Network working group
Internet Draft
Category: Informational

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

Expires: February 2013

August 23, 2012

Use Cases for DC Network Virtualization Overlays

draft-mity-nvo3-use-case-02

Status of this Memo

Copyright Notice

Abstract

This draft describes the generalized NVO3 use cases. The work intention is to help validate the NVO3 framework and requirements as along with the development of the solutions.

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

Compute Virtualization has dramatically and quickly changed IT industry in terms of efficiency, cost, and the speed in providing a new applications and/or services. However, in a Data Center, the configuration on virtual machines is often tied to the physical network configuration, which forces multi-tenant applications to deal with physical network limitations in a virtual environment. This limitation hinders the flexibility in constructing cloud applications and virtual Data Centers.

IETF NVO3 WG works on a solution for DC network virtualization overlays to relax this limitation. The solution will decouple the virtual machines (VM) and DC physical networks and make both VMs and its networking exist in a virtual environment. This will enable to build an IT application in a true virtual environment and isolate the traffic among different applications. The solution will allow constructing many tenant virtual networks on a common network infrastructure and provides: 1) traffic isolation among one another; 2) independent address space in each virtual network and address isolation from the infrastructure's; 3) VM placement and move from one server to another without any physical network limitation. These characteristics will help speed up the configuration of multi-tenant cloud applications and virtual Data Center, and bring the potentials for a new DC application.

Although NVO3 enables a true virtual environment where VMs and net service appliances communicate, NVO3 solution has to address how to communicate between a virtual network and physical network. This is because 1) many traditional DCs already exist and will not disappear any time soon; 2) a lot of DC applications serve to Internet and/or cooperation users; 3) some applications like Big Data analytics which are CPU bound may not want to the virtualization capability.

This document provides three sets of generalized use cases for Network Virtualization Overlays. These use cases are intended to help validate the NVO3 framework and requirements as along with the development of the solutions.

One set of use cases is to connect many tenant end systems in one Data Center and form one L2 or L3 communication domain. A virtual network segregates its traffic from others and allows the VMs in the network moving from one server to another. The case may be used for DC internal applications that constitute the DC East-West traffic.

The second set of use cases is for a DC provider to offer a secure DC service to an enterprise customer and/or Internet users. In these cases, the enterprise customer may use a traditional VPN provided by a carrier or IPsec tunnel over Internet connecting to an overlay virtual network offered by a Data Center provider. This is mainly constitutes DC North-South traffic.

The third set of use cases is to enable the designs of various DC applications using the net service appliance, virtual compute, storage, and networking. In this case, NVO3 provides the virtual networking functions for the applications.

The document uses the architecture reference model and terminologies defined in [NVO3FRWK] to describe the use cases.

2. Terminology

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

VNIF: Internal Virtual Network Interconnection Interface

L2 VNI: L2 Virtual Network Instance

L3 VNI: L3 Virtual Network Instance

ARP: Address Resolution Protocol

CPE: Customer Premise Equipment

DNS: Domain Name Service

DMZ: DeMilitarized Zone

NAT: Network Address Translation

POD: Performance Optimized Data Center

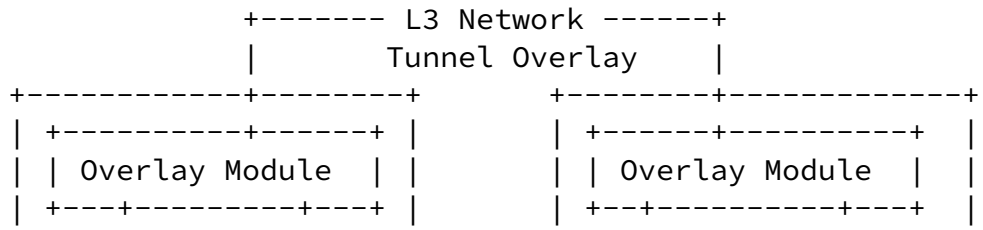
3. Virtual Network in One Data Center

A tenant virtual network may exist in one DC. The network interconnects many tenant end systems that communicate as a closed user group.

Figure 1 depicts this case by using the framework model. [NVO3FRWK] NVE1 and NVE2 are two network virtual edges and each may exist on a

server or ToR. Each NVE may be the members of multiple virtual networks that may have different topologies and run at L2 or L3 individually. In this illustration, three virtual networks with VN context Ta, Tn, and Tm are shown. The VNIa terminates on both NVE1 and NVE2; The VNIn terminates on NVE1 and the VNIm at NVE2 only. Each NVE has one overlay module to perform frame encapsulation/decapsulation and tunneling initiation/termination. In this scenario, a tunnel between NVE1 and NVE2 is necessary for the virtual network Ta. Note that it is possible that one TES participates in more than one virtual network via one VAP for each; further if individual virtual networks use different address spaces, the TES participating in them will be configured with multiple addresses as well. A TES as a gateway is an example.

A VNI on an NVE is a forwarding table that caches and/or maintains the mapping of an end system and its attached NVE. The table entry may be updated by the control plane or data plane or the combination of both. A TES associates to one VNI via a VAP. One tenant virtual network may terminate on many NVEs and interconnect several thousands of TESs, the capability of supporting a lot of TESs per tenant instance and TES mobility is critical for NVO3 solution no matter where an NVE resides.



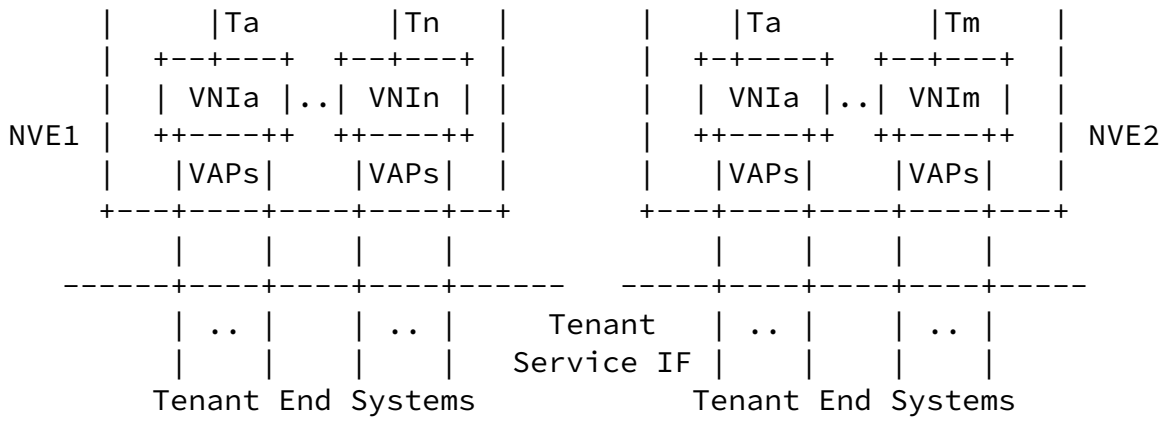


Figure 1 NVo3 for Tenant End-System interconnection

Individual virtual networks may use its own address space and the space is isolated from DC infrastructure. This eliminates the route changes in the DC underlying network when VMs move. Note that the NVO3 solutions still have to address VM move in the overlay network, i.e. the TES/NVE association change when a VM moves.

It is worth mentioning two scenarios regarding to the NVE location. At first an NVE resides on a server, a server manager system such as vCenter [VMWARE] is responsible to create NVE/VNs and VMs, and also responsible to assign a VM to a VN that has unique identification, the server software just makes it works properly and securely. Second an NVE resides on physical switch such as ToR, when a server manger system creates a VM and add it to a VN, the server will send a notification of TES participating in a VN to the local NVE. [ESYS][VDP] Note that if non-virtualized server is used, local configuration on NVE is necessary to attach the TES (server) to a VN. In both cases, when a local NVE notices the new attached TES in a VN, it will announce the TES to remote NVEs or to a mapping server via a control plane protocol. In the case of using mapping server, the remote NVEs can query the server for any TES location and cache it in the VNI.

If a tenant virtual network spans across multiple DC sites, one design is to allow the corresponding NVO3 instance seamlessly span across those sites without DC gateway routers' termination (see section 4.3). In this case, the tunnel between a pair of NVEs may in turn be tunneled over other intermediate tunnels over the Internet

or other WANs, or the intra DC and inter DC tunnels are stitched together to form an end-to-end tunnel between two NVEs.

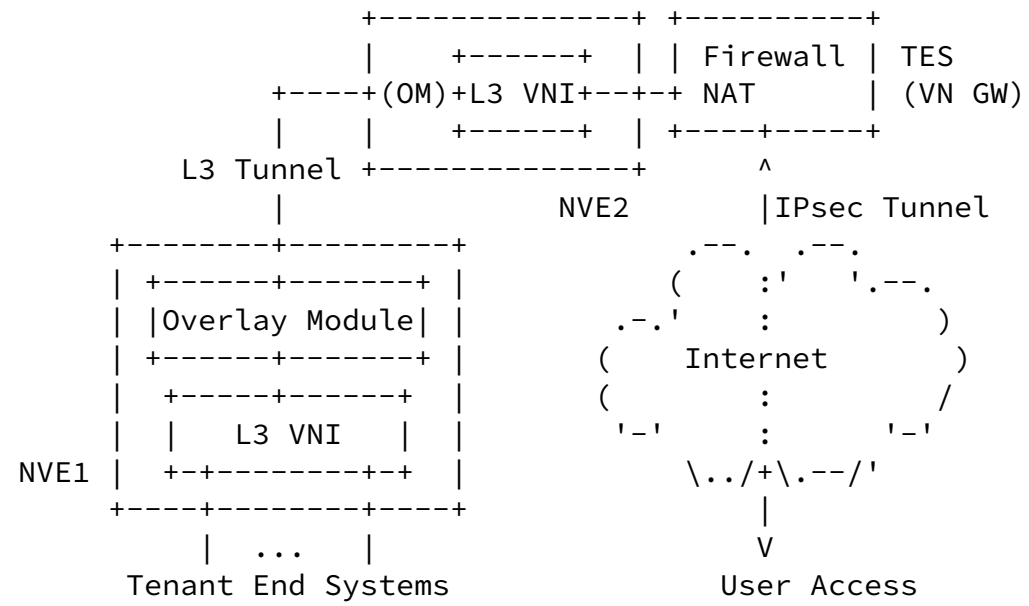
4. Interconnection between DC Virtual Network and External Users

In this scenario, the customers (an enterprise or individuals) utilize 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 or Internet. Three cases are described here.

4.1. DC Virtual Network Access via Internet

A user or an enterprise customer may want to connect to a DC virtual network via Internet but securely. Figure 2 illustrates this case. An L3 virtual network is configured on NVE1 and NVE2 and two NVEs are connected via an L3 tunnel in the Data Center. A set of tenant end systems attach to NVE1. The NVE2 connects to one (may be more) TES that runs the VN gateway and NAT applications (known as net service appliance). A user or customer can access the VN via

Internet by using IPsec tunnel.[RFC4301] The encrypted tunnel is established between the VN GW and the user machine or CPE at enterprise location. The VN GW provides authentication scheme and encryption. Note that VN GW function may be performed by a net service appliance or on a DC GW.



OM: Overlay Module;

Figure 2 DC Virtual Network Access via Internet

4.2. One Virtual Network Method for DC Connectivity

If both the 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 BGP/MPLS VPN solutions [RFC4364] and GRE Tunnel, the NVE in DC and the PE in WAN can be members of one VN instance. Figure 3 illustrates this scenario. The left side of the figure presents an NVE (NVE1) in DC Provider site connecting to tenant end-systems; the right side 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 GRE tunnel (IPsec)[RFC4023] 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 use the same encapsulation semantics; it is just a matter of different terminologies in NVO3 framework [NVO3FRWK] and L3VPN [RFC4364].

The DC and WAN networks may belong to different ASs. Control plane or management plane protocols can facilitate the VN configuration. Routing and forwarding between NVE1 and TES are mentioned in section 3; Routing and forwarding between NVE1 and PE1 and between PE1 and CE in Figure 3 are as specified in RFC4364 [RFC4364]. Note that the draft just uses BGP/MPLS L3VPN as an example for this case, not lead to this specific solution. Trade-off of this solution is described in [NVO3PRBM].

)

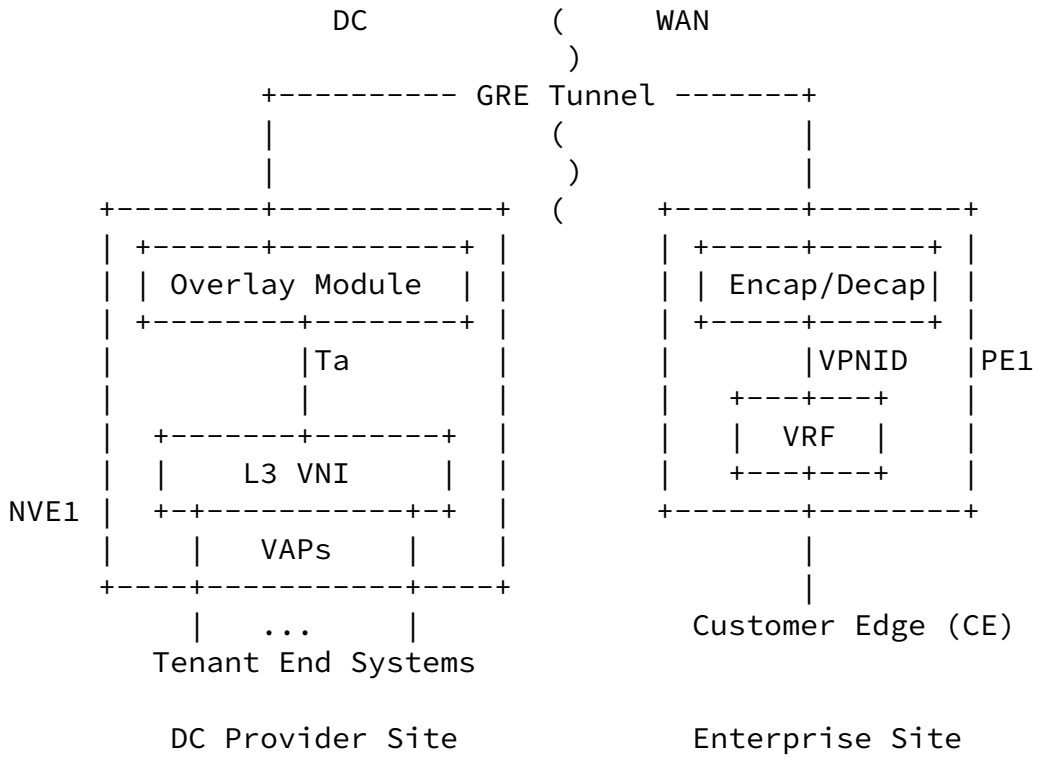


Figure 3 One VN solution across DCs and Carrier Networks

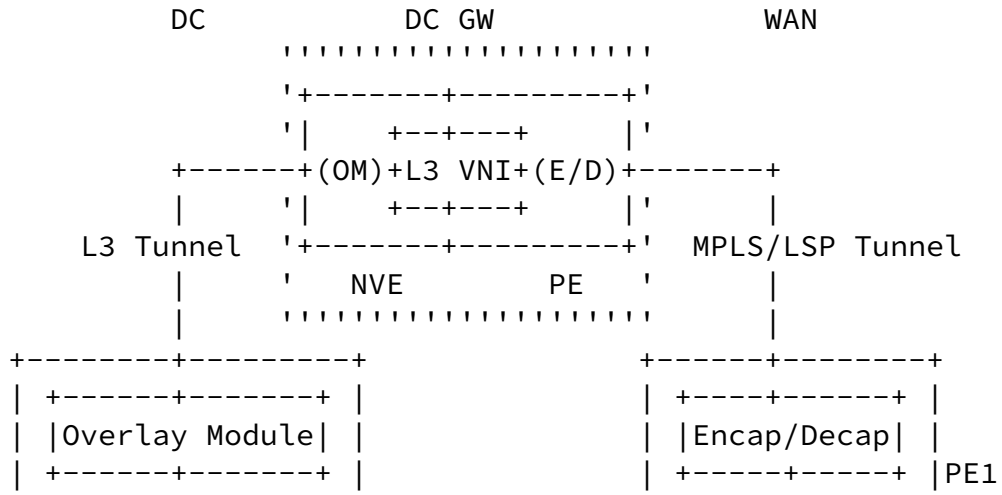
4.3. NVO3 and VPN Interconnection at DC Gateway

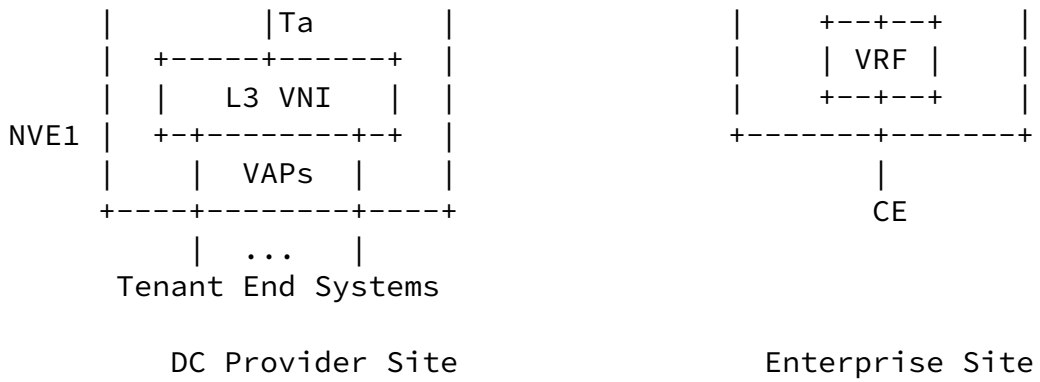
The DC Provider and Carrier may build a tenant VN and VPN for an enterprise customer independently and interconnect the two together at the DC GW. Figure 4 depicts this case. The GW supports both NVE and PE capability. Here an L3 VN instance is built between NVE1 on a server and the NVE2 on DC GW and an L3VPN instance is configured on DC GW and PE1, respectively. The NVE2 on DC GW performs L3 VNI functions, NVO3 encapsulation, and tunneling toward the DC; it also performs L3VPN functions toward the WAN. Both L3 tunnel and LSP Tunnel terminate at the DC GW. The packets are processed at the L3 VNI on DC GW. Operators may choose use of one routing table for both instances as shown in the figure or choose one for each.

This implementation is more complex than the one in section 4.2. However it provides DC network and WAN network demarcation clearly and allows each network use of different VN implementations, which is necessary in many situations. Note that the nvo3 solution can be

simpler than traditional VPNs. Furthermore, two VNs may use different address spaces and let DC GW to perform the address translation.

The alternative of this case is to 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 the L3VPN instance terminates at a PE in the 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, which will be the same as VRF-Lite [VRF-LITE].





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

Figure 4 L3 VNI and L3VPN interconnection across multi networks

If an enterprise only has a few locations, it may use P2P VPWS [RFC4664] or L2TP [RFC5641].

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

5. DC Applications Using NVO3

NVO3 brings DC operators the flexibility to design different applications in a virtual environment without worry about physical network configuration in the Data Center. DC operators may build several virtual networks and interconnect them directly to form a

tenant virtual network; or may allocate some VMs to run tenant applications and some to run net service applications such as Firewall, DNS for the tenant. Several use cases are given in this section.

5.1. Supporting Multi Technologies in a Data Center

Most likely servers deployed in a large data center are rolled in at different times and may have different capacities/features. Some servers may be virtualized, some may not; some may be equipped with virtual switches, some may not. For the ones equipped with hypervisor based virtual switches, some may support VxLAN [VXLAN]

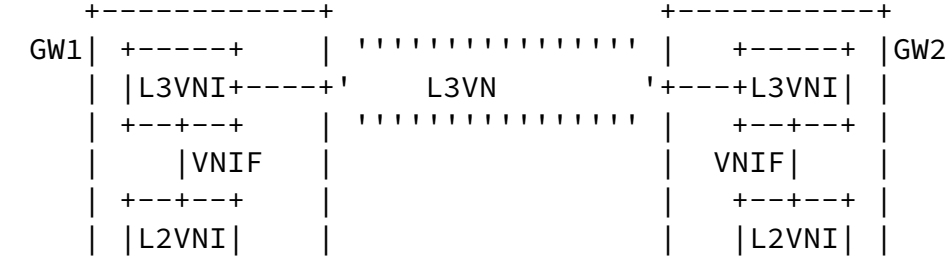
encapsulation, some may support NvGRE encapsulation [NVGRE], and some may not support any types of encapsulation.

To enable the communications among these VMs or servers in a virtual environment, it is necessary to have an entity, either on Gateway or standalone one, to map the services and identifiers and change the packet encapsulation semantics among the Virtual Networks with different encapsulations.

5.2. Tenant Virtual Network with Bridging/Routing

A tenant virtual network may span across multiple Data Centers. DC operator may want to use L2VN within a DC and L3VN outside DCs for a tenant. This is very similar to today's DC physical network configuration. L2 bridging has the simplicity and endpoint awareness while L3 routing has advantages in aggregation and scalability. For this configuration, the virtual gateway function is necessary to interconnect L2VN and L3VN in each DC. Figure 5 illustrates this configuration.

Figure 5 depicts two DC sites. The site A constructs an L2VN that terminates on NVE1, NVE2, and GW1. An L3VN is configured between the GW1 at site A and the GW2 at site Z. An internal Virtual Network Interconnection Interface (VNIF) connects to L2VNI and L3VNI on GW1. Thus the GW1 is the members of the L2VN and L3VN. The L2VNI is the MAC/NVE mapping table and the L3VNI is IP prefix/NVE mapping table. Note that a VNI also has the mapping of TES and VAP at the local NVE. The site Z has the similar configuration. A packet coming to the GW1 from L2VN will be descapulated and converted into an IP packet and then encapsulated and sent to the site Z. The Gateway uses ARP protocol to obtain MAC/IP mapping. Note that both the L2VN and L3VN in the figure are carried by the tunnels supported by the underlying networks which are not shown in the figure.



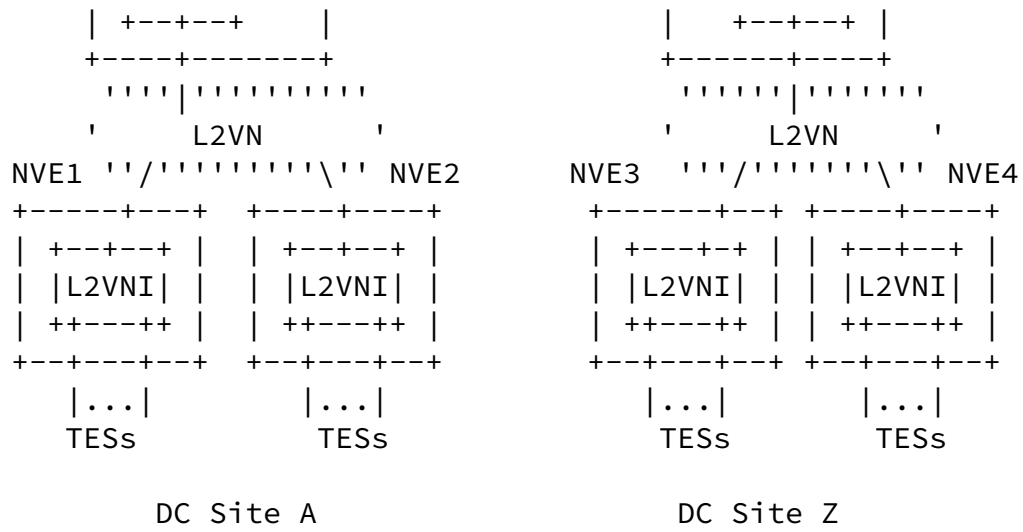


Figure 5 Tenant Virtual Network with Bridging/Routing

5.3. Virtual Data Center

Enterprise DC's today may often use several routers, switches, and service devices to construct its internal network, DMZ, and external network access. A DC Provider may offer a virtual DC to an enterprise customer to run enterprise applications such as website/emails. Instead of using many 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 service applications per customer basis. The service applications may include firewall, gateway, DNS, load balancer, NAT, etc.

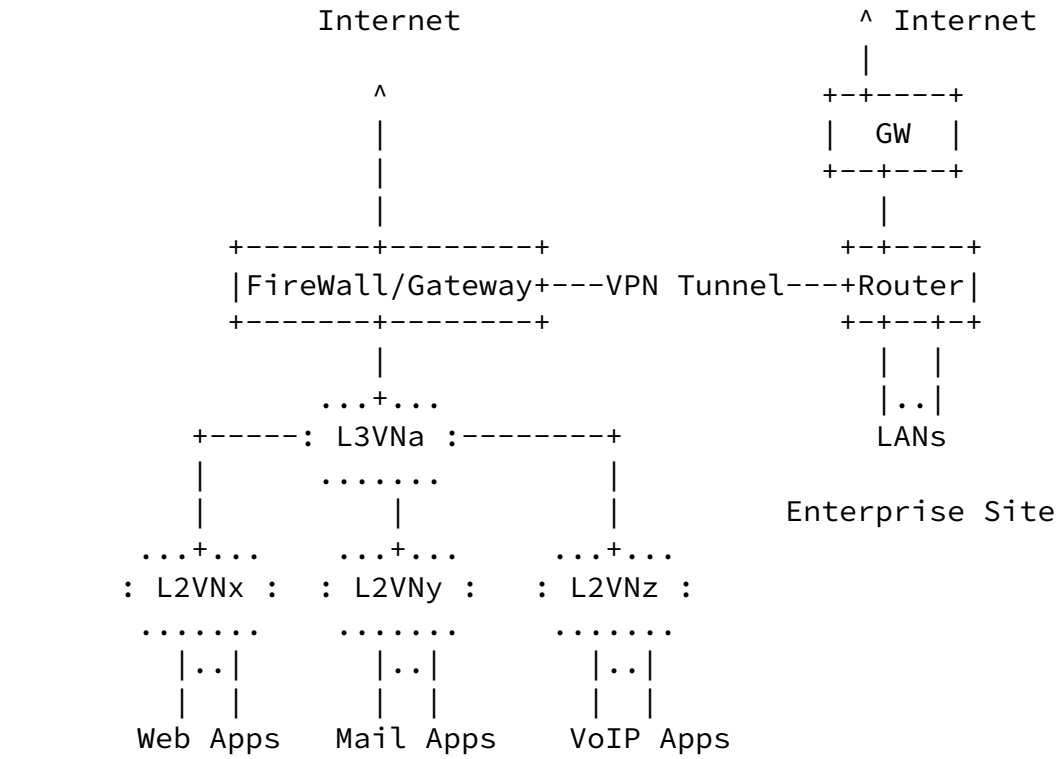
Figure 6 below illustrates this scenario. For the simple illustration, it only shows the L3VN or L2VN as virtual and overlay routers or switches. In this case, DC operators construct several L2 VNs (L2VNx, L2VNy, L2VNz in figure 6) to group the end tenant systems together per application basis, create an L3VNa for the internal routing. A server or VM runs firewall/gateway applications and connects to the L3VNa and Internet. A VPN tunnel is also built between the gateway and enterprise router. The design runs

Enterprise Web/Mail/VoIP applications at the provider DC site; lets the users at Enterprise site to access the applications via the VPN tunnel and Internet via a gateway at the Enterprise site; let Internet users access the applications via the gateway in the provider DC. The enterprise operators can also use the VPN tunnel or

IPsec over Internet to access the vDC for the management purpose. The firewall/gateway provides application-level and packet-level gateway function and/or NAT function.

The Enterprise 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 and gateway function. DC operators may further set different QoS levels for the different applications for a customer.

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



* firewall/gateway may run on a server or VMs

Figure 6 Virtual Data Center 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, RFC5880, 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 describes some basic potential use cases of NVO3. The combination of these cases should give operators flexibility and power to design more sophisticated cases for various purposes.

The main differences between other overlay network technologies and NVO3 is that the client edges of the NVO3 network are individual and virtualized hosts,not network sites or LANs. NVO3 enables these virtual hosts communicating in a true virtual environment without considering physical network configuration.

NVO3 allows individual tenant virtual networks to use their own address space and isolates the space from the network infrastructure. The approach not only segregates the traffic from multi tenants on a common infrastructure but also makes VM placement and move easier.

DC applications are about providing virtual processing/storage, applications, and networking in a secured and virtualized manner, in which the NV03 is just a portion of an application. NVO3 decouples the applications and DC network infrastructure configuration.

NVO3's underlying network provides the tunneling between NVEs so that two NVEs appear as one hop to 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 WANs. It is also possible that intra DC and inter DC tunnels are stitched together to form an end-to-end tunnel between two NVEs.

A DC virtual network may be accessed via an external network in a secure way. Many existing technologies can achieve this.

The key requirements for NVO3 are 1) traffic segregation; 2) supporting a large scale number of virtual networks in a common infrastructure; 3) supporting highly distributed virtual network with sparse memberships 3) VM mobility 4) auto or easy to construct a NVE and its associated TES; 5) Security 6) NVO3 Management [NVO3PRBM].

8. Security Considerations

Security is a concern. DC operators need to provide a tenant a secured virtual network, which means the tenant traffic isolated from other tenant's and non-tenant VMs not placed into the tenant virtual network; they also need to prevent DC underlying network from any tenant application attacking through the tenant virtual network or one tenant application attacking another tenant application via DC networks. For example, a tenant application attempts to generate a large volume of traffic to overload DC underlying network. The NVO3 solution has to address these issues.

9. IANA Considerations

This document does not request any action from IANA.

10. Acknowledgements

Authors like to thank Sue Hares, Young Lee, David Black, Pedro Marques, Mike McBride, David McDysan, and Randy Bush for the review, comments, 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] Allan, D., Nadeau, T., "A Framework for Multi-Protocol Label Switching (MPLS) Operations and Management (OAM)", rfc4378, February 2006

[RFC4023] Worster, T., etc, "Encapsulating MPLS in IP or Generic Routing Encapsulation (GRE)", rfc4023, March 2005

[RFC4301] Kent, S., "Security Architecture for the Internet Protocol", rfc4301, December 2005

[RFC4664] Andersson, L., "Framework for Layer 2 Virtual Private Networks (L2VPNs)", rfc4664, September 2006

[RFC5641] McGill, N., "Layer 2 Tunneling Protocol Version 3 (L2TPv3) Extended Circuit Status Values", rfc5641, April 2009.

[RFC5880] Katz, D. and Ward, D., "Bidirectional Forwarding Detection (BFD)", rfc5880, June 2010.

11.2. Informative References

[ESYS] Marques,P., "End-System support for BGP-signaled IP/VPNs", draft-marques-l3vpn-end-system-07, August 2012

[NVGRE] Sridharan, M., "NVGRE: Network Virtualization using Generic Routing Encapsulation", draft-sridharan-virtualization-nvgre-01, July 2012

[NVO3PRBM] Narten, T., etc "Problem Statement: Overlays for Network Virtualization", draft-narten-nvo3-overlay-problem-statement-04, August 2012

[NVO3FRWK] Lasserre, M., Motin, T., and etc, "Framework for DC Network Virtualization", draft-lasserre-nvo3-framework-03, July 2012

[VDP] "IEEE P802.1Qbg Edge Virtual Bridging".

[VMWARE] VMware, "vCenter", http://www.vmware.com

[VRF-LITE] Cisco, "Configuring VRF-lite", http://www.cisco.com

[VXLAN] Mahalingam,M., Dutt, D., etc "VXLAN: A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", draft-mahalingam-dutt-dcops-vxlan-02.txt, August 2012

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

Linda Dunbar
Huawei Technologies,
4320 Legacy Dr.
Plano, Tx75025 US

Phone: +1-469-277-5840
Email: linda.dunbar@huawei.com