

Network working group
Internet Draft
Category: Informational

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

Expires: December 2012

July 16, 2012

Use Cases for DC Network Virtualization Overlays

[draft-mity-nvo3-use-case-01](#)

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on December, 2012.

Copyright Notice

Copyright (c) 2009 IETF Trust and the persons identified as the document authors. All rights reserved.

Internet-Draft

NV03 Use Case

June 2012

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Abstract

This draft describes NV03 use cases. The work intention is to help validate the NV03 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.....	3
2.	Terminology.....	4
3.	Virtual Network in One Data Center.....	4
4.	Interconnection between DC Virtual Network and External Users..	6
4.1.	One Virtual Network Method for DC Connectivity.....	6
4.2.	NV03 and VPN Interconnection at DC Gateway.....	7
4.3.	Connecting a DC Virtual Network via Internet.....	9
5.	DC Applications Using NV03.....	10
5.1.	Tenant Network with Bridging/Routing and Internet Access.	10
5.2.	Virtual Data Center.....	11
6.	OAM Considerations.....	13
7.	Summary.....	13
8.	Security Considerations.....	14
9.	IANA Considerations.....	14
10.	Acknowledgements.....	14
11.	References.....	14
11.1.	Normative References.....	14
11.2.	Informative References.....	15
	Authors' Addresses.....	15

1. Introduction

This document provides the use cases for Network Virtualization Overlay, i.e. NV03, which is driven by Data Center Networks. These use cases are intended to help validate the framework and requirements as along with the development of the solutions.

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 NV03 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 Data Center (DC) environment. The main differences between other overlay network technologies and NV03 is that the client edges, of the NV03 network, are individual virtualized hosts and not network sites, and the hosts and the network edge may be on the same physical device. Other differentiating characteristics may include (1) virtual host access and mobility which causes association between hosts to NV03 edge nodes to be non-fixed (2) Less chance for loop among VMs attached to NV03 edge due to simple topology.

NV03 use cases can be highly varied. This document outlines some basic scenarios and groups them into three sets.

One set of use cases is to connect many tenant end systems in one Data Center and form an L2 or L3 communication domain. Overlay tenant virtual networks segregate tenant traffic and allow individual tenants having its own address space and isolating the space from DC infrastructure. In addition, they allow VM moves from one server to another.

The second set of NV03 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 VPN provided by a carrier or IPsec over Internet connecting to an overlay virtual network offered by a Data Center provider.

The third set of NV03 use cases is to enable the designs of various DC applications using the service applications, compute, storage, and networking. In this case, NV03 provides the networking functions for the applications.

The document uses the reference model and terminologies defined in [NV03FRWK] to describe the use cases.

[2.](#) Terminology

This document uses the terminologies defined in [[NV03FRWK](#)], [[RFC4364](#)]. Some additional terms used in the document are listed here.

VNIF: VNI Interconnection Interface on an NVE

L2 VNI: L2 Virtual Network Instance

L3 VNI: L3 Virtual Network Instance

ARP: Address Resolution Protocol

DNS: Domain Name Service

DMZ: DeMilitarized Zone

NAT: Network Address Translation

[3.](#) Virtual Network in One Data Center

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

Figure 1 depicts this case. NVE1 and NVE2 are two network virtual edges that may exist on a server or ToR. Each NVE may be 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. 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 VNI terminated on an NVE may locally associate to one or more VAPs each of which may associate with one or more TESSs. 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 TESSs, the ability of supporting a large number of TESSs per tenant instance and TES mobility is critical for NV03 solution.

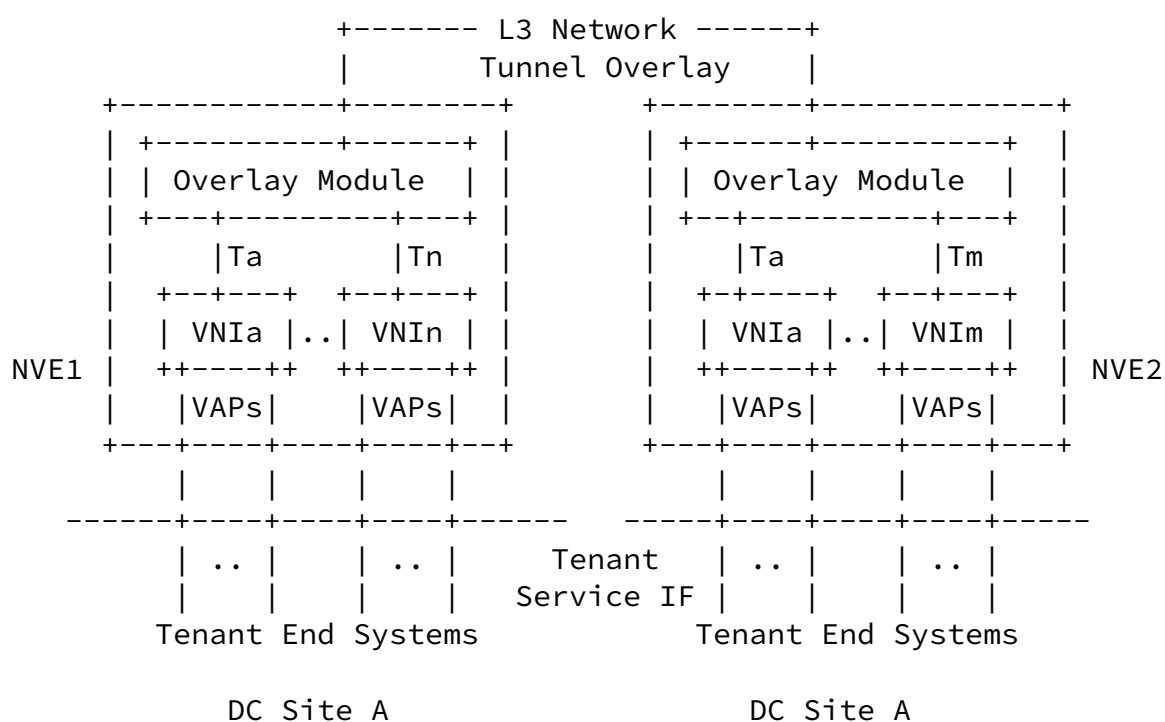


Figure 1 NV03 for Tenant End-System interconnection

Individual virtual network instances 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: the NV03 solutions still have to address VM move in overlay network.

When a DC operator creates a VM on a server, he/she has a plan which VN the VM belongs to and assigns the VM to the VN via an

administration system such as vCenter. When a VM is alive/off, i.e. power-on/off, or relocated to another server, its associated NVE should be notified. NV03 solution is necessary to support these features. [\[TESNVE\]](#) [\[SV2NVE\]](#)

If a tenant virtual network spans across multiple DC sites, one design is to allow the corresponding NV03 instance seamlessly span across those sites without DC gateway routers' termination. In this case, the tunnel 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.](#) 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 MPLS-based VPN solutions [\[RFC4364\]](#) and GRE 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 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 perform the same functions; it is just a matter of using different terminologies in NV03 framework [[NV03FRWK](#)] and L3VPN [[RFC4364](#)].

The DC and WAN networks may belong to different ASs. Control plane or management plane protocols can facilitate the VN configuration. Note: If an NVE is on a server and TESSs are VMs on the server, it is no need any routing protocol between NVE and TESSs; TES-NVE association is configured by DC operators. When a VM is "power-on", the NVE populates it in the forwarding table; When the VM is "power-off", the NVE removes it from the table. The forwarding between the NVE and TESSs is simply an internal table lookup and delivery process on the server. If an NVE is on ToR, TESSs may be either non-virtualized servers or VMs on virtualized servers. For the latter the routing between NVE and TESSs may use Petro's proposal [[ESYS](#)] or a routing protocol such as OSPF per VN. The forwarding between two

is like CE-PE's. Routing and forwarding between NVE1 and PE1 and between PE1 and CE in Figure 2 are as specified in [RFC4364](#) [[RFC4364](#)].

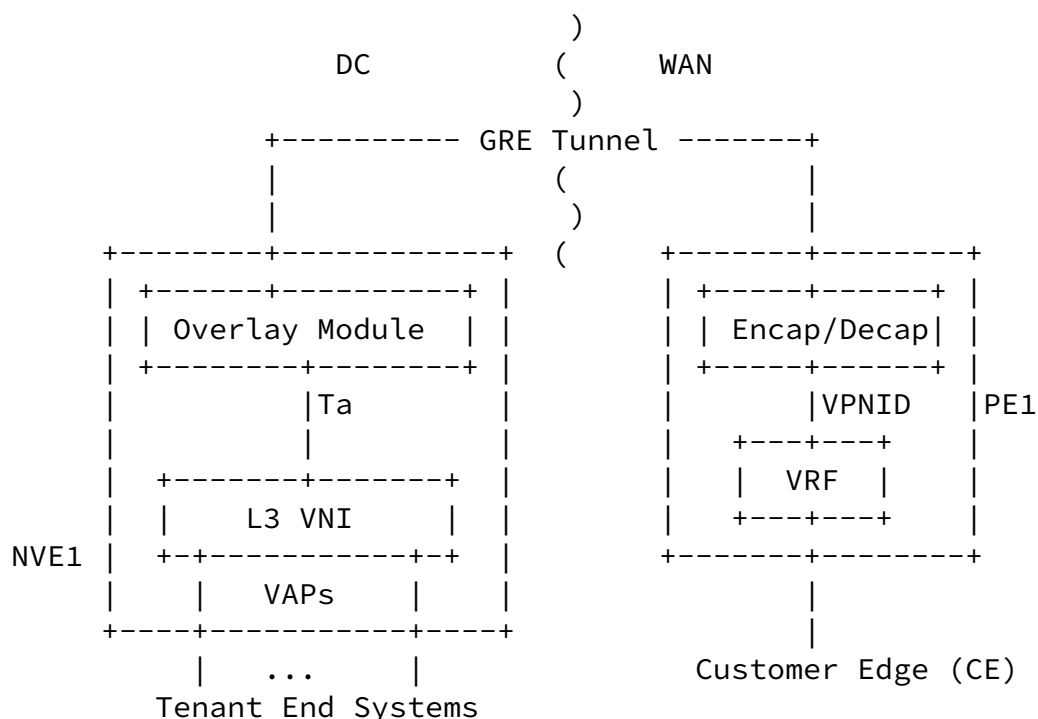


Figure 2 One VN solution across DCs and Carrier Networks

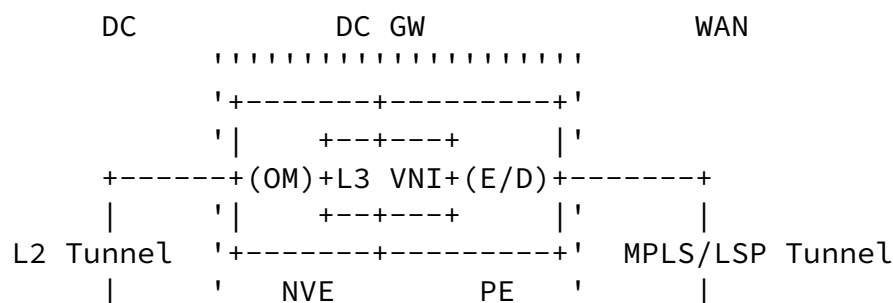
4.2. NV03 and VPN Interconnection at DC Gateway

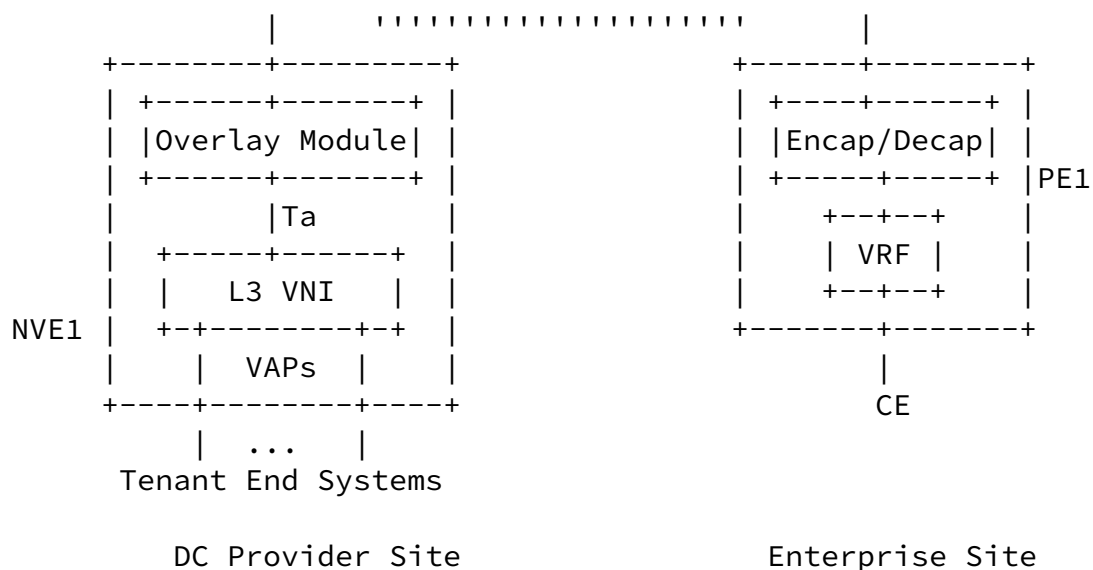
The DC Provider and Carrier may build a tenant VN and VPN for an enterprise customer independently and interconnect the two together at the DC GW. Figure 3 depicts this case. The GW supports both NVE and PE capability. Here an L3 VN instance is built between NVE1 and the NVE on DC GW and an L3VPN instance is configured on the PE of DC GW and PE1, respectively. The NVE on DC GW performs L3 VNI functions, NV03 encapsulation, and tunneling toward the DC; it also performs L3VPN functions toward the WAN. Both L2 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 they can choose one for each.

This implementation is more complex than the 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. Note: the nvo3 solution can be

simpler than L3VPN [[RFC4364](#)] due to TES and NVE functionality. 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](#)]





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

Figure 3 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 that provides the connectivity to the VMs belonging to the same tenant.

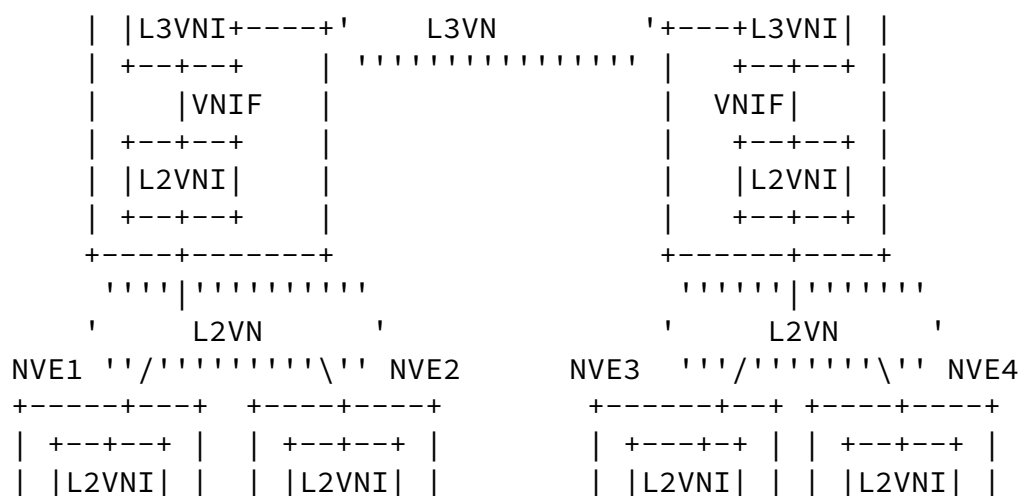
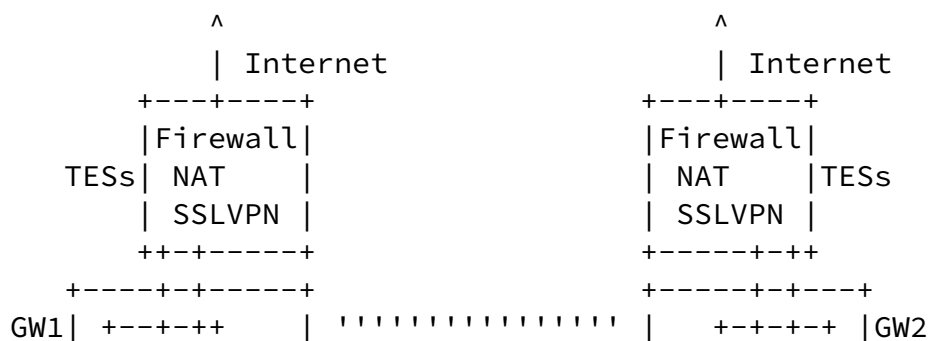
[4.3](#). Connecting a DC Virtual Network via Internet

A user may want to connect to a DC virtual network via Internet but securely. Figure 4 illustrates this case. A L3 virtual network is configured on NVE1 and NVE2 and two NVEs are connected via a L2 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. A user can access the VN via Internet by using IPsec.[[RFC4301](#)] The encrypted tunnel is used between the VN GW and the user. The VN GW provides authentication scheme and encryption.

Figure 4 DC Virtual Network Access via Internet

internally are interconnected on GW1 via Virtual Network Interconnection Interface (VNIF). The site Z has the similar configuration. 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.

This configuration provides a private cloud network in/across Data Center site A and Z and consists of three virtual networks. Within each Data Center, the L2VN provides the L2 connectivity to all the associated TESSs and the GW. The GW1 or GW2 terminates the L2VN traffic and forwards the packets as IP packets to remote DC, which forms a private cloud network among all the TESSs. The GW1 or GW2 also forwards/receives the IP packets from TESSs running firewall/NAT/SSLVPN; the TESSs connect to Internet via DC underlying network. This lets the cloud network connecting to Internet in a secure way. DC operator can choose an address space for the cloud network and rely on the NAT application to perform address translation. This configuration allows a VM move within the L2VN but not across DCs due to different IP subnet on each GW.



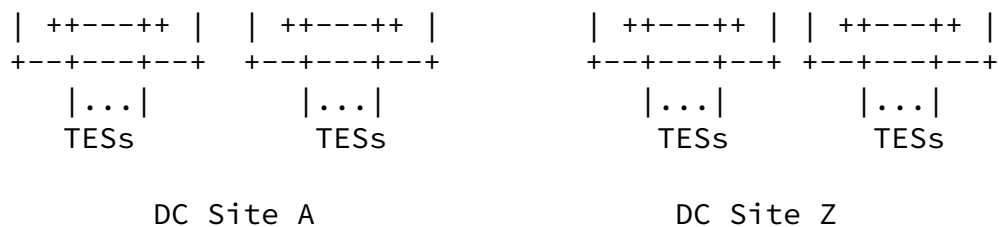


Figure 5 Tenant Network with Bridging/Routing and Internet Access

5.2. 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 NV03, 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.

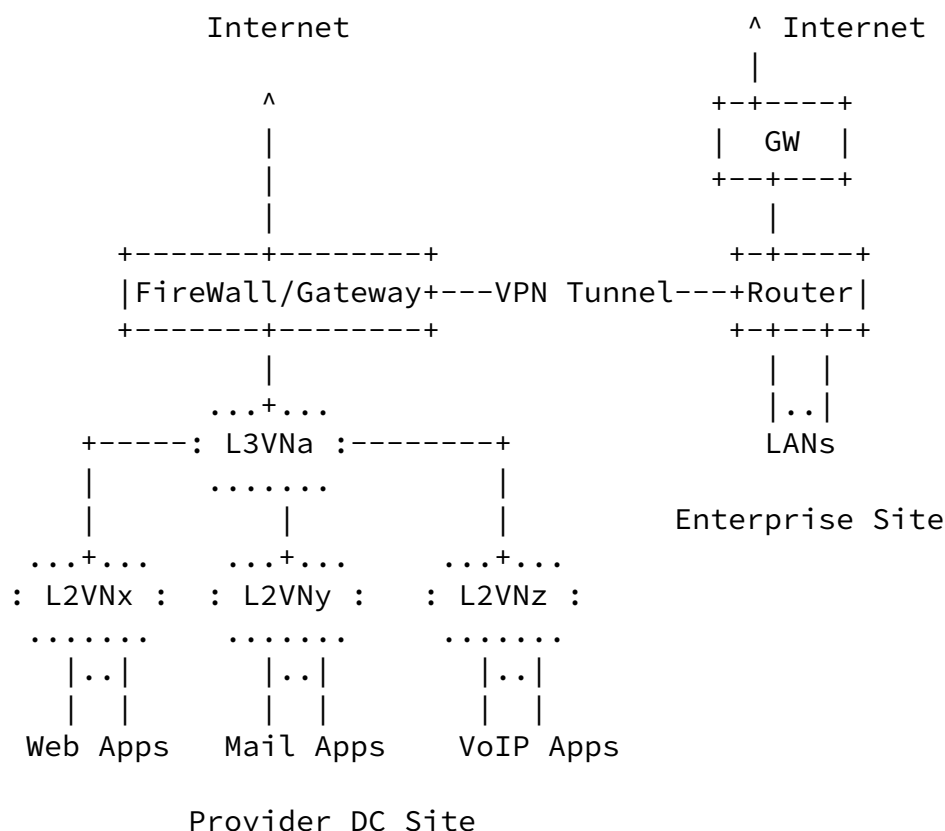
Figure 6 below illustrates this scenario. For the simple illustration, it only shows the L3VN or L2VN as virtual and overlay routers or switches. In this case, DC operators construct several L2 VNs (L2VNx, L2VNy, L2VNz in figure 6) to group the end tenant systems together per application basis, create 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 NV03 solution to provide the DC operator an easy way to create NVEs and VNIs for any design and to quickly assign TESSs 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 NV03

6. OAM Considerations

NV03 brings the ability for a DC provider to segregate tenant traffic. A DC provider needs to manage and maintain NV03 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 differences between other overlay network technologies and NVo3 is that the client edges of the NVo3 network are individual and virtualized hosts and not network sites or LANs. NVo3 no longer treats 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 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 dynamic placement easier.

DC applications are about providing virtual processing/storage, applications, and networking in a secured and virtualized manner, in which the NVo3 is just a portion of an application. NVo3 decouples the applications and DC network infrastructure.

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 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.

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 NV03 are 1) traffic segregation; 2) support a large scale number of TEs in a virtual network; 3) VM mobility 4) auto or easy to construct a NVE and its associated TE; 5) Security 6) NV03 Management.

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 NV03 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, and Mike McBride 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] 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.

11.2. Informative References

- [NV03FRWK] Lasserre, M., Motin, T., and etc, "Framework for DC Network Virtualization", [draft-lasserre-NV03-framework-02](#), June 2012.
- [ESYS] Marques, "End-System support for BGP-signaled IP/VPNs", [draft-marques-l3vpn-end-system-02](#), October 2011
- [NVGRE] Sridharan, M., "NVGRE: Network Virtualization using Generic Routing Encapsulation", [draft-sridharan-virtualization-nvgre-00](#), September 2011
- [SV2NVE] Kompella, K., Rekhter, Y., etc "Using Signaling to Simplify Network Virtualization Provisioning", [draft-kompella-nvo3-server2nve-00](#), July 2012
- [TESNVE] Gu., Y. "The mechanism and protocol between VAP and TES to facilitate NV03", July 2012
- [VRF-LITE] Cisco, "Configuring VRF-lite", <http://www.cisco.com>

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

