Network working group
Internet Draft
Category: Informational

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

Expires: April 2013 October 22, 2012

Use Cases for DC Network Virtualization Overlays

draft-mity-nvo3-use-case-04

Status of this Memo

Copyright Notice

Abstract

This draft describes the general 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 the problems in today's data center hinder the support of an elastic cloud service and dynamic virtual tenant networks [NVO3PRBM]. The goal of DC Network Virtualization Overlays, i.e. NVO3, is to decouple a communication among tenant end systems (VMs) from DC physical networks and to allow the network infrastructure to provide: 1) traffic isolation among one virtual network and another; 2) independent address space in each virtual network and address isolation from the infrastructure's; 3) Flexible VM placement and move from one server to another without any physical network limitation. These characteristics will help address the issues in the data centers.

Although NVO3 may enable a true virtual environment where VMs and net service appliances communicate, the NVO3 solution has to address how to communicate between a virtual network and a physical network. This is because 1) many traditional DCs 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 the virtualization capability.

This document is to describe general NVO3 use cases that apply to various data center networks to ensure nvo3 framework and solutions can meet the demands. Three types of the use cases are:

- A virtual network connects many tenant end systems within a 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.

- A DC provider offers 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 an IPsec tunnel over Internet connecting to an overlay virtual network offered by a Data Center provider. This is mainly constitutes DC North-South traffic.

- A DC provider uses NVO3 to design a variety of DC applications that make use of the net service appliance, virtual compute, storage, and networking. In this case, the 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.

CUG: Closed User Group

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

VNIF: Internal Virtual Network Interconnection Interface

3. Basic Virtual Networks in a Data Center

A virtual network may exist within a DC. The network enables a communication among tenant end systems (TESs) that are in a Closed User Group (CUG). A TES may be a physical server or virtual machine (VM) on a server. A virtual network has a unique virtual network identifier (may be local or global unique) for switches/routers to properly differentiate it from other virtual networks. The CUGs are formed so that proper policies can be applied when the TESs in one CUG communicate with the TESs in other CUGs.

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 member of one or more virtual networks. Each virtual network may be L2 or L3 basis. In this illustration, three virtual networks with VN context Ta, Tn, and Tm are shown. The VN 'Ta' terminates on both NVE1 and NVE2; The VN 'Tn' terminates on NVE1 and the VN 'Tm' at NVE2 only. If an NVE is a member of a VN, one or more virtual network instances (VNI) (i.e. routing and forwarding table) exist on the NVE. 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.

A TES attaches to a virtual network (VN) via a virtual access point (VAP) on an NVE. One TES may participate in one or more virtual networks via VAPs; one NVE may be configured with multiple VAPs for a VN. Furthermore if individual virtual networks use different address spaces, the TES participating in all of them will be configured with multiple addresses as well. A TES as a gateway is an example for this. In addition, multiple TESes may use one VAP to attach to a VN. For example, VMs are on a server and NVE is on ToR, some VMs may attach to NVE via one VLAN.

A VNI on an NVE is a routing and forwarding table that caches and/or maintains the mapping of a tenant end system and its attached NVE. The table entry may be updated by the control plane or data plane or management plane. It is possible that an NVE has more than one VNIs associated with a VN.

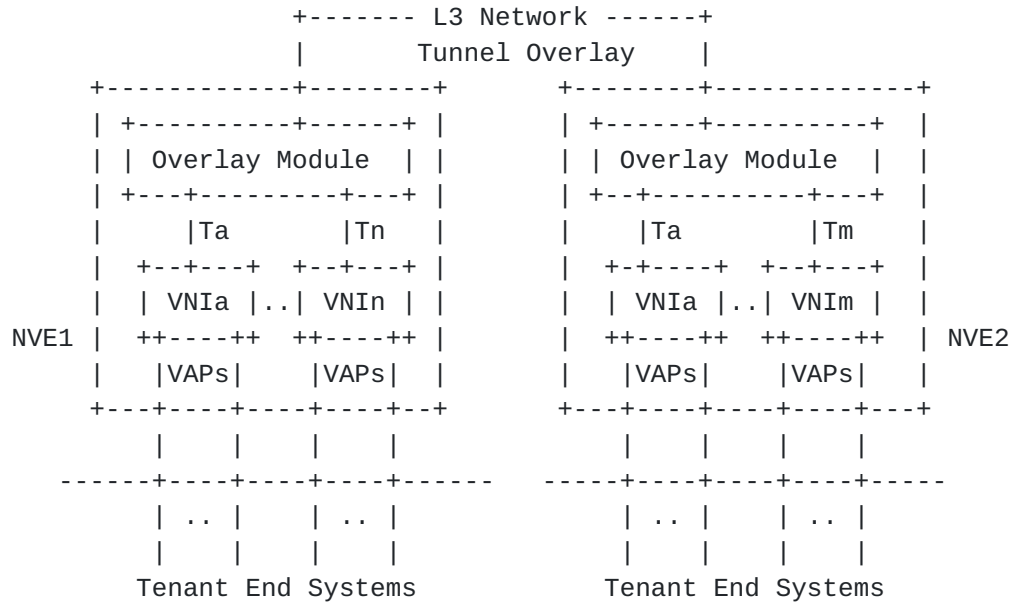


Figure 1 NVo3 for Tenant End-System interconnection

One virtual network may have many NVE members and interconnect several thousands of TESs (as a matter of policy), 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.

It is worth to mention two distinct cases here. The first is when TES and NVE are co-located on a same physical device, which means that the NVE is aware of the TES state at any time via internal API. The second is when TES and NVE are remotely connected, i.e. connected via a switched network or point-to-point link. In this case, a protocol is necessary for NVE to know TES state.

Note that if all NVEs are co-located with TESes in a CUG, the communication in the CUG is in a true virtual environment. If a TES connects to a NVE remotely, the communication from this TES to other TESes in the CUG is not in a true virtual environment. The packets to/from this TES are exposed to a physical network directly, i.e. on a wire.

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.

If a 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 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. Interconnecting DC Virtual Network and External Networks

For customers (an enterprise or individuals) who want to utilize the DC provider's compute and storage resources to run their applications, they need to access those end systems hosted in a DC through Carrier WANs or Internet. A DC provider may want to use an NVO3 virtual network to connect these end systems; then it, in turn, becomes the case of interconnecting DC virtual network and external networks. Two 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.

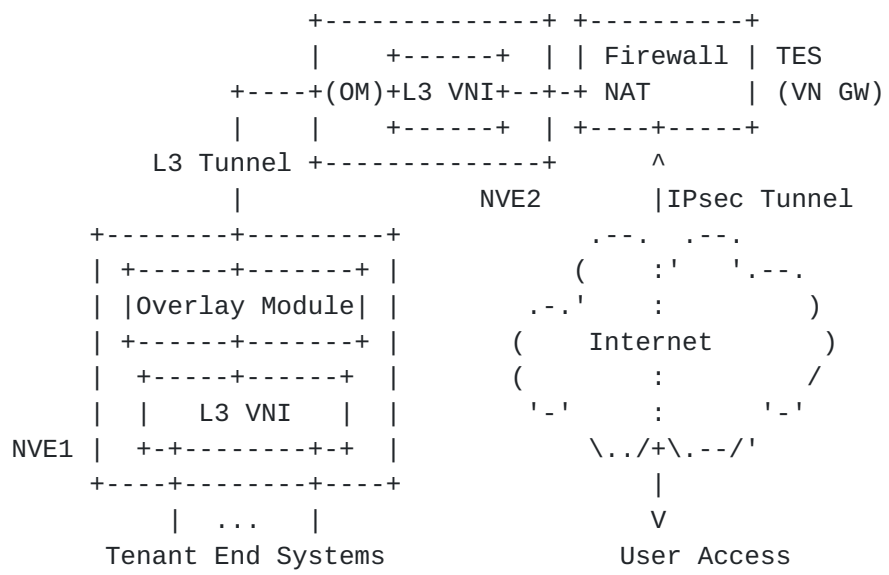


Figure 2 DC Virtual Network Access via Internet

4.2. DC Virtual Network and WAN VPN Interconnection

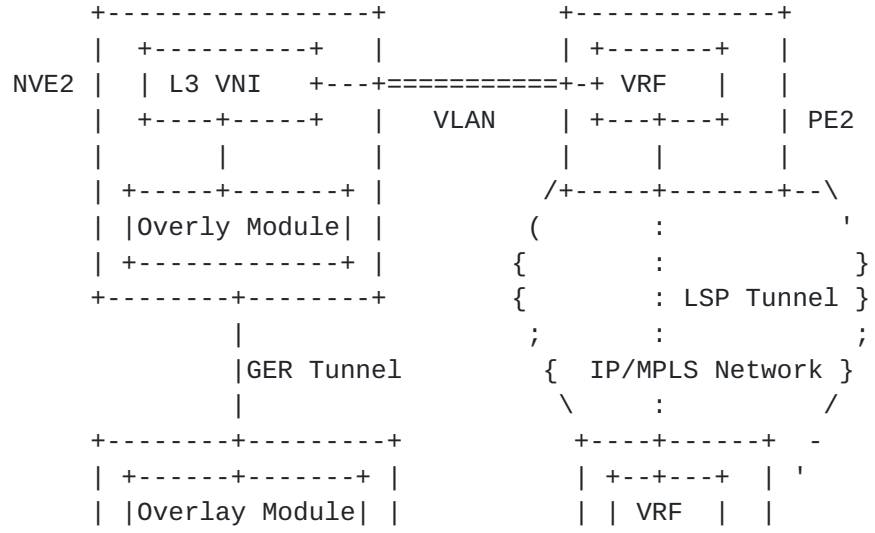
A DC Provider and Carrier may build a VN and VPN independently and interconnect the two at the DC GW and PE for an enterprise customer. Figure 3 depicts this case in a L3 overlay (L2 overlay is the same). The DC provider constructs an L3 VN between the NVE1 on a server and the NVE2 on the DC GW in the DC site; the carrier constructs an L3VPN between PE1 and PE2 in its IP/MPLS network. An Ethernet Interface physically connects the DC GW and PE2 devices. The local VLAN over the Ethernet interface [VRF-LITE] is configured to connect the L3VNI/NVE2 and VRF, which makes the interconnection between the

L3 VN in the DC and the L3VPN in IP/MPLS network. An Ethernet Interface may be used between PE1 and CE to connect the L3VPN and enterprise physical networks.

This configuration allows the enterprise networks communicating to the L3 VN as if its own networks but not communicating with DC provider underlying physical networks as well as not other overlay networks in the DC. The enterprise may use its own address space on the L3 VN. The DC provider can manage the VM and storage assignment to the L3 VN for the enterprise customer. The enterprise customer can determine and run their applications on the VMs. From the L3 VN perspective, an end point in the enterprise location appears as the end point associating to the NVE2. The NVE2 on the DC GW has to perform both the GRE tunnel termination [RFC4797] and the local VLAN termination and forward the packets in between. The DC provider and Carrier negotiate the local VLAN ID used on the Ethernet interface.

This configuration makes the L3VPN over the WANs only has the reachbility to the TES in the L3 VN. It does not have the reachability of DC physical networks and other VNs in the DC. However, the L3VPN has the reachbility of enterprise networks. Note that both the DC provider and enterprise may have multiple network locations connecting to the L3VPN.

The eBGP protocol can be used between DC GW and PE2 for the route population in between. In fact, this is like the Option A in [RFC4364]. This configuration can work with any NVO3 solution. The eBGP, OSPF, or other can be used between PE1 and CE for the route population.



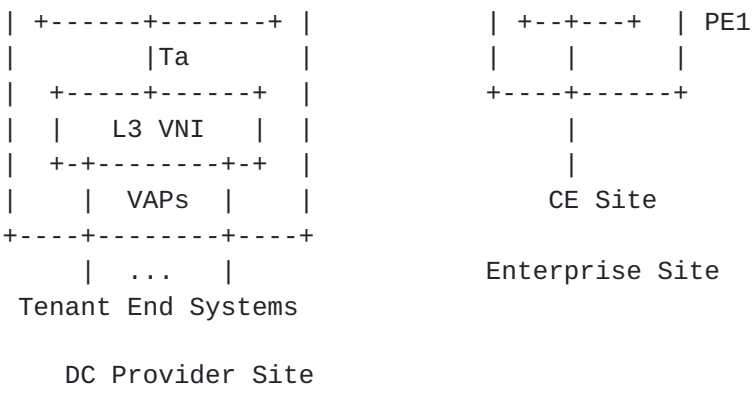


Figure 3 L3 VNI and L3VPN interconnection across multi networks

If an enterprise only has one location, it may use P2P VPWS [RFC4664] or L2TP [RFC5641] to connect one DC provider site. In this case, one edge connects to a physical network and another edge connects to an overlay network.

The interesting feature in this use case is that the L3 VN and compute resource are managed by the DC provider. The DC operator can place them at any location without notifying the enterprise and carrier because the DC physical network is completely isolated from the carrier and enterprise network. Furthermore, the DC operator may move the compute resources assigned to the enterprise from one sever to another in the DC without the enterprise customer awareness, i.e. no impact on the enterprise 'live' applications running these resources. Such advanced feature brings some requirements for NVO3 [NVO3PRBM].

5. DC Applications Using NVO3

NVO3 brings DC operators the flexibility to design different applications in a true 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 and implement the communication rules through policy; 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 construct a tenant virtual network among these servers and the ToRs, it may use two virtual networks and a gateway to allow different implementations working together. For example, one virtual network uses VxLAN encapsulation and another virtual network uses traditional VLAN.

The gateway entity, either on VMs or standalone one, participates in to both virtual networks, and maps the services and identifiers and changes the packet 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.

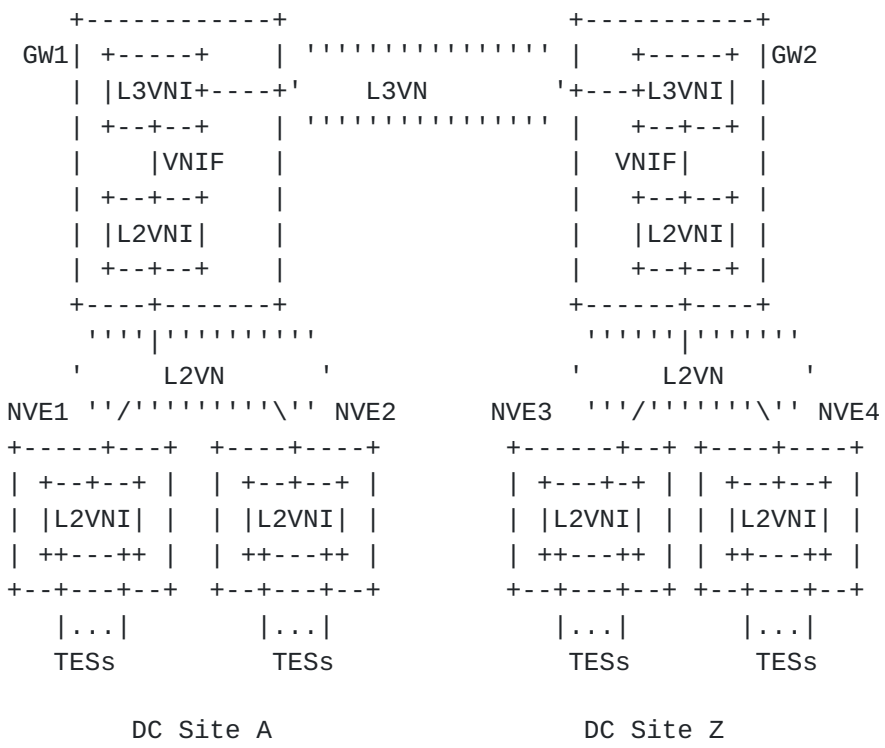


Figure 4 Tenant Virtual Network with Bridging/Routing

5.3. Virtual Data Center (VDC)

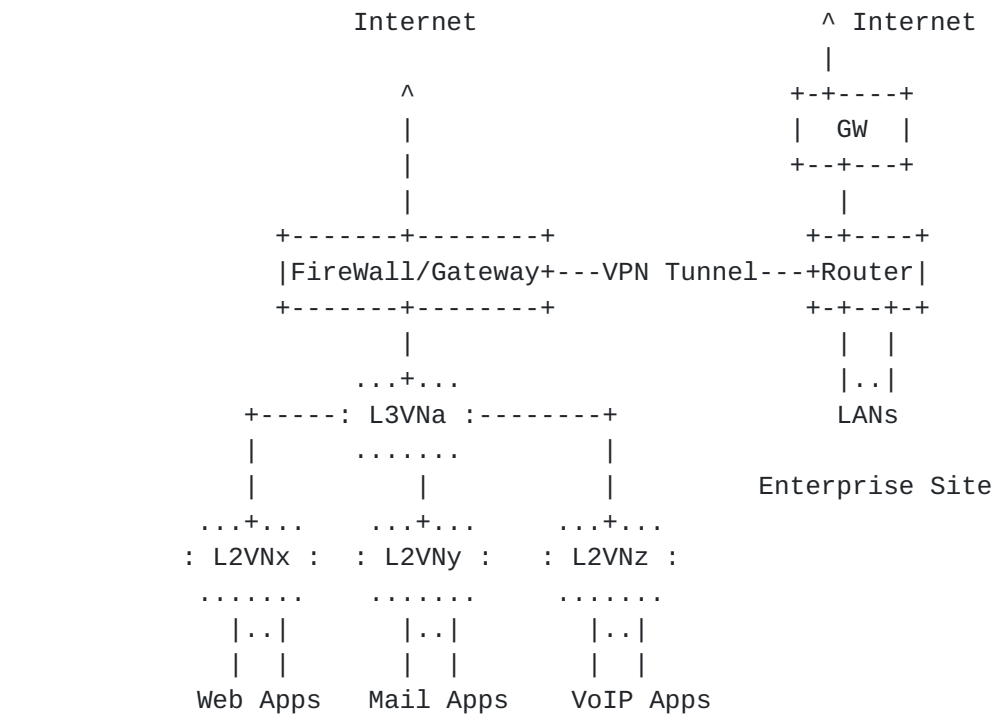
Enterprise DC's today may often use several routers, switches, and service devices to construct its internal network, DMZ, and external network access. A DC Provider may offer a virtual DC to an enterprise customer to run enterprise applications such as website/emails. Instead of using many hardware 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 5 Virtual Data Center by Using NVO3

5.4. Federating NV03 Domains

Two general cases are 1) Federating AS managed by a single operator; 2) Federating AS managed by different Operators. The detail will be described in next version.

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

[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

[RFC4797] Rekhter, Y., etc, "Use of Provider Edge to Provider Edge (PE-PE) Generic Routing Encapsulation (GRE) or IP in BGP/MPLS IP Virtual Private Networks", RFC4797, January 2007

[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

[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-ietf-nvo3-overlay-problem-statement-00, September 2012

[NVO3FRWK] Lasserre, M., Motin, T., and etc, "Framework for DC Network Virtualization", draft-ietf-nvo3-framework-01, October 2012

[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