Network working group
Internet Draft
Category: Informational

L. Yong
Huawei
M. Toy
Comcast
A. Isaac
Bloomberg
V. Manral
Hewlett-Packard

Expires: December 2012

June 22, 2012

Use Cases for DC Network Virtualization Overlays

draft-mity-nvo3-use-case-00

Status of this Memo

Copyright Notice

Abstract

This draft describes NVO3 use cases. The framework draft [NVO3FRWK] layouts the NVO3 architecture reference model, functional modules, and NVO3 and VPN interworking aspects and challenges. This draft presents the use cases that help in validating the framework and requirements as well as the solution development.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [RFC2119].

Table of Contents

1. Introduction

This document provides the use cases for Network Virtualization Overlay, i.e. NVO3, which is driven by Data Center Networks. These use cases are intended to help in validating the framework and requirements as well as the solution development.

The advent of the hypervisor eliminated the tight coupling of an endpoint from the physical computer on which it ran. This change has allowed the physical computer to become a service point rather than a client. The goal of NVO3 is to no longer treat the physical computer as a client of the network but as a native service point of the network.

Although overlay networks have been around for many years, hypervisor-aware overlay networks have certain characteristics that are suited for the DC environment. The main characteristic difference between other overlay network technology and NVO3 is that the client edges of the NVO3 network are individual virtualized hosts and not network sites or LANs. Other differentiating characteristics include (1) the potential for lower overall costs when compared to comparable network-based overlays, (2) better loop-free scaling over traditional DC network virtualization solutions (3) better virtual host access and mobility.

The NVO3 framework [NVO3FRWK] provides the architecture reference model, functional modules, and the overlay/underlying network aspects, which empowers Data Center service designs. Data Center service is to provide applications and/or virtual compute and/or storage and networking. The key requirements for Data Center service networks are to enable secure, virtual, and segregation networks that are highly scalable in number and size, native extension of these virtual networks to the virtual interface of virtual machines, flexible operator-defined virtual network topologies, natural interconnection of networks and interposing of network services, and the simple and rapid enablement of network access and services. Either an L3 or L2 virtual network instance can be constructed over the underlying networks.

Use cases for the NVO3 framework [NVO3FRWK] can be highly varied. This document outlines some basic scenarios and groups them into three set.

One set of NVO3 use cases is connecting many, many tenant end systems in one and/or multiple DC sites and forming an L2 or L3 communication domain. Using overlay tenant virtual network instances

segregate the traffic from different tenants, allows each tenant instance having its own address space and isolating the space from DC infrastructure. In addition, support VM move.

The second set of NVO3 use cases is for a DC provider to offer a secure DC service to an enterprise customer. In these cases, the enterprise customer may use a traditional L3/L2 VPN provided by a carrier connecting to an overlay virtual network instance offered by a Data Center provider, and offload its applications on to the servers located in provider Data Center sites.

The third set of NVO3 use cases is to enable DC provider design various cloud services through the applications, compute, storage, and the networking. In this case, NVO3 provides the networking functions of a service rather than the service itself.

2. Terminology

This document uses the terminologies defined in [NVO3FRWK], [RFC4364]. Some additional terms used in the document are listed here.

VNIF: VNI Interconnection Interface on an NVE

L2 VPI: L2 Virtual Network Instance

L3 VNI: L3 Virtual Network Instance

ARP: Address Resolution Protocol

DNS: Domain Name Service

NAT: Network Address Translation

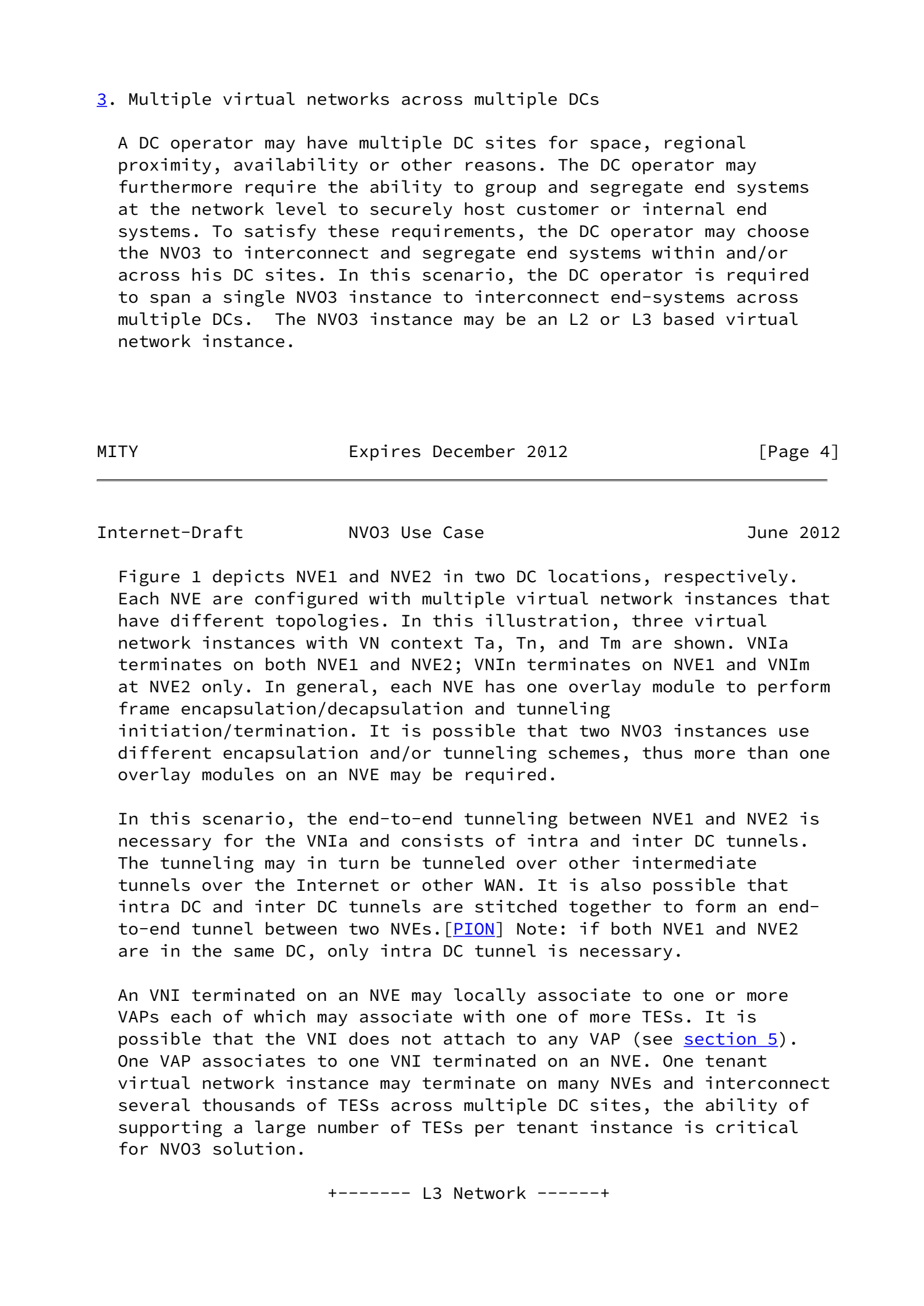
3. Multiple virtual networks across multiple DCs

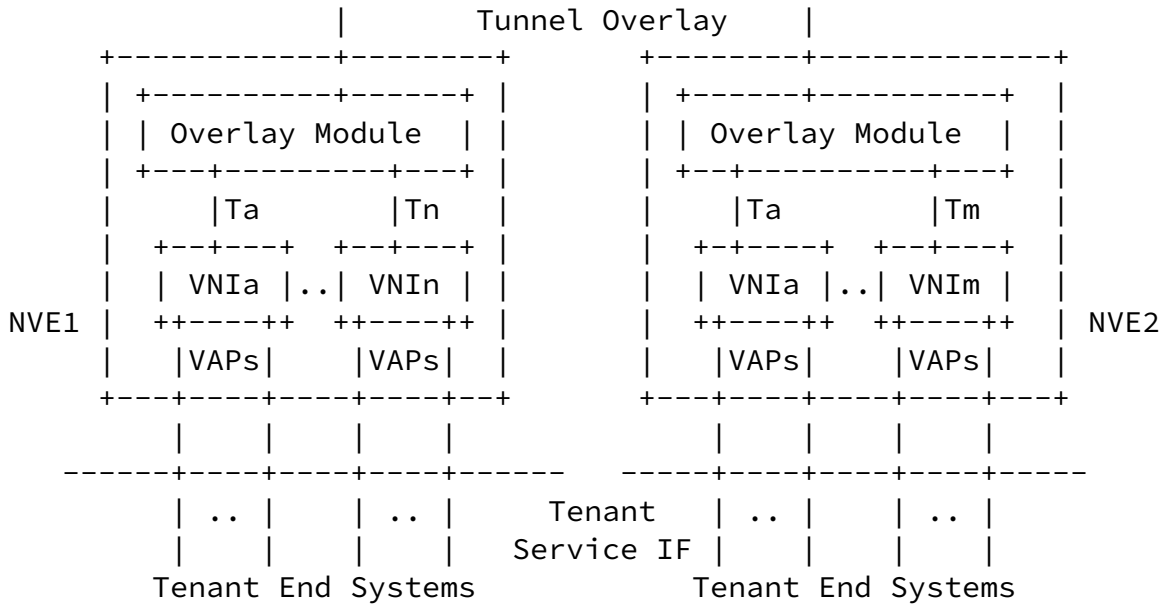
A DC operator may have multiple DC sites for space, regional proximity, availability or other reasons. The DC operator may furthermore require the ability to group and segregate end systems at the network level to securely host customer or internal end systems. To satisfy these requirements, the DC operator may choose the NVO3 to interconnect and segregate end systems within and/or across his DC sites. In this scenario, the DC operator is required to span a single NVO3 instance to interconnect end-systems across multiple DCs. The NVO3 instance may be an L2 or L3 based virtual network instance.

Figure 1 depicts NVE1 and NVE2 in two DC locations, respectively. Each NVE are configured with multiple virtual network instances that have different topologies. In this illustration, three virtual network instances with VN context Ta, Tn, and Tm are shown. VNIa terminates on both NVE1 and NVE2; VNIn terminates on NVE1 and VNIm at NVE2 only. In general, each NVE has one overlay module to perform frame encapsulation/decapsulation and tunneling initiation/termination. It is possible that two NVO3 instances use different encapsulation and/or tunneling schemes, thus more than one overlay modules on an NVE may be required.

In this scenario, the end-to-end tunneling between NVE1 and NVE2 is necessary for the VNIa and consists of intra and inter DC tunnels. The tunneling may in turn be tunneled over other intermediate tunnels over the Internet or other WAN. It is also possible that intra DC and inter DC tunnels are stitched together to form an end-to-end tunnel between two NVEs.[PION] Note: if both NVE1 and NVE2 are in the same DC, only intra DC tunnel is necessary.

An VNI terminated on an NVE may locally associate to one or more VAPs each of which may associate with one of more TESs. It is possible that the VNI does not attach to any VAP (see section 5). One VAP associates to one VNI terminated on an NVE. One tenant virtual network instance may terminate on many NVEs and interconnect several thousands of TESs across multiple DC sites, the ability of supporting a large number of TESs per tenant instance is critical for NVO3 solution.





DC Site A DC Site Z

Figure 1 NVo3 for DC interconnection

Individual virtual network instances may use its own address space and the space is isolated from DC infrastructure. This eliminates IP subnet constraints in the infrastructure when moving VMs. Note: the NVO3 solution still have to address the discovery of VM move.

4. Interconnection of DC Virtual Network and Carrier VPN

In this scenario, the enterprise customer utilizes the DC provider's compute and storage resources to run its applications, and the DC provider allows the customer to access his hosted end systems through a Carrier WAN. This use case requires a tenant instance in the DC to interconnect with a carrier VPN for customer access. The enterprise customer may also utilize the Carrier's VPN to connect its corporate locations, and/or the DC provider's locations where his tenant instance exists. The DC provider creates a tenant instance for the enterprise customer that interconnects all the VMs and storage resources allocated to the customer. The customer unique instance provides the customer with traffic segregation from other customers and free selection of the address space. Both DC VNI and carrier VPN instance can run at either L2 or L3.

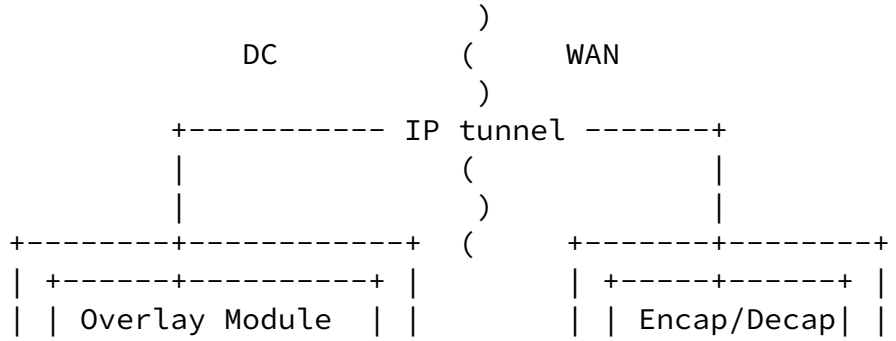
4.1. One virtual network method across DCs and WAN networks

If both DC Provider and Carrier use the same encapsulation and tunneling technology, it is possible to configure one overlay virtual network instance across DC networks and Carrier networks. For example, if both DC provider and Carrier use existing MPLS-based VPN solutions [RFC4364] and IP tunnel, the NVE in DC and the PE in WAN can be members of one VN instance. Figure 2 illustrates this scenario. The left side of the figure presents a NVE (NVE1) in DC Provider site and the right shows Provider Edge (PE1) in a WAN network connecting to Customer Edge (CE) at an Enterprise site. The CE is often a network site and contains routers and/or switches and terminal systems.

In this case, an L3 VNI and L3VPN instance are configured on NVE1 and PE1, respectively. If the MPLS label is used as VN context/VPN identifier and IP tunnel is established between NVE1 and PE1, the configuration will provide the L3 connectivity between a TES and CE. The MPLS label for the L3 VNI identifier (Ta) on NVE1 can be different from the MPLS label for the L3VPN identifier (VPNID) on PE1 since MPLS labels are locally significant. Although the figure

shows Overlay Module on NVE1 and Encap/Decap (Encapsulation/Decapsulation) on PE1, both perform the same functions; it is just matter of using different terminologies in NVO3 framework [NVO3FRWK] and L3VPN [RFC4364].

The DC and WAN networks may belong to different ASs, control plane protocol or management plane can facilitate the VN configuration. Note: A TES is an end system, not a network site as the CE in figure 2; more than likely it does not run any routing protocol. Thus the routing between NVE1 and TES are static routing, which may be done by the API or software assisted configuration in a secured way. Routing and forwarding between NVE1 and PE1 and between PE1 and CE are specified in RFC4364 [RFC4364].



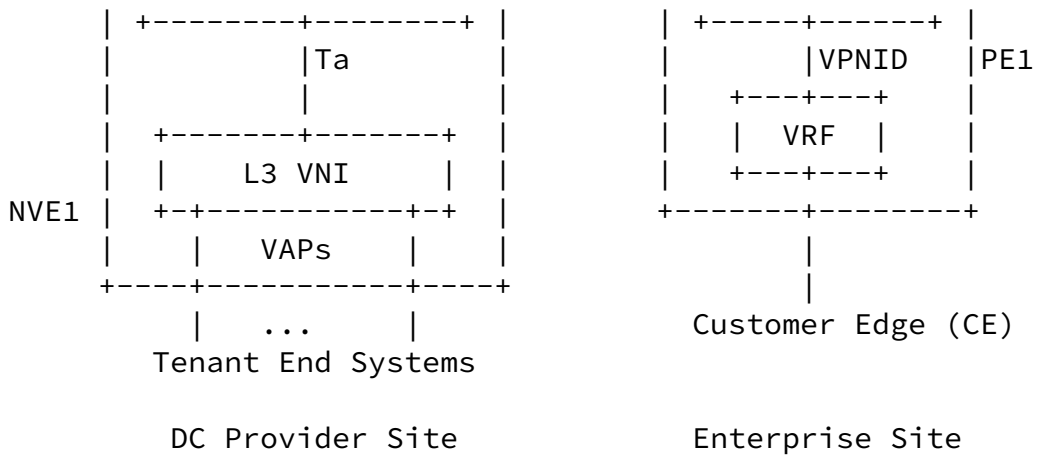


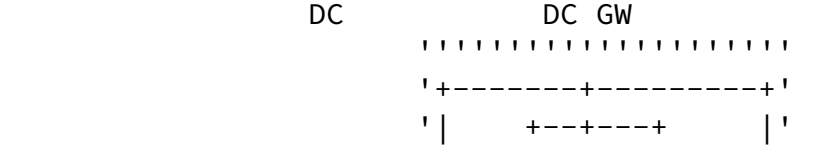
Figure 2 One VN solution across DCs and Carrier Networks

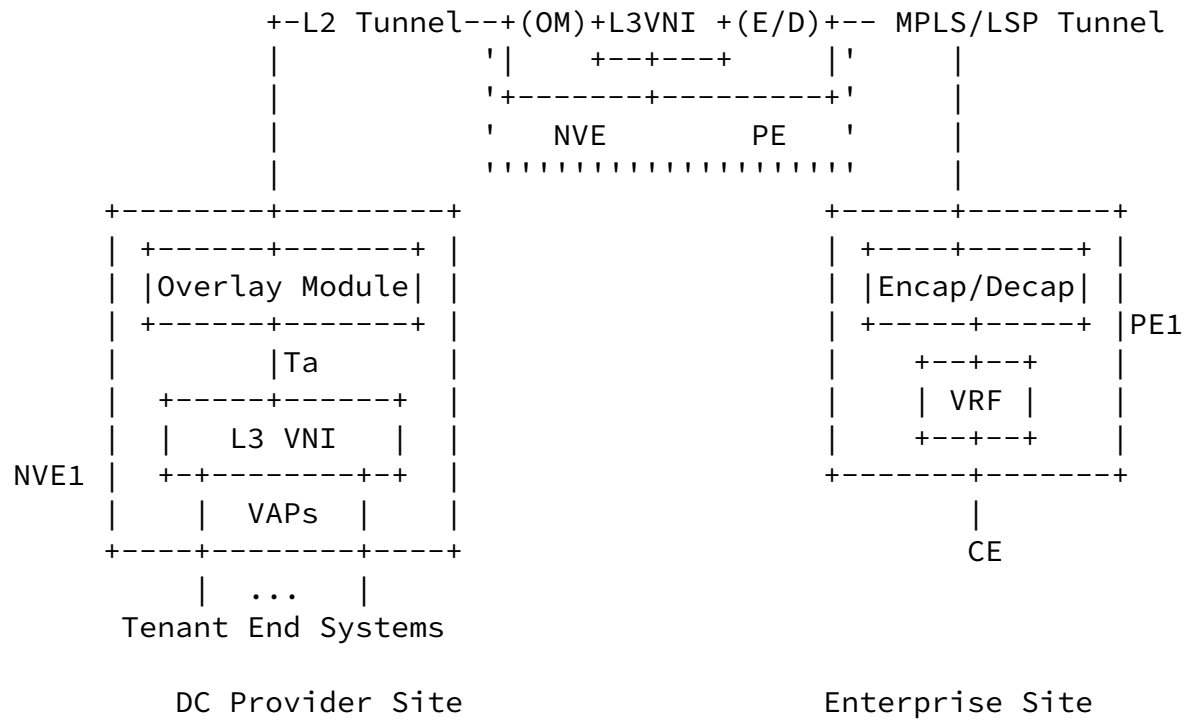
4.2. NVO3 and VPN Interconnections between DC and WAN networks

DC Provider and Carrier may build a tenant instance and VPN instance for an enterprise customer independently and interconnect two together at DC GW. Figure 3 depicts this case. The GW supports both NVE and PE capability. Here an L3 VNI instance is built between NVE1 and DC GW and an L3VPN instance is configured on DC GW and PE1,

respectively. The GW performs L3 VNI function, NVO3 encapsulation, and tunneling toward the DC; it also performs L3VPN function toward the WAN. Both L2 tunnel and LSP Tunnel terminate at DC GW. The packets are processed at the L3 VNI on DC GW. Operator may choose use of one routing table for both instances as shown in the figure or one for each. This implementation is more complex than one in figure 2. However it provides DC network and WAN network demarcation clearly and allows each network use of different VN implementations, which is necessary in some situations.

The alternative of this case is physically split the gateway function on to DC GW and WAN PE devices. In this case, the tenant instance is terminated on the DC GW and L3VPN instance terminates at a PE in WAN. An Ethernet interface is used to physically connect to the DC GW and PE devices and an Ethernet VLAN is configured on both devices for interconnecting two instances.





Note: OM: Overlay Module; E/D: Encap/Decap

Figure 3 L3 VNI and L3VPN interconnection across multi networks

If an enterprise only has few locations, it may use P2P VPWS connecting to the DC GW. Further some enterprises may connect to DC GW via Internet by using IPsec. In this case, DC GW needs to provide some authentication scheme for the security.

Such interconnection may also apply to one or across multiple DC sites. During the migration process, it is possible that some portion of DC site may be able to support NVE and other may not. Such gateway function may be used to interconnect a tenant instance and a regular underlying VPN that provides the connectivity to the VMs belonging to the same tenant.

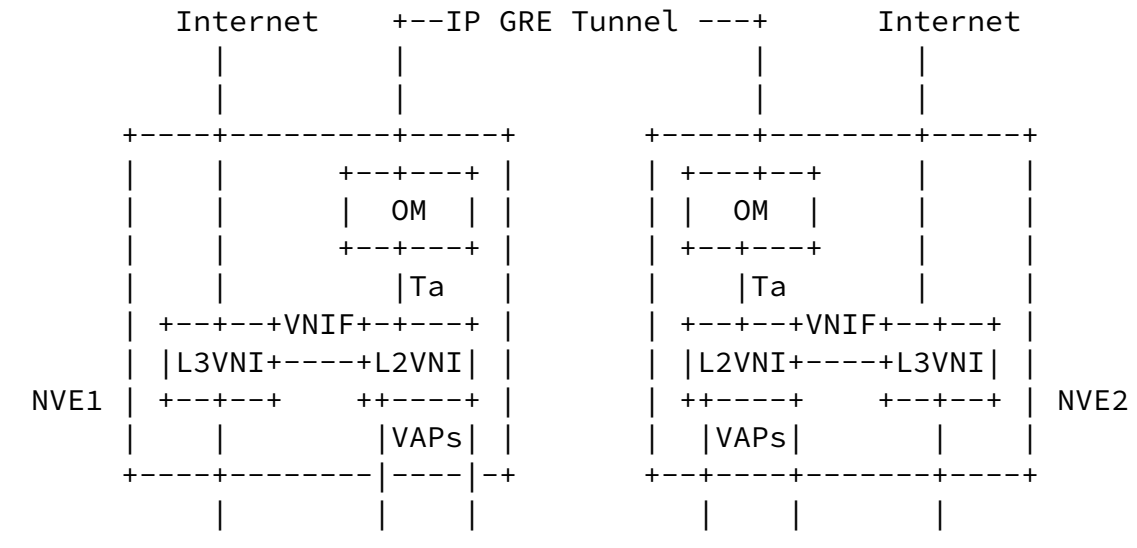
5. Cloud Services Using NVO3

NVO3 brings DC operators power and flexibility to design different applications for cloud services. DC operators construct VMs, storage, applications, and interconnect them together via an VNI. Since DC operator manages both tenant end systems and NVEs, they can assign some functions on VMs and some on NVEs. For example, designate a VM for running firewall and/or a VM for DNS for a tenant; apply

forwarding policies and/or access list on an VNI terminated on an NVE. Operators may dedicate some VMs and/or storages for these applications and allocate other VMs/storages for running tenant specific applications. Furthermore, the design of a cloud service for a customer may often require creating both L2 VNI and L3 VNI per a tenant on one or more NVEs and interconnecting the VNIs together like physical router and switch devices in a traditional DC.

5.1. Public Cloud Service

Figure 4 depicts that an L2 VNI is used to connect all the TESs together and provides a simple LAN among TESs; an L3 VNI is used for public Internet Access and firewall access. Note: An L3 VNI provides virtualized IP routing and forwarding and an L2 VNI provides Ethernet LAN emulation. The firewall application runs on a tenant end system connecting to the L3 VNI terminated on NVE1 and NVE2. Ta in Figure 4 is the VN context of the L2 VNI. Internal VNI interconnecting interface (VNIF) is used to connect the L2 VNI and L3 VNI on NVE1 and NVE2. In this case, the L3 VNI connects to external switch directly, not to an overlay module. Operator defined policies can directly apply to the L3 VNI terminated on NVE. Furthermore, the L3 VNI does not have to co-exist with the L2 VNI on every NVEs; the several NVEs where the L2 VNI terminates can connect to one NVE where the L3 VNI terminates. In this case, VNIF reachability should be known by these NVEs.



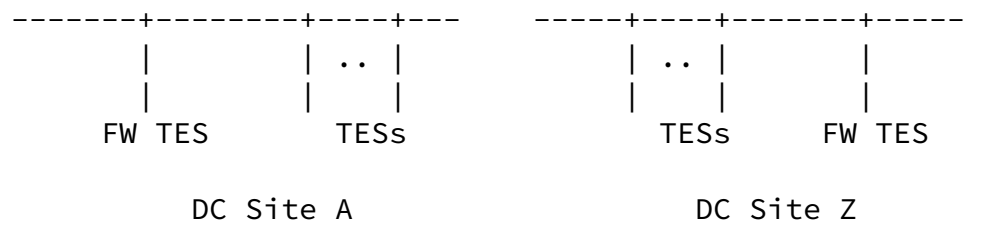
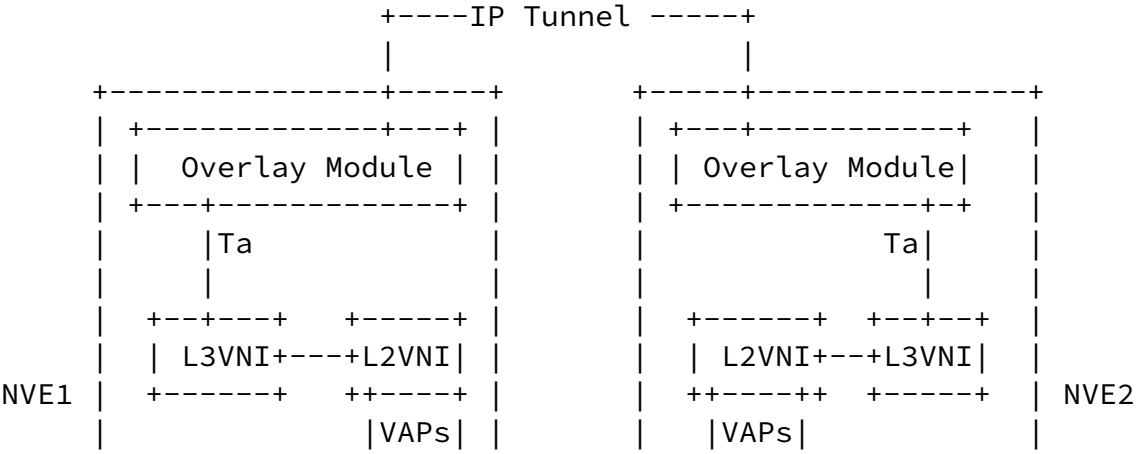


Figure 4 L2 VNI for connecting TESs and L3 VNI for public Internet access

In this case, if the tenant uses private IP addresses on Tenant End Systems, the NAT is needed for public Internet access. Operator may allocate another TES connecting to L3 VNI and run NAT application on it.

5.2. Private Cloud Service

Figure 5 below illustrates another overlay bridging and routing example. A L2 VNI is used for local LAN switching and L3 VNI is used for interconnecting NVEs remotely and providing IETF IP VPN functionality. In this case, the L2 VNI terminated on an NVE only has local significance.



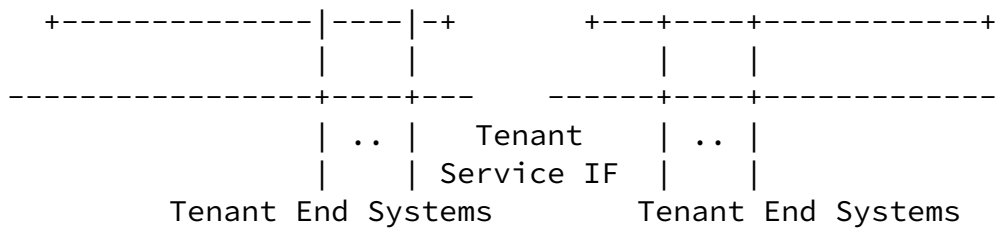


Figure 5 NVO3 for Virtual Bridging and Routing Per Tenant

The DC provider may also create an NVE on DC GW and configure the L3 VNI on the NVE (instead of co-resident with L2 VNI on the same NVE). Thus L2 overlay is used within a DC, and L3 overlay is used for inter-DCs.

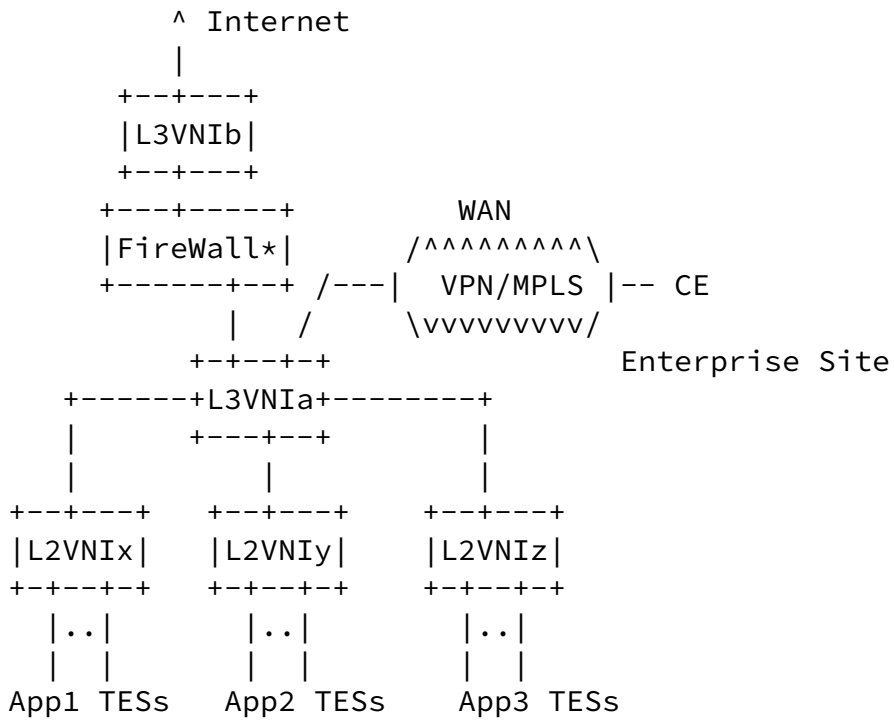
5.3. Virtual Data Center

Enterprise DC today may often uses several routers and switches devices to construct a secure network for intranet and extranet. [SANS] DC Provider may want to offer Virtual DC called Infrastructure as service (IaaS) to such enterprise customers. Instead of using many router/switch hardware devices, with the overlay and virtualization technology of NVO3, DC operators can build them on top of a common network infrastructure for many customers and run applications per customer basis. The applications may include firewall, DNS, load balancer, NAT, etc.

Figure 6 below illustrates this scenario. For the simplification, it only shows the VNIs as logic routers or switches. In this case, DC operators construct several L2 VNIs (x, y, z in figure 6) to group the end tenant systems together per application basis, create an L3 VNI (L3 VNIa in the figure) for internal routing and another L3 VNI (L3 VNIb in the figure) for broad routing. Allocate a TES to run a firewall application between L3 VNIa and L3 VNIb. The design implements the security policy with the appropriate firewall rules and Layer 3 access list. Configure an interconnection between the L3VNIa to an L3VPN over the WAN to reach Enterprise Sites.

The Enterprise as a customer decides which applications are accessed by intranet only and which by both intranet and extranet; DC operators then design and configure the proper security policy. DC operators can further set different QoS levels for the different applications for a customer.

This application requires NVO3 solution to provide DC operator an easy way to create NVEs and VNIs for any design and to quickly assign TESs to an VNI, and configure policies on an NVE easily.



* firewall may run on VMs

Figure 6 Secure Cloud Service by Using NVO3

6. OAM Considerations

NVO3 brings the ability for a DC provider to segregate tenant traffic. A DC provider needs to manage and maintain NVO3 instances. Similarly, the tenant needs to be informed about tunnel failures impacting tenant applications.

Various OAM and SOAM tools and procedures are defined in [IEEE 802.1ag, ITU-T Y.1731, RFC4378, ITU-T Y.1564] for L2 and L3 networks, and for user, including continuity check, loopback, link trace, testing, alarms such as AIS/RDI, and on-demand and periodic measurements. These procedures may apply to tenant overlay networks and tenants not only for proactive maintenance, but also to ensure support of Service Level Agreements (SLAs).

As the tunnel traverses different networks, OAM messages need to be translated at the edge of each network to ensure end-to-end OAM.

It is important that failures at lower layers which do not affect NVo3 instance are to be suppressed.

7. Summary

The document intends to illustrate some basic potential use cases. The combination of these cases should give operators flexibility and power to design more sophisticated cases for various purposes.

The main characteristic difference between other overlay network technologies and NVO3 is that the client edges of the NVO3 network are individual virtualized hosts and not network sites or LANs. NVO3 is to no longer treat the physical computer as a client of the network but as a native service point of the network. The same operator manages both NVO3 network and its clients.

NVO3 lets individual virtual network instances use its own address space and isolate the space from the network infrastructure. The approach not only segregates the traffic from multi tenants on a common infrastructure but also makes VM mobility easier within a tenant.

Cloud services are about providing virtual processing/storage, applications, and networking in a secured and virtualized manner, in which the NV03 is just a portion of a service, not a service itself. NVO3 decouples cloud services and DC network infrastructure.

NVO3 underlying network provides the tunneling between NVEs so that two NVEs appear as one hop each other. Many tunneling technologies can serve this function. The tunneling may in turn be tunneled over other intermediate tunnels over the Internet or other WAN. It is

also possible that intra DC and inter DC tunnels are stitched together to form an end-to-end tunnel between two NVEs.

The key requirements for NVO3 are 1) traffic segregation; 2) support a large scale number of TESs in an VNI; 3) VM mobility 4) easy to construct an VNI and its associated TES; 5) NVO3 Management.

8. Security Considerations

Please see it in NVO3 Framework. [NVO3FMWK]

9. IANA Considerations

This document does not request any action from IANA.

10. Acknowledgements

Authors like to thank Linda Dunbar, Sue Hares, and Young Lee for the review and suggestions.

11. References

11.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[IEEE 802.1ag] Virtual Bridged Local Area Networks - Amendment 5: Connectivity Fault Management, December 2007.

[ITU-T G.8013/Y.1731] OAM Functions and Mechanisms for Ethernet based Networks, 2011.

[ITU-T Y.1564] Ethernet service activation test methodology, 2011.

[RFC4378], A Framework for Multi-Protocol Label Switching (MPLS) Operations and Management (OAM)

11.2. Informative References

[NVO3FRWK] Lasserre, M., Motin, T., and etc, "Framework for DC Network Virtualization", draft-lasserre-NVO3-framework-02, June 2012.

[PION] Jin, L. and Khasnabish, B., "Architecture of PSN Independent Overlay Network (PION)", draft-kj-nvo3-pion-architecture-00, May 2012

[SANS] Daniel Oxenhander, "Designing a Secure Local Area Network", 2003

Authors' Addresses

Lucy Yong
Huawei Technologies,
4320 Legacy Dr.
Plano, Tx75025 US

Phone: +1-469-277-5837
Email: lucy.yong@huawei.com

Mehmet Toy
Comcast
1800 Bishops Gate Blvd.,
Mount Laurel, NJ 08054

Phone : +1-856-792-2801
E-mail : mehmet_toy@cable.comcast.com

Aldrin Isaac
Bloomberg
E-mail: aldrin.isaac@gmail.com

Vishwas Manral
Hewlett-Packard Corp.
191111 Pruneridge Ave.
Cupertino, CA 95014

Phone: 408-447-1497
Email: vishwas.manral@hp.com