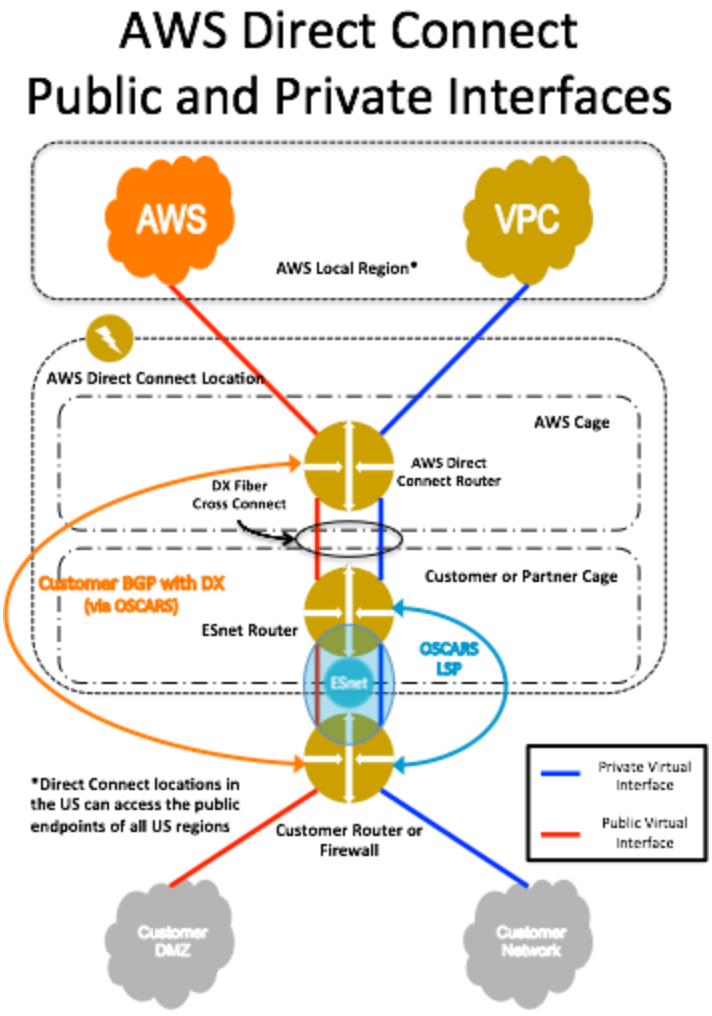


Fig. 3.0

2 Observations about IaaS

Enterprise IaaS is the prevailing architecture in use at National Laboratories today. Enterprise IaaS describes an approach where an organization purchases cloud infrastructure services that they consume internally, within the enterprise. While cloud services implemented in this way are likely fulfilling current requirements for many ESnet customer sites, it is important to understand where performance and scalability issues may arise in the future when consumption of cloud resources expand beyond Enterprise IaaS (one user to the cloud) and move into a Collaborative IaaS (many users to the cloud) paradigm.

2.1 Enterprise IaaS

Using the Enterprise IaaS model, a compute center maintains direct ESnet customer support between their LAN edge and the cloud. Once inside the cloud, the IaaS provider can be relied on for customer support. The Enterprise IaaS approach should not expose the customer to "support gaps" or intractable performance issues on the network connecting to the cloud since ESnet connects directly to major cloud providers in the US. As ingress and egress paths to the cloud remain on ESnet, Enterprise IaaS implementations do not benefit in terms of performance by using Direct Connect (DX) or Virtual Private Cloud (VPC). There are inherent security advantages to using VPC, but in terms of performance the ESnet path will be comparable for Enterprise IaaS over DX or the general ESnet routed path.

Benefits of Enterprise IaaS

- Simple use of AWS address ranges, does not require network level configuration, ie: a new AWS BGP peering
- High Performance ESnet customer transit directly to a cloud provider's general Internet peering
- ESnet technical assistance, performance metrics & maintenance Notifications

Limitations of Enterprise IaaS

- Does not support high-performance R&E data transfer to a global collaboration
- Site perimeter security integration is not possible at the network layer
- Lack of control over the AWS address ranges and routing to third party collaborators

2.2 Collaborative Infrastructure as a Service (IaaS)

In contrast to Enterprise IaaS, Collaborative IaaS implies that data and compute could be shared beyond the borders of the enterprise as grid computing resources are shared today. Since the collaborating compute centers can not be assumed to be ESnet connected sites, the path from AWS to these collaborators are likely to egress the cloud onto the general Internet, following a routing plan defined by Amazon's upstream transit providers. Once data egresses the cloud onto the general Internet, science flows will be exposed to a series of networks designed for a huge numbers of small transfers without a clear chain of support for when things go wrong. The ESnet AWS pilot findings illustrate how the use of VPC over DX ensure that science flows remain on R&E networks and can even transit between heterogeneous cloud service providers in support of Collaborative IaaS.

Key assumption:
"When geographically dispersed collaborations begin to transit large datasets directly out of the cloud, VPC services will be essential in order to control routing between those collaborating institutions as well as any integrated cloud providers that they may choose."

ESnet has already observed a trend toward distributed use of cloud services within the LHC collaborations. We predict that over time, as our customers become more accomplished in using cloud services, large multi-institution collaborations will begin sharing cloud resources as they now share local resources and so, be forced to move beyond simple Enterprise IaaS toward a Collaborative IaaS model.

Assumptions:

- Global collaborations will begin to share data directly from "The Cloud".
- Science flows will traverse the general Internet unless steps are taken to ingress and egress onto R&E networks instead of cloud provider transit networks.
- Scientific (elephant) flows are likely to impact performance of general purpose (mouse) flows and will require special traffic engineering steps be taken to perform properly.
- The public internet is highly fragmented and not engineered to support the kind of Scientific flows required by the globally distributed computing model.
- In particular the LHC globally distributed computing model requirements can not be met deterministically by the general Internet.

2.3 Key Points When Leveraging ESnet Support for AWS Services

- ESnet has established high capacity BGP peering with AWS at multiple locations in the US. ESnet does not peer with AWS outside of the US.
- Within the context of the Enterprise IaaS, ESnet directly connected AWS links provide a deterministic path from AWS through ESnet to the customer site in the US.
- AWS BGP route advertisements provide reachability only to the availability zones in the local geographical region where the BGP peering is established.
- If the Enterprise IaaS model is modified introducing a third party collaborator, the non-homogeneous, regionally compartmentalized AWS BGP routing implementation will leave many collaborators with only commercial commodity internet paths to most AWS regions, this being particularly acute between the US and Europe.
- R&E networks only provide commercial network (ie:AWS) transit to customers, not to each other (peers). We do not suggest that changing this would have a great benefit.
- The fact that ESnet peers with AWS in the US does not help GEANT customers reach AWS regions in the US. European collaborators will use one or more commercial internet transit provider networks to reach AWS regions in the US, a different set of providers in each direction is common.

3 The AWS Cloud is not built to be a transit network

The general Internet is often described as a "network of networks" with the vast majority assigned their own unique autonomous system number (ASN) and routing consistently using a single strategy that ensures reachability and prevents loops. The Amazon AWS network is different in that while it is a single ASN, from an external perspective it is not contiguous and employs a "regional" rather than uniform single routing strategy.

3.1 AWS network ignores a number of longstanding Internet Architecture rules

- AWS ASN (16509) is not contiguous, as Autonomous Systems are intended to be. This "Cloud" is a set of regional Data Centers.
- BGP with AWS establishes connectivity with only the geographically close regions, **NOT** the entire AS. Establishing BGP with AWS, ASN 16509 in one region does not establish connectivity to other regions, continents etc.
- Surprisingly, routing to remote AWS regions will NOT use your established BGP with AWS.
- Similarly, the reverse path from remote regions are configured to prefer commodity public Internet over R&E paths.

3.2 Amazon is not in the long-haul network transport business

- Amazon Web Services (AWS) offers an extensive portfolio of computing resources, but they are not in the "networking" business.
- AWS recharges customers for egress traffic out of the cloud. Some researchers have described these fees as "holding their data hostage".
- AWS upstream transit to the Internet is provided through traditional upstream transit providers and so they pass these costs on to their customers.
- Their inter-region long haul circuits are used for internal control and management rather than customer transport.
- AWS offers a service that will migrate data between regions, but this is strictly controlled and scheduled by AWS, not customers.
- Since AWS peers for free with R&E networks, they offer customers of these networks an "Egress Traffic Fee Waiver".

In summary, AWS peerings on one AWS region do not offer routes beyond the geographically close neighboring regions-this is a challenge for collaborations designed around a globally distributed computing model. However, ESnet and R&E networks in general remain in a strong position to continue supporting the research community as a reliable, deterministic and responsive high performance alternative to the general Internet. In this role, Research & Education Networks have an opportunity to scale cloud services beyond the local region for their customers.

4 Controlling WAN routing with Virtual Private Cloud (VPC)

VPC is a type of cloud service that encapsulates compute instances in a rich networking layer, containing features not available using public AWS services. In many ways like a physical LAN, VPC enables a network perimeter to be constructed with defined ingress and egress routing.

Two pilot participants (BNL, FNAL) successfully provisioned VPC BGP peerings using the AWS self-service portal. This configuration enabled the strict routing control of their own IP prefix/range that they mapped into the AWS cloud, providing an "on-ramp" to high performance R&E networks like ESnet.

As a test of "Private Cloud IaaS" we mapped an ESnet CIDR address block (192.188.23.0/24) into a Virtual Private Cloud in the US-West-1 region using Direct Connect (DX) VPC services. We then instantiated a LINUX compute instance using address 192.188.23.110.

From the perspective of a European NREN, GARR in this case, the BGP route from a European NREN to an AWS compute instance in the US-West-1 region routes using only R&E network paths just as it would if the host was located at an ESnet customer site. VPC provides the ability to distribute data from a cloud instance to a global collaboration using the same R&E networks that have been built specifically for this purpose, offering an elegant solution to cloud based data distribution.

Example using the GARR (Italian NREN) looking Glass:
show route 192.188.23.110
inet.0: 554383 destinations, 1143162 routes (554352 active, 27 holddown, 11 hidden)
*+ = Active Route, - = Last Active, * = Both*
*192.188.23.0/24 *[BGP/170] 11w2d 02:52:32, localpref 300, from 90.147.84.9*
AS path: 20965 293 292 I

R&E AS PATH: GEANT(20965) ESnet(293) ESnet-west(292)

This architecture provides a comprehensive solution, enabling cloud instances to route as if they were physically located at the customer site. The resulting implementation scales very well, requiring no additional network configuration or changes by any of the collaborating networks or compute centers. This is in stark contrast to complex end to end schemes that attempt to use cloud provider Internet address space instead.

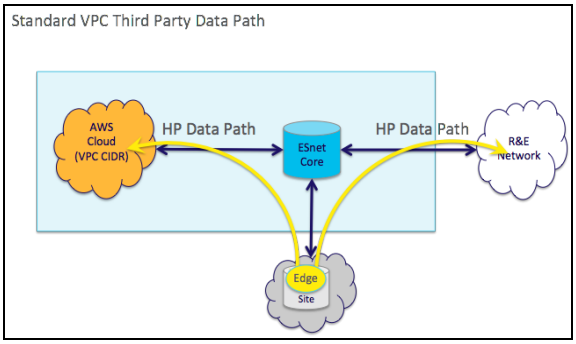


Fig. 4.1 VPC Data Distribution

Figure 4.1 illustrates a direct BGP peering from the site edge to the AWS cloud over an IPsec tunnel. In VPC implementations there is a single path in and out of the AWS cloud and that path is through the BGP peering connecting to the customer site.

If VPC had not been used in this case, R&E paths could not be used to transport the data. This is because “Public Cloud IaaS” uses AWS assigned IPv4 addressing and traffic will egress on any number of Amazon commercial transit provider peerings to the general internet.

Since ESnet and GEANT do not exchange routes to non-customers ie:AWS, It would not matter that GEANT peers with ESnet over 100G interconnects, or that both networks have contracted multiple 100G circuits across the Atlantic, these paths simply could not be used. Instead, the GEANT paid transit path would become the best path from Europe to the US and ESnet paid transit would be chosen for the return path. In fact the number of intermediate commercial networks handling the data depends on the source/destination compute centers and cloud providers. Without VPC, the fact that ESnet peers with AWS in the US does not help GEANT customers at all.

VPC is the principle technology that provides routing control and DX is simply one approach to supporting it. The principle benefit of DX from our perspective is to enable VPC over a dedicated physical connection from an ESnet site directly into the AWS network.

4.1 Why exchanging cloud provider routes between NRENs is not adequate

The seemingly simple solution of exchanging cloud service provider route prefixes will not satisfy the requirement to route between cloud providers. We cannot assume that all collaborators will contract the same cloud provider, so at some point, cloud instances will need to move data between each other's public address space and these providers will not use ESnet or any other R&E network to connect in this way.

4.2 Pros and Cons of Direct Connect

Pro

- Data sharing directly from a Virtual Private Cloud over R&E networks
    - Deterministic
    - Egress fee control
- Shortest BGP path over direct AWS to site BGP peering
- Dedicated physical path into the AWS network
    - Doesn't compete with Amazon retail or any other public streams

Con

- Costly, additional DX cost up to $23K per 10GE annually when compared to public peerings
- 10G network infrastructure vs 100G Amazon public peerings
- IPv4 address availability limits VPC address mapping scalability
- S3 storage can not be mapped into a VPC today
- Requires scripting of BGP filter policies based on region.

The tremendous IPv6 address range would allow VPC address mapping to scale more effectively. IPv6 is not supported by AWS.

5 AWS Billing Overview

In support of DX as a potential shared service among ESnet customer sites, we needed to confirm that the AWS billing process was able to direct all DX specific charges to ESnet while attributing compute related fees to our customers directly. AWS was able to successfully separate these charges as ESnet required. This was confirmed on the cross connects at the Pacific Northwest GigaPOP. Note however that BNL ATLAS paid for their own dedicated DX service in the AWS Eastern region and those charges are present on their billing statement.

AWS Service Charges are broken down into the following categories

- Data Transfers
- Direct Connect (DX)
- Elastic Compute Cloud (EC2)
- Simple Storage Service (S3)
- Support (Business)
- CT Consumption Tax (Japan)
- GST - Goods and Service Tax (Canada and Australia)
- US Sales Tax to be collected
- VAT - Value Added Tax (European Union, Swiss, Liechtenstein, Norwegian, Icelandic customers)

5.1 Amazon Grant awards

ESnet received two *AWS in Education Grant* awards $20,000 on 2/23/2015 and $12,000 on 10/14/2015. These two grants covered all AWS charges during the pilot with the exception of $3,030.00 in cross connect fees to Pacific Northwest GigaPOP for the dedicated 2x10GE ports to AWS Direct Connect (DX) service.

5.2 ESnet AWS Spend Rate

AWS Pilot ESnet's average monthly spend rate in the following areas were:

- AWS Support (Business) $334.34
- AWS Direct Connect $3,285.29
- AWS Amazon Elastic Compute Cloud $57.96

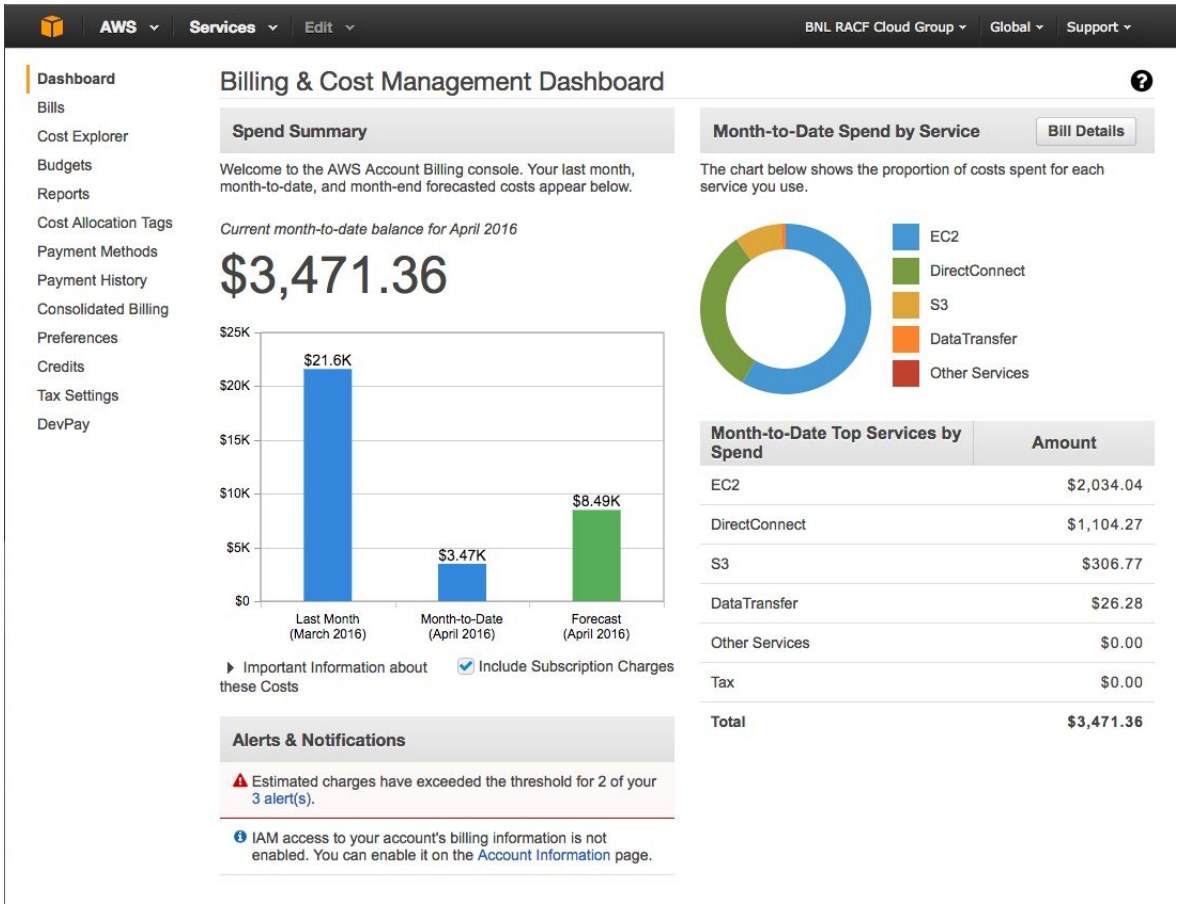
** For a detailed example of the monthly ESnet AWS service charges see Appendix A.*

5.3 BNL ATLAS AWS Billing Overview

In March 2016 BNL's BNL RACF Cloud Group ran a 45,000 Core Test at a cost of $21,598.26.

- The Core Test used AWS resources from the following three regions
  - US East (Northern Virginia) Region
  - US West (Northern California) Region
  - US West (Oregon) Region

** For a detailed example of the monthly BNL ATLAS AWS service charges see Appendix A.*



BNL ATLAS Billing & Cost Management Dashboard April 2016
45,000 Core testing in March, 2016

6 Conclusions

The AWS Pilot provided a perspective into a future state of cloud computing achieved when cloud services are ultimately incorporated into the prevailing large scale scientific Globally Distributed Computing Model. This report suggests that the required wide area network components of this model are not being adequately supported by commercial cloud providers today and that this state of affairs is likely to continue into the foreseeable future. This observation is in many ways similar to lack of support for large scale science in the general Internet today and that Research and Education networks will remain in a unique position to satisfy the needs of large-scale science in a global heterogeneous cloud based computing environment.

Virtual private cloud services can be provisioned with tunnels over the public peering infrastructure, without the additional cost of **Direct Connect** (DX). ESnet does not plan to purchase DX service directly, but we will support DX service purchased by our customers when required. In addition we view the current implementation of DX on separate lower bandwidth infrastructure as a limiting factor that will inevitably impede flexibility and growth.

6.1 Detailed Conclusions

- Virtual Private Cloud services are essential as an "on-ramp" connecting the cloud to R&E transport between global collaborations.
- VPC services enable transport between non-homogeneous cloud service providers.
- Using tunneling techniques, VPC can be provisioned over public AWS peerings.
- The AWS network AS(16509) is not uniformly reachable from a single BGP peering location as networks generally are configured to be.
- Peering with AWS in Europe does not provide reachability to AWS regions in the US.
- Enterprise IaaS to the nearest region is successful today for many ESnet customers, but this is not a particularly compelling or unique service model requiring services specific to ESnet.
- Research & Education networks will continue to play an essential role in supporting the "Globally Distributed Compute Model" as it embraces Collaborative IaaS.
- The HEP community is working to expand support with standard tools for cloud deployments, ie: CONDOR, OpenStack.
- Egress transit fees can be avoided by using VPC in conjunction with Collaborative IaaS.
- ESnet must upgrade public Internet connections to bring them in line with R&E interconnects in order to support science traffic to and from the cloud.

- ESnet is not often colocated within the same type of facility as Amazon. ESnet may have to expand our presence to peer at higher levels with AWS. (R&E vs commercial COLO)
- Track customer cloud service choices in an effort to provide support and connectivity when and where required.
- The "AWS Egress Fee Waiver" is not provided in writing and may be terminated without notice.
- AWS billing does separate Network DX charges from Compute, this allows ESnet to cover DX charges for customers if desired.

7 Recommendations

- Aggressively upgrade ESnet interconnects to the public Internet, bringing performance in line with R&E network paths.
- Investigate the performance characteristics of various tunneling options available for supporting VPC over AWS.
- Investigate AWS Virtual Network Infrastructure to understand if these service offerings might be used in satisfying Esnet customer requirements.
- Determine which cloud service providers beyond AWS are in use and of interest to the ESnet customer base, then verify that connection levels to these providers are resourced appropriately.
- Evaluate the Collaborative Cloud Service (CCS) Proposal below.

7.1 Proposal: ESnet Collaborative Cloud Service (CCS) Architecture

The Collaborative Cloud Service CCS could offer virtual router instances to our customer sites that would effectively place a portion of their LAN at the edge of one or more cloud regions. This service is an optimization of the VPC approach to Collaborative IaaS and improves performance by shortening routed paths, reducing latency and steering all egress traffic out of the cloud and onto ESnet avoiding extra hops through their campus perimeter.

Benefits:

- The path latency between collaborators and the cloud data movers is reduced, increasing transfer rates.
- ESnet can save on expensive customer local loop upgrades by implementing higher capacity cloud links.
- Sites may save money on expensive edge routing equipment upgrades.

- Lower bandwidth customer locations could collaborate quickly at a high performance level, directly from the cloud.

SDN site router - A virtual router that is provisioned to BGP peer with the cloud provider using the customer ASN.

- Multiple Virtual Routers per (VR) hardware device, one VR per customer.
- IPsec or similar tunneling protocol will connect the SDN Site router to the cloud in order to support VPC over public AWS peerings
- Collaborator transit paths will traverse the best ESnet path to the cloud
- The CCS proposal is compatible with, but does not require DX

The potential exists to:

- Become an important application for SDN and a new ESnet service
- Transform the ESnet site connection architecture by alleviating upgrade pressure on local loop circuits

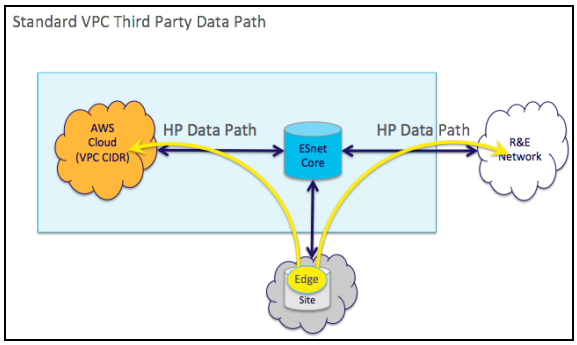


Fig 7.1 VPC Data Distribution without CCS

Fig 7.1 illustrates controlling the ingress and egress routing to a VPC using a direct BGP peering from the site ege to the AWS cloud over an IPsec tunnel.

Pro:

- Routing is controlled by the site not Amazon
- The VPC can be incorporated into a perimeter security architecture
- Third party collaborators can reach this VPC using High Performance R&E networks

Con:

- Data will traverse the site local loop twice in each direction between collaborator and cloud.

Fig. 7.2 & 7.3 both illustrate the insertion of an ESnet provided SDN Router in order to serve multiple Virtual Router instances for customer sites requiring data transport directly from a cloud provider edge.

The data distribution path to a remote collaborator is direct in Fig. 7.2, providing a more efficient routed path when compared with Fig 7.1.

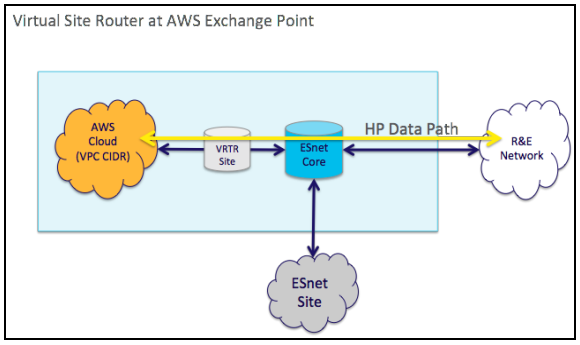


Fig 7.2 CCS Routing Directly From the Cloud using CCS

Fig. 7.3 indicates detailed protocol information for each link in a CCS installation.

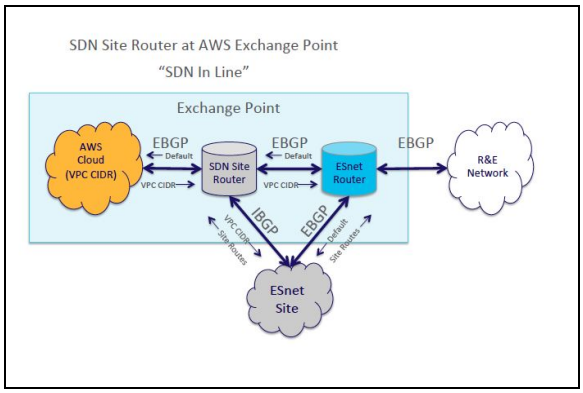


Fig 7.3 CCS Protocol Description

7.2 CCS Future Work

1. Virtual router service architecture, SDN, whitebox etc.
2. Investigate IPsec and other kinds of tunnel performance between ESnet and AWS over public peerings in support of VPC.
3. Investigate various virtual Network Device offerings from AWS, such as routers, switches and firewalls.

Appendix A

1. **ESnet AWS Service Charges 08/2015**

2. **BNL ATLAS AWS Service Charges 03/2016**