AWS Pilot Report

Connecting Clouds to NRENS

Michael O'Connor moc@es.net

ESnet

Network Engineering

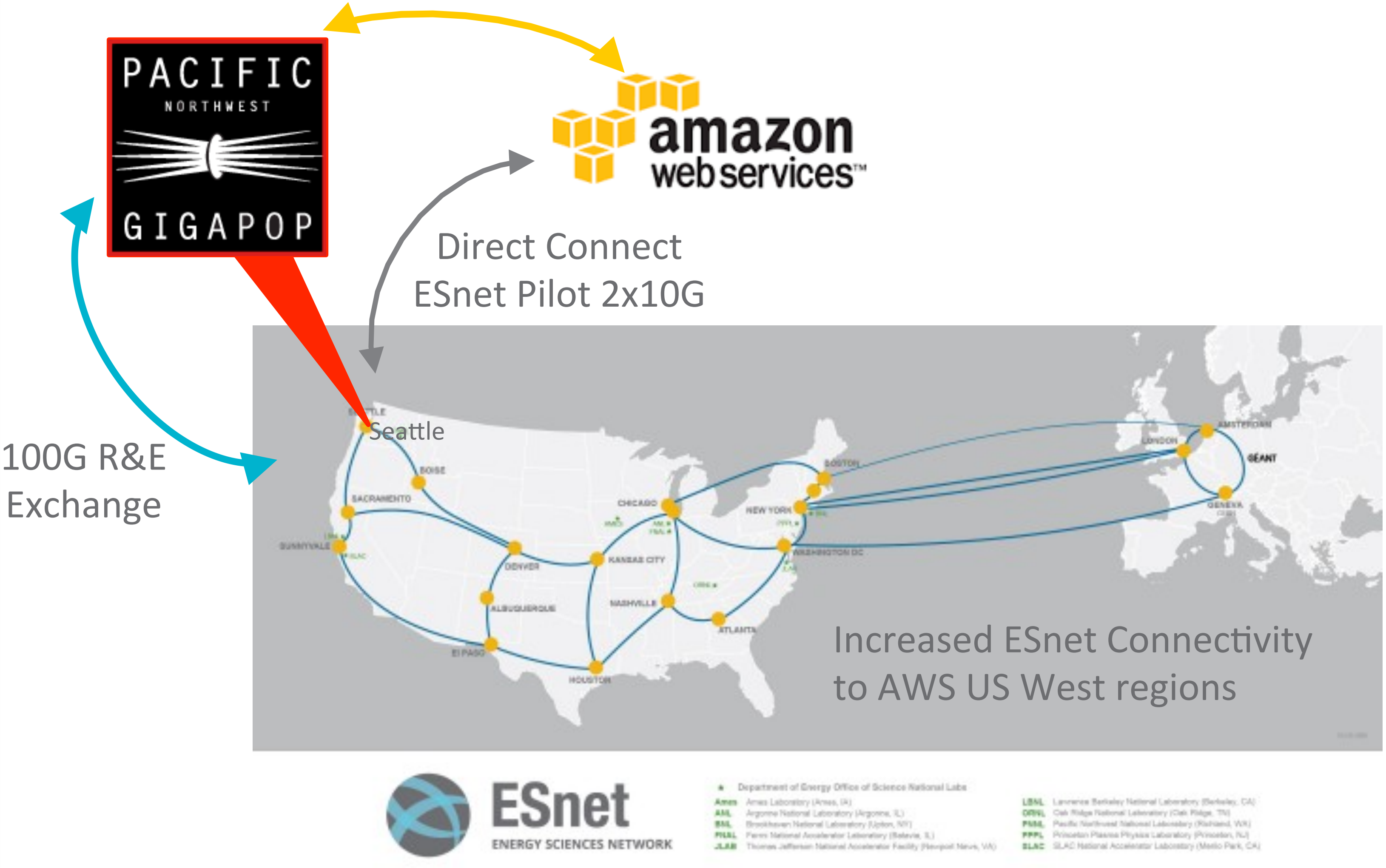
LHCOPN-LHCONE meeting

Helsinki (FI)

Sep. 19, 2016

ESnet AWS Pilot

AWS
100G to PNWG


PACIFIC NORTHWEST GIGAPOP
amazon web services™
Seattle
SEATTLE
BOISE
SACRAMENTO
SUNNYVALE
LBNL
SLAC
DENVER
ALBUQUERQUE
EL PASO
HOUSTON
KANSAS CITY
CHICAGO
AMES
ANL
FNAL
NASHVILLE
ORNL
ATLANTA
WASHINGTON DC
JLAB
NEW YORK
BNL
PPPL
BOSTON
LONDON
AMSTERDAM
GÉANT
GENEVA
CERN
ESnet
ENERGY SCIENCES NETWORK
Department of Energy Office of Science National Labs
Ames Ames Laboratory (Ames, IA)
ANL Argonne National Laboratory (Argonne, IL)
BNL Brookhaven National Laboratory (Upton, NY)
FNAL Fermi National Accelerator Laboratory (Batavia, IL)
JLAB Thomas Jefferson National Accelerator Facility (Newport News, VA)
LBNL Lawrence Berkeley National Laboratory (Berkeley, CA)
ORNL Oak Ridge National Laboratory (Oak Ridge, TN)
PNNL Pacific Northwest National Laboratory (Richland, WA)
PPPL Princeton Plasma Physics Laboratory (Princeton, NJ)
SLAC SLAC National Accelerator Laboratory (Menlo Park, CA)

Direct Connect
ESnet Pilot 2x10G

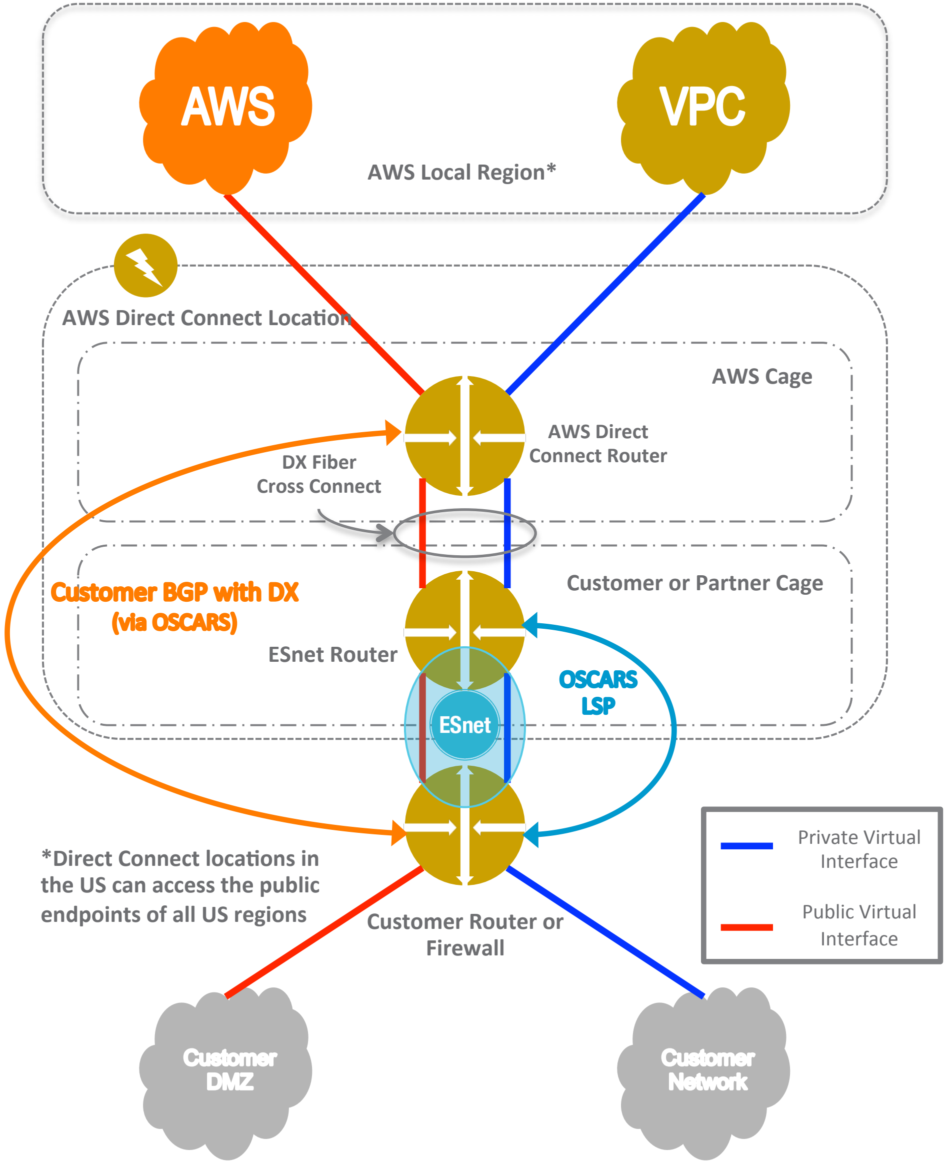
100G R&E
Exchange

Increased ESnet Connectivity
to AWS US West regions

M. O'Connor, Y. Hines, Amazon Web Services Pilot Report, ESnet Report, September 2016,

http://es.net/news-and-publications/publications-and-presentations/

AWS Direct Connect Public and Private Interfaces



*Direct Connect locations in the US can access the public endpoints of all US regions

Pros and Cons of DX

Pro

- Data sharing directly from a Virtual Private Cloud over R&E networks
  - Deterministic
  - Completely controls Egress fees
- Shortest BGP path over direct AWS to site BGP peering
- Dedicated physical path into the AWS network
  - doesn't compete with commercial traffic

Con

- Costly, additional DX cost up to $20K annually
- 10G network infrastructure vs 100G Amazon public peerings
- IPv4 address range constraints limit VPC address mapping scalability
- S3 storage can not currently be mapped into a VPC
- Requires scripting of BGP filter policies

The AWS Cloud
Changing the rules of the Network Game

The Amazon Web Service network bends a number or 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 region(s), **NOT** the entire AS. Establishing BGP with AWS, ASN 16509 in one region doesn't establish connectivity to other regions, continents etc.
- Surprisingly, routing to remote AWS regions will NOT use your established BGP peering with AWS (ASN 16509)

This is a challenge for globally distributed computing, but it's not necessarily a bad thing for NREN customers

Amazon is not in the “Networking” 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 must pay for upstream transit to the Internet and so they pass this on to their customers.
- Their inter-region long haul circuits are primarily 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”.

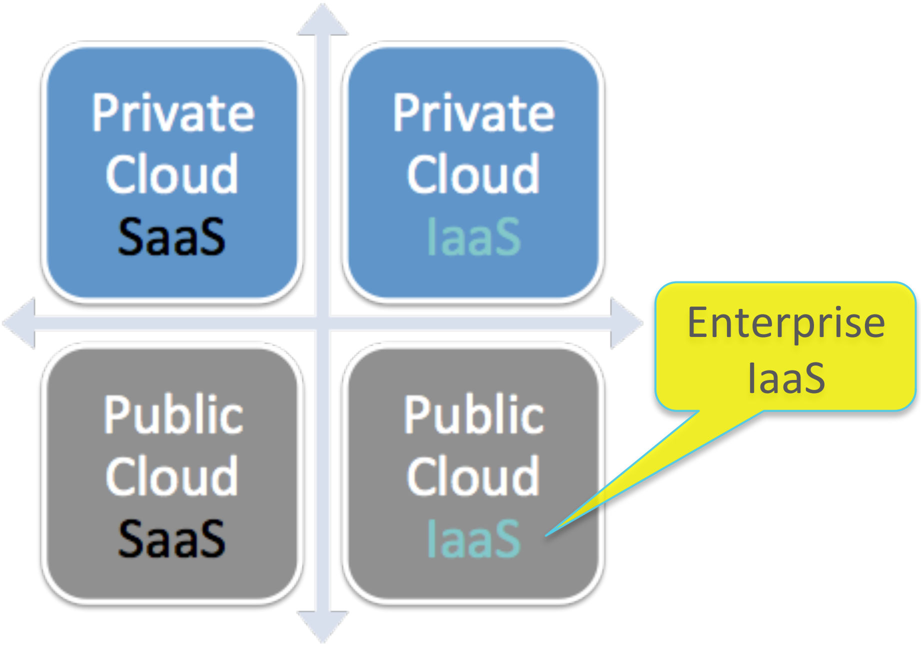
Research & Education Networks have an opportunity to scale “Cloud” beyond the local region for their customers.

Enterprise IaaS

The **Enterprise IaaS** approach is in use by many NREN customers today, where all data transfers are only between that customer site and AWS using Amazon public IP addressing

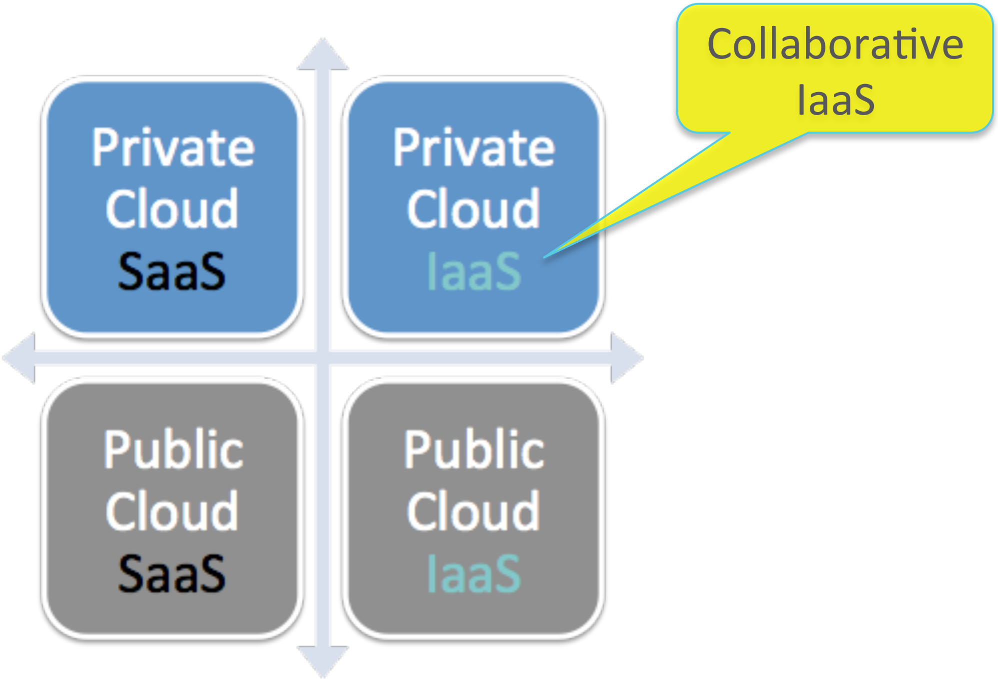
Limitations of Enterprise IaaS

- Difficult to scale to 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



Collaborative Infrastructure as a Service (IaaS)

The principal assumption: When geographically dispersed collaborations begin to transit large datasets directly out of the cloud, **VPC services** will be essential in order to control global routing between collaborating institutions



VPC – Virtual Private Cloud

Why VPC?

- Global collaborations will eventually 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.
- The public internet is highly fragmented and not engineered to support the kind of Scientific flows required by the globally distributed computing model.
- The LHC globally distributed computing model requirements can not be met deterministically by the general Internet.
- **Without VPC, all European NRENs could peer directly with AWS and it would not keep flows to US AWS regions off the public Internet**

VRF/Overlay Solution: Scaling

Alternative to VPC

Scalability: This model requires NRENs and sites to "join" a "cloud VRF" and reciprocate by advertising all of their cloud prefixes into the shared overlay network

Technical issues:

- Connecting a common VRF across multiple NRENs and their customers over large geographic distances is not trivial. (ie: LHCONE)
- In order to preserve a closed network, NRENs would peer with all cloud providers a second time at each existing location on the Cloud VRF.
- Participating sites may now have to deal with policy routing multiple overlay networks that can not reasonably be expect to remain distinct from each other.
- Difficulties coexisting with LHCONE
    - Will the Cloud VRF be completely isolated? (simple but less useful)
    - Will prefixes overlap with LHCONE? (Complex but more useful)
    - Will it peer with LHCONE? (In that case, why not simply use LHCONE instead)
    - Which network should be preferred? (difficult to reach consensus)

ESnet

VRF/Overlay Solution: Policy

Alternative to VPC

Policy issues:

- Is encapsulating traffic in a VRF a compelling enough reason to allow changes to NREN policies regarding commercial transit and appropriate use?
- When NRENs export commercial transit routes into an R&E VRF, how badly does the case for flexible AUPs erode as the VRF table exponentially increases in size and becomes predominately commercial.
- Will enough NRENs agree to change their transit policies in order to participate?
- **Fermilab criteria (Phil Demar)** – Will this solution support transit over R&E networks between cloud providers?
- Almost certainly NOT, cloud providers do not routinely accept transit to other cloud providers over an R&E VRF network. For instance, would Amazon agree to use ESnet to reach Google? (Not Likely)
- Transport between cloud providers will use the general Internet.

VRF/Overlay Solution: Functionality

Alternative to VPC

Functional Requirements:

- Huge effort, requiring a common VRF implemented by all participating NRENS and their customers.
- No way to ensure ALL egress traffic will transit NRENs for all collaborators. This solution may incur egress traffic fees.
- How will NRENs distinguish cloud service routes from on-line retail or other cloud customers? Should on-line retail routes be allowed in the routing table of a Scientific controlled access network?

Security:

- There is no way to restrict unaffiliated cloud customers from having access to collaborating sites through this network. For example, any random cloud customer that happens to reside in the right CIDR block will have a high performance path to hosts at collaborating institutions.
- Would sites resort to managing host based access tables on every cloud instance, might lack of host table coordination create issues reaching new sites?

Open Exchange Model

Improve Efficiency & Simplify

We can break the problem into smaller pieces using open exchange points (OPXs)

- NREN ensures connectivity for user institutions (as usual)
- NREN connects to one or more OPXs (and each-other)
- Cloud provider has NREN connectivity through:
  - Direct connections to NREN
  - Connections to one or more OXPs

From: NORDUnet's views on cloud and cloud providers, Taipei, March 2016

Open Exchange Model: Limitations

- By encouraging cloud providers to join common exchange points, only regional path efficiency can be improved
- Regional cloud to cloud transfers over public networks can benefit if all cloud providers peer with each other at every exchange
- Optimizing regional connectivity is only a partial solution, it may improve the single customer cloud model but only some cloud to cloud connections
- This approach will not support global data distribution over R&E network transport or the Globally Distributed Computing Model
- Adding VPC to this architecture would address many of the outstanding issues
- Scaling this model to work with GNA will be difficult since it is only a regional model.

From: NORDUnet's views on cloud and cloud providers, Taipei, March 2016

Collaborative IaaS
Cloud "On-Ramp" to R&E Networking

Virtual Private Cloud over R&E networks enhance global data sharing

- A "Collaborative IaaS" model will facilitate cloud based data transport beyond the Enterprise
- VPC enables a customer to control the routing of their data by providing a means of mapping customer IP allocations into the cloud
- Networks that do not control their own addressing lack the authority to define routing policy for the addresses in "their" network.
- **Even if all European NRENs peer directly with AWS, it won't keep flows to US AWS regions off the public Internet**

Global collaborations need to control transport of large data sets

Collaborative IaaS

Based on VPC

Technical issues:

- Scales well since there are no additional requirements for collaborating NSPs or their customers to coordinate or implement any changes
- Externally, VPC cloud networks are treated exactly the same as any other prefix sourced by the compute center
- Each customer would BGP peer with the cloud provider in each region that they intent to use (L2 Circuits)
- No additional VRF and no third party peering requirements

Fermilab criteria (Phil Demar) – Will this solution support transit over R&E networks between cloud providers?

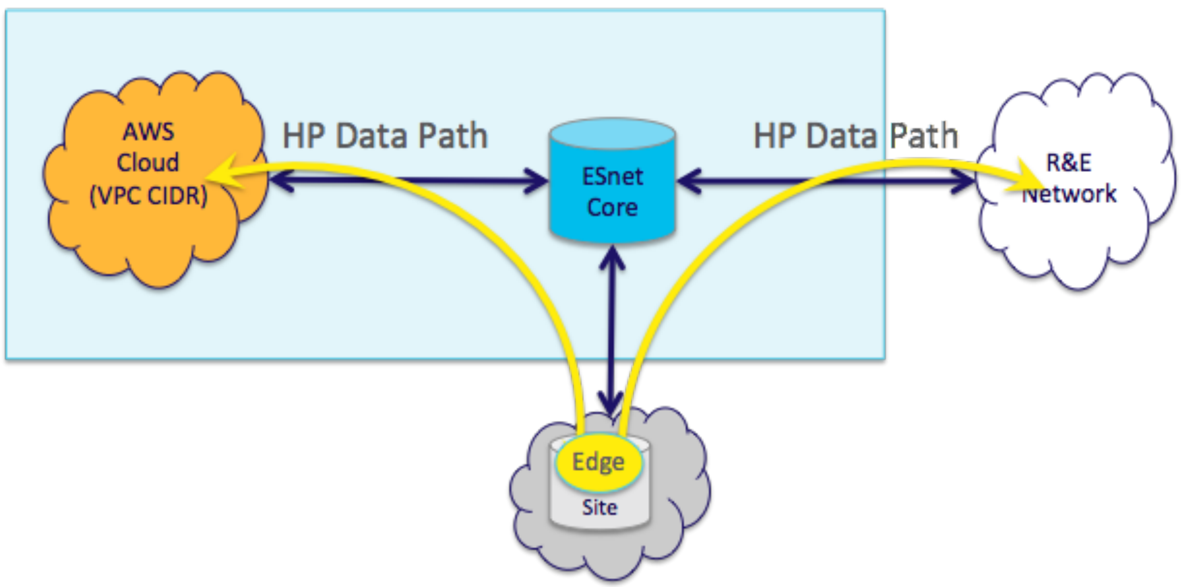
  - YES!

- **Integrates seamlessly with LHCONE**
  - Any site controlled VPC prefix can be routed over LHCONE
  - Commercial cloud providers don't need to peer at any particular location or with any other cloud provider
- The cloud provider must support Virtual Private Cloud services
- The customer will map a portion of their assigned or allocated IP space into the service providers network

Collaborative IaaS

Based on Virtual Private Cloud

Standard VPC Third Party Data Path

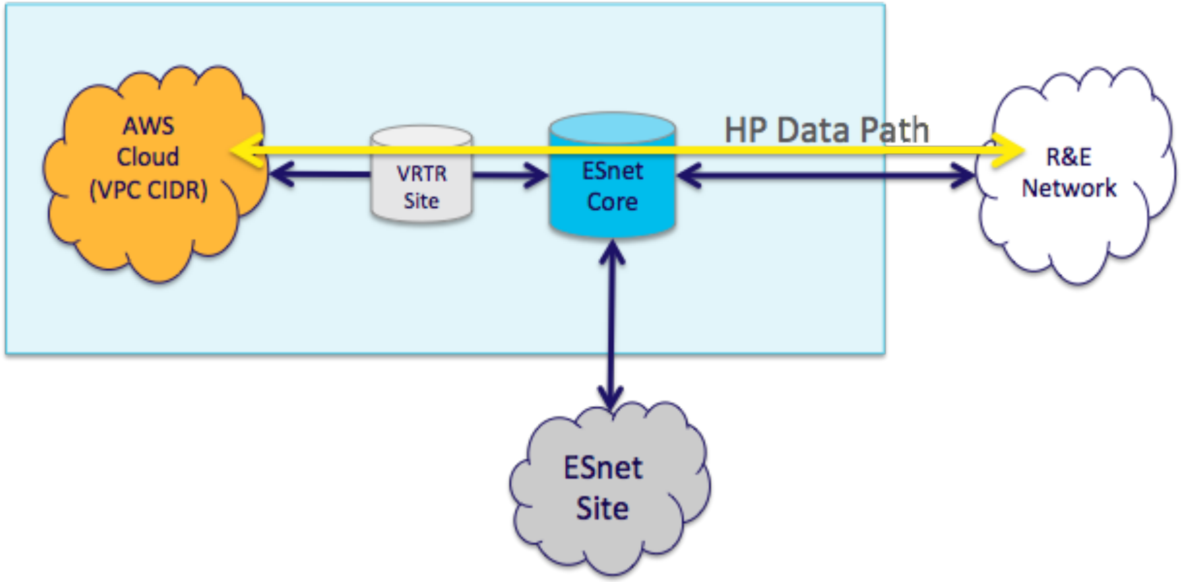


- Standard VPC implementation provides an **“On-Ramp”** for cloud to R&E Network Transport
- Implemented over common public peering with AWS

Virtual "Site Router" (VRTR) Service

Improving the Path Efficiency of Collaborative IaaS

Virtual Site Router at AWS Exchange Point



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

- Multiple virtual routers per hardware device
- IPsec will connect the SDN Site router to the cloud in order to support VPC
- Collaborator transit paths take the best path to the cloud and back
- Improves path efficiency and takes pressure off of the site local-loop

Collaborative IaaS Conclusions

Functional Requirements:

- IPv4 addressing is limited and insufficient for large VPC footprints
- R&E networks should push cloud providers for IPv6 support
- All traffic will egress onto NRENs for all NREN hosted collaborators, eliminating egress traffic fees
- The small controlled LHCONE routing table will remain small and controlled

Security:

- Restricting unaffiliated cloud customers from direct network access is a feature, not a problem
- VPC LANs can reside in an enterprise secure perimeter, ingress and egress in the same way as any other LAN
- Using a more standard network access control model reduces the complexity of relying solely on host based access tables

ESnet AWS Testing

Successful pilot testing in the following areas

- Physical Network
  - Cross connection & LOA process
- Layer two tagged VLAN interface with multiple VLANs
- Customer site connecting to AWS over ESnet physical infrastructure
- Routing
  - BGP routing and policy
  - Public & Private cloud
  - VPC CIDR mapping
  - Access Filters
- Billing
  - Egress waiver verified
  - Separate the DX charges onto the ESnet account
  - Compute fees charged to customer site
- Portal functions
  - Provisioning
  - Reporting
  - Account management

Future Requirements & Testing

- Cloud Provider Requirements
  - IPv6 support for VPC
  - Cloud storage VPC support, ie: S3
- IPsec tunnel performance testing
- SDN/Virtual Router
  - Features & Functionality
  - Performance
  - Configuration & Management

The remaining requirements and testing are squarely aligned with existing network initiatives in the R&E community

ESnet
ENERGY SCIENCES NETWORK

Thank You

Michael O'Connor moc@es.net

ESnet

Network Engineering

LHCOPN-LHCONE meeting

Helsinki (FI)

September 19, 2016

Additional Slides

AWS Egress Waiver

Intended to address concerns raised by the R&E community about Amazon holding their data hostage.

AWS has extended an **egress traffic fee waiver** to all ESnet, I2 & GEANT customers

- R&E networks are considered "zero cost transit peers"
- Up to 15% of monthly bill total waived for traffic that exits the AWS network via ESnet
- Caveat: traffic egressing "non-zero cost transit peers" will incur fees *(when AWS begins measuring it)*

Despite repeated ESnet requests, AWS will not provide a written service description of the egress waiver offered to ESnet and it's customers