Network Working Group
Internet Draft
Category: Informational

L. Yong
L. Dunbar
Huawei
M. Toy
Verizon
A. Isaac
Juniper Networks
V. Manral
Ionos Networks

Expires: June 2017 December 8, 2016

Use Cases for Data Center Network Virtualization Overlay Networks

draft-ietf-nvo3-use-case-14

Abstract

This document describes data center network virtualization overlay (NVO3) network use cases that can be deployed in various data centers and serve different data center applications.

Table of Contents

1. Introduction

Server virtualization has changed the Information Technology (IT) industry in terms of the efficiency, cost, and speed of providing new applications and/or services such as cloud applications. However traditional data center (DC) networks have some limits in supporting cloud applications and multi tenant networks [RFC7364]. The goal of data center network virtualization overlay (NVO3) networks is to decouple the communication among tenant systems from DC physical infrastructure networks and to allow one physical network infrastructure:

- Carry many NVO3 networks and isolate different NVO3 network traffic on a physical network that carries NVO3 network traffic.

- Independent address spaces in individual NVO3 networks such as MAC, IP, TCP/UDP etc.

- Flexible Virtual Machines (VM) and/or workload placement including the ability to move them from one server to another without requiring VM address changes and physical infrastructure network configuration changes, and the ability to perform a "hot move" with no disruption to the live application running on VMs.

These characteristics of NVO3 networks help address the issues that cloud applications face in data centers [RFC7364].

An NVO3 network may interconnect with another NVO3 network on the same physical network, or another physical network (i.e., not the physical network that the NVO3 network is carried over), via a gateway. The use case examples for the latter are: 1) DCs that migrate toward an NVO3 solution will be done in steps, where a portion of tenant systems in a VN is on virtualized servers while others exist on a LAN. 2) many DC applications serve to Internet users who are on physical networks; 3) some applications are CPU bound, such as Big Data analytics, and may not run on virtualized resources. Some inter-VN policies can be enforced at the gateway.

This document describes general NVO3 network use cases that apply to various data centers. The use cases described here represent DC provider's interests and vision for their cloud services. The document groups the use cases into three categories from simple to advance in term of implementation. However the implementations of these use cases are outside the scope of this document. These three categories are highlighted below:

- Basic NVO3 networks (Section 2). All Tenant Systems (TS) in the network are located within the same DC. The individual networks can be either Layer 2 (L2) or Layer 3 (L3). The number of NVO3 networks in a DC is much higher than what traditional VLAN based virtual networks [IEEE 802.1Q] can support. This case is often referred as to the DC East-West traffic.

- A virtual network that spans across multiple Data Centers and/or to customer premises where NVO3 networks are constructed and interconnect another virtual or physical network outside the data center. An enterprise customer may use a traditional carrier VPN or an IPsec tunnel over the Internet to communicate with its systems in the DC. This is described in Section 3.

- DC applications or services require an advanced network that contains several NVO3 networks that are interconnected by the gateways. Three scenarios are described in Section 4: 1) supporting multiple technologies; 2) constructing several virtual networks as a tenant network; 3) applying NVO3 to a virtual Data Center (vDC).

The document uses the architecture reference model defined in [RFC7365] to describe the use cases.

1.1. Terminology

This document uses the terminologies defined in [RFC7365] and [RFC4364]. Some additional terms used in the document are listed here.

DMZ: Demilitarized Zone. A computer or small sub-network that sits between a trusted internal network, such as a corporate private LAN, and an un-trusted external network, such as the public Internet.

DNS: Domain Name Service [RFC1035]

DC Operator: A role who is responsible to construct and manage cloud service instances in their life-cycle and manage DC infrastructure that runs these cloud instances.

DC Provider: A company that uses its DC infrastructure to offer cloud services to its customers.

NAT: Network Address Translation [RFC3022]

vGW: virtual Gateway; a gateway component used for an NVO3 virtual network to interconnect with another virtual/physical network.

2. Basic NVO3 Networks

An NVO3 network provides communications among Tenant Systems (TS) in a DC. A TS can be a physical server/device or a virtual machine (VM) on a server, i.e., end-device [RFC7365]. A DC provider often uses NVO3 networks for its internal applications in which each application runs on many VMs or physical services and requires application segregation.

A Network Virtual Edge (NVE) is an NVO3 architecture component [RFC7365]]. It is responsible to forward and encapsulate the NVO3 traffic in outbound direction; and decapsulate and forward the NVO3 traffic in inbound direction [NVO3ARCH]. A Network Virtualization Authority (NVA) is another NVO3 architecture component [RFC7365]. An NVE obtains the reachability information of tenant systems in a NVO3 network from the NVA. The tenant systems attached to the same NVE may belong to a same or different NVO3 networks.

The network virtualization overlay in this context means that a virtual network is implemented with an overlay technology, i.e., within a DC, NVO3 traffic is encapsulated at an NVE and carried by a tunnel to another NVE where the packet is decapsulated and sent to a target tenant system [NVO3ARCH]. This architecture decouples a NVO3 network construction from the DC physical network configuration, which provides the flexibility for VM placement and mobility. It also means that the nodes in the infrastructure network (except tunnel end point nodes) carry encapsulated NVO3 traffic but not aware of the existence of NVO3 networks. In the architecture [NVO3ARCH], one tunnel can carry NVO3 traffic belonging to different NVO3 networks; a virtual network identifier is used in an NVO3 encapsulation protocol to differentiate NVO3 traffic.

An NVO3 network may be an L2 or L3 domain. The network provides switching (L2) or routing (L3) capability to support host (i.e. tenent systems) communications. An NVO3 network may required to carry unicast traffic and/or multicast, broadcast/unknown (for L2 only) traffic from/to tenant systems. There are several ways to transport NVO3 network BUM traffic [NVO3MCAST].

It is worth mentioning two distinct cases regarding to NVE location. The first is where TSs and an NVE are co-located on a single end host/device, which means that the NVE can be aware of the TS's state at any time via an internal API. The second is where TSs and an NVE are not co-located, with the NVE residing on a network device; in this case, a protocol is necessary to allow the NVE to be aware of the TS's state [NVO3HYVR2NVE].

One NVO3 network can provide connectivity to many TSs that attach to many different NVEs in a DC. TS dynamic placement and mobility results in frequent changes of the binding between a TS and an NVE. The TS reachability update mechanisms need be fast enough so that the updates do not cause any communication disruption/interruption. The capability of supporting many TSs in a virtual network and many more virtual networks in a DC is critical for the NVO3 solution.

If a virtual network spans across multiple DC sites, one design using NVO3 is to allow the network to seamlessly span across the sites without DC gateway routers' termination. In this case, the tunnel between a pair of NVEs can be carried within other intermediate tunnels over the Internet or other WANs, or an intra DC tunnel and inter DC tunnel(s) can be stitched together to form an end-to-end tunnel between the pair of NVEs that are in different DC sites. Both cases will form one virtual network across multiple DC sites.

3. DC NVO3 Network and External Network Interconnection

Many customers (an enterprise or individuals) who utilize a DC provider's compute and storage resources to run their applications need to access their systems hosted in a DC through Internet or Service Providers' Wide Area Networks (WAN). A DC provider can construct a NVO3 network that provides connectivity to all the resources designated for a customer and allows the customer to access the resources via a virtual gateway (vGW). This, in turn, becomes the case of interconnecting an NVO3 network and the virtual private network (VPN) on the Internet or wide-area networks (WAN). Note that a VPN is not implemented by NVO3 solution. Two use cases are described here.

3.1. DC NVO3 Network Access via the Internet

A customer can connect to an NVO3 network via the Internet in a secure way. Figure 1 illustrates an example of this case. The NVO3 network has an instance at NVE1 and NVE2 and the two NVEs are connected via an IP tunnel in the Data Center. A set of tenant systems are attached to NVE1 on a server. NVE2 resides on a DC Gateway device. NVE2 terminates the tunnel and uses the VNID on the packet to pass the packet to the corresponding vGW entity on the DC GW (the vGW is the default gateway for the virtual network). A customer can access their systems, i.e., TS1 or TSn, in the DC via the Internet by using an IPsec tunnel [RFC4301]. The IPsec tunnel is configured between the vGW and the customer gateway at the customer site. Either a static route or iBGP may be used for prefix advertisement. The vGW provides IPsec functionality such as

authentication scheme and encryption; iBGP protocol traffic is carried within the IPsec tunnel. Some vGW features are listed below:

- The vGW maintains the TS/NVE mappings and advertises the TS prefix to the customer via static route or iBGP.
- Some vGW functions such as firewall and load balancer can be performed by locally attached network appliance devices.
- If the NVO3 network uses different address space than external users, then the vGW needs to provide the NAT function.
- More than one IPsec tunnel can be configured for redundancy.
- The vGW can be implemented on a server or VM. In this case, IP tunnels or IPsec tunnels can be used over the DC infrastructure.
- DC operators need to construct a vGW for each customer.

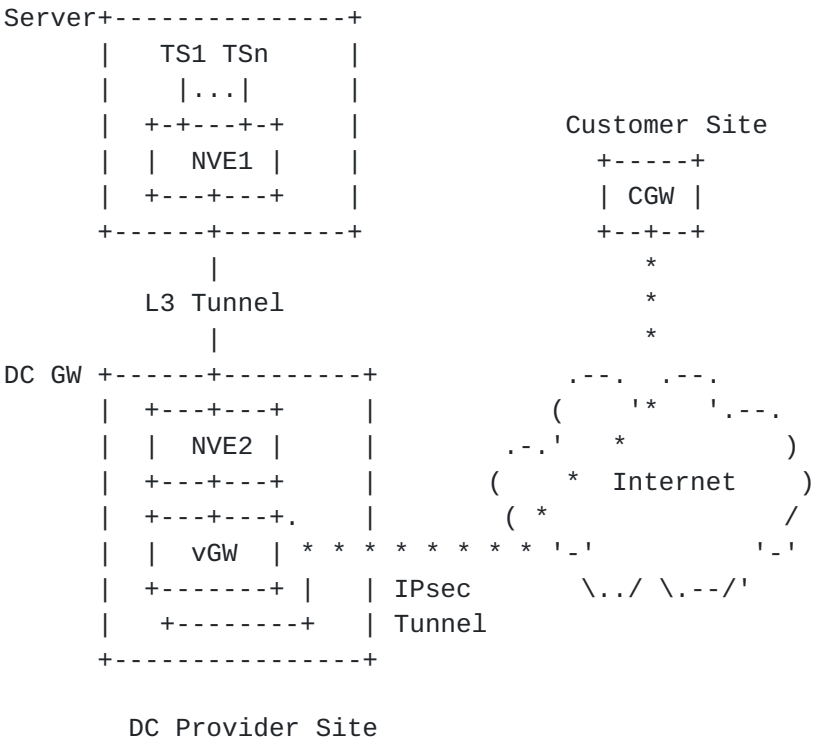

Figure 1 - DC Virtual Network Access via the Internet

3.2. DC NVO3 Network and SP WAN VPN Interconnection

In this case, an Enterprise customer wants to use a Service Provider (SP) WAN VPN [RFC4364] [RFC7432] to interconnect its sites with an NVO3 network in a DC site. The Service Provider constructs a VPN for the enterprise customer. Each enterprise site peers with an SP PE. The DC Provider and VPN Service Provider can build an NVO3 network and a WAN VPN independently, and then interconnect them via a local link, or a tunnel between the DC GW and WAN PE devices. The control plane interconnection options between the DC and WAN are described in RFC4364 [RFC4364]. Using Option A with VRF-LITE [VRF-LITE], both ASBRs, i.e., DC GW and SP PE, maintain a routing/forwarding table (VRF). Using Option B, the DC ASBR and SP ASBR do not maintain the VRF table; they only maintain the NVO3 network and VPN identifier mappings, i.e., label mapping, and swap the label on the packets in the forwarding process. Both option A and B allow the NVO3 network and VPN using own identifier and two identifiers are mapped at DC GW. With option C, the VN and VPN use the same identifier and both ASBRs perform the tunnel stitching, i.e., tunnel segment mapping. Each option has pros/cons [RFC4364] and has been deployed in SP networks depending on the applications in use. BGP is used with these options for route distribution between DCs and SP WANs. Note that if the DC is the SP's Data Center, the DC GW and SP PE in this case can be merged into one device that performs the interworking of the VN and VPN within an AS.

The configurations above allow the enterprise networks to communicate with the tenant systems attached to the NVO3 network in the DC without interfering with the DC provider's underlying physical networks and other NVO3 networks in the DC. The enterprise can use its own address space in the NVO3 network. The DC provider can manage which VM and storage elements attach to the NVO3 network. The enterprise customer manages which applications run on the VMs without knowing the location of the VMs in the DC. (See Section 4 for more)

Furthermore, in this use case, the DC operator can move the VMs assigned to the enterprise from one sever to another in the DC without the enterprise customer being aware, i.e., with no impact on the enterprise's 'live' applications. Such advanced technologies bring DC providers great benefits in offering cloud services, but add some requirements for NVO3 [RFC7364] as well.

4. DC Applications Using NVO3

NVO3 technology provides DC operators with the flexibility in designing and deploying different applications in an end-to-end

virtualization overlay environment. The operators no longer need to worry about the constraints of the DC physical network configuration when creating VMs and configuring a network to connect them. A DC provider may use NVO3 in various ways, in conjunction with other physical networks and/or virtual networks in the DC for a reason. This section highlights some use cases for this goal.

4.1. Supporting Multiple Technologies

Servers deployed in a large data center are often installed at different times, and may have different capabilities/features. Some servers may be virtualized, while others may not; some may be equipped with virtual switches, while others may not. For the servers equipped with Hypervisor-based virtual switches, some may support VxLAN [RFC7348] encapsulation, some may support NVGRE encapsulation [RFC7637], and some may not support any encapsulation. To construct a tenant network among these servers and the ToR switches, operators can construct one traditional VLAN network and two virtual networks where one uses VxLAN encapsulation and the other uses NVGRE, and interconnect these three networks via a gateway or virtual GW. The GW performs packet encapsulation/decapsulation translation between the networks.

Another case is that some software of a tenant is high CPU and memory consumption, which only makes a sense to run on metal servers; other software of the tenant may be good to run on VMs. However provider DC infrastructure is configured to use NVO3 to connect to VMs and VLAN [IEEE802.1Q] connect to metal services. The tenant network requires interworking between NVO3 and traditional VLAN.

4.2. DC Application with Multiple Virtual Networks

A DC application may necessarily be constructed with multi-tier zones, where each zone has different access permissions and runs different applications. For example, a three-tier zone design has a front zone (Web tier) with Web applications, a mid zone (application tier) where service applications such as credit payment or ticket booking run, and a back zone (database tier) with Data. External users are only able to communicate with the Web application in the front zone; the back zone can only receive traffic from the application zone. In this case, communications between the zones must pass through a GW/firewall. Each zone can be implemented by one NVO3 network and a GW/firewall can be used to between two NVO3 networks, i.e., two zones. As a result, a tunnel carrying NVO3 network traffic must be terminated at the GW/firewall where the NVO3 traffic is processed.

4.3. Virtual Data Center (vDC)

An enterprise data center today may deploy routers, switches, and network appliance devices to construct its internal network, DMZ, and external network access; it may have many servers and storage running various applications. With NVO3 technology, a DC Provider can construct a virtual Data Center (vDC) over its physical DC infrastructure and offer a virtual Data Center service to enterprise customers. A vDC at the DC Provider site provides the same capability as the physical DC at a customer site. A customer manages its own applications running in its vDC. A DC Provider can further offer different network service functions to the customer. The network service functions may include firewall, DNS, load balancer, gateway, etc.

Figure 2 below illustrates one such scenario at service abstraction level. In this example, the vDC contains several L2 VNs (L2VNx, L2VNy, L2VNz) to group the tenant systems together on a per-application basis, and one L3 VN (L3VNa) for the internal routing. A network firewall and gateway runs on a VM or server that connects to L3VNa and is used for inbound and outbound traffic processing. A load balancer (LB) is used in L2VNx. A VPN is also built between the gateway and enterprise router. An Enterprise customer runs Web/Mail/Voice applications on VMs within the vDC. The users at the Enterprise site access the applications running in the vDC via the VPN; Internet users access these applications via the gateway/firewall at the provider DC site.

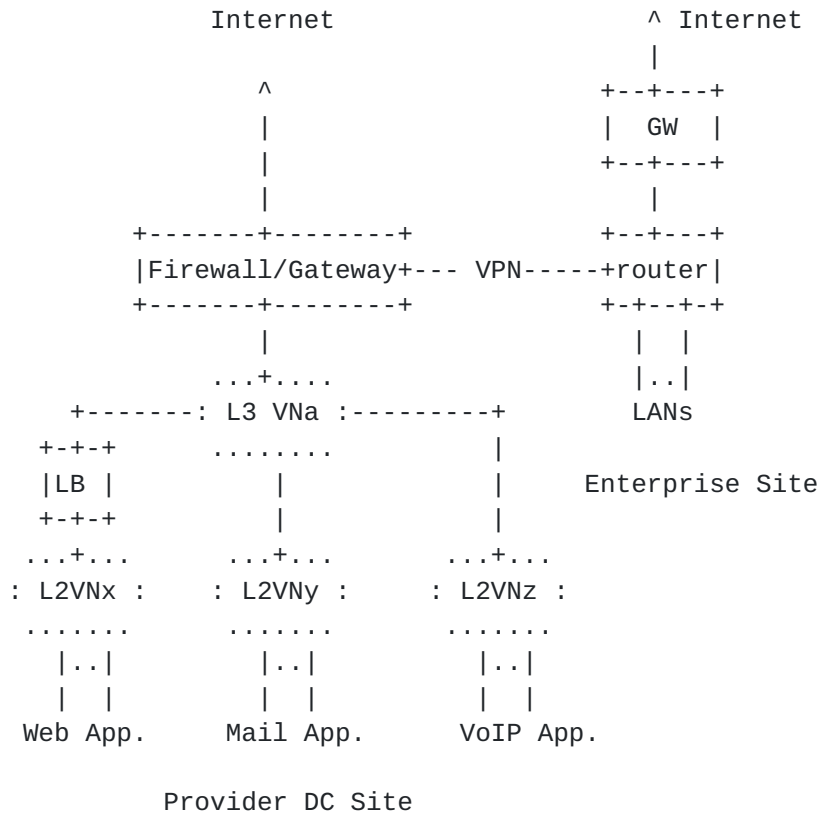

Figure 2 - Virtual Data Center Abstraction View

The enterprise customer decides which applications should be accessible only via the intranet and which should be assessable via both the intranet and Internet, and configures the proper security policy and gateway function at the firewall/gateway. Furthermore, an

enterprise customer may want multi-zones in a vDC (See section 4.2) for the security and/or the ability to set different QoS levels for the different applications.

The vDC use case requires an NVO3 solution to provide DC operators with an easy and quick way to create an NVO3 network and NVEs for any vDC design, to allocate TSs and assign TSs to the corresponding NVO3 network, and to illustrate vDC topology and manage/configure individual elements in the vDC in a secure way.

5. Summary

This document describes some general and potential NVO3 use cases in DCs. The combination of these cases will give operators the flexibility and capability to design more sophisticated cases for various cloud applications.

DC services may vary, from infrastructure as a service (IaaS), to platform as a service (PaaS), to software as a service (SaaS). In these services, NVO3 networks are just a portion of such services.

NVO3 uses tunnel techniques to deliver NVO3 traffic over DC physical infrastructure network. A tunnel encapsulation protocol is necessary. An NVO3 tunnel may in turn be tunneled over other intermediate tunnels over the Internet or other WANs.

An NVO3 network in a DC may be accessed by external users in a secure way. Many existing technologies can help achieve this.

6. Security Considerations

Security is a concern. DC operators need to provide a tenant with a secured virtual network, which means one tenant's traffic is isolated from other tenants' traffic as well as from underlay networks. DC operators also need to prevent against a tenant application attacking their underlay DC network; further, they need to protect against a tenant application attacking another tenant application via the DC infrastructure network. For example, a tenant application attempts to generate a large volume of traffic to overload the DC's underlying network. An NVO3 solution has to address these issues.

7. IANA Considerations

This document does not request any action from IANA.

8. Informative References

[IEEE802.1Q] IEEE, "IEEE Standard for Local and metropolitan area networks -- Media Access Control (MAC) Bridges and Virtual Bridged Local Area", IEEE Std 802.1Q, 2011.

[NVO3HYVR2NVE] Li, Y., et al, "Hypervisor to NVE Control Plane Requirements", draft-ietf-nvo3-hpvr2nve-cp-req-05, work in progress.

[NVO3ARCH] Black, D., et al, "An Architecture for Overlay Networks (NVO3)", draft-ietf-nvo3-arch-08, work in progress.

[NVO3MCAST] Ghanwani, A., Dunbar, L., et al, "A Framework for Multicast in Network Virtualization Overlays", draft-ietf-nvo3-mcast-framework-05, work in progress.

[RFC1035] Mockapetris, P., "DOMAIN NAMES - Implementation and Specification", RFC1035, November 1987.

[RFC3022] Srisuresh, P. and Egevang, K., "Traditional IP Network Address Translator (Traditional NAT)", RFC3022, January 2001.

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

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

[RFC7348] Mahalingam, M., Dutt, D., et al, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC7348 August 2014.

[RFC7364] Narten, T., et al "Problem Statement: Overlays for Network Virtualization", RFC7364, October 2014.

[RFC7365] Lasserre, M., Motin, T., et al, "Framework for DC Network Virtualization", RFC7365, October 2014.

[RFC7432] Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A. and J. Uttaro, "BGP MPLS Based Ethernet VPN", RFC7432, February 2015

[RFC7637] Garg, P., and Wang, Y., "NVGRE: Network Virtualization using Generic Routing Encapsulation", RFC7637, Sept. 2015.

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

Contributors

Vinay Bannai
PayPal
2211 N. First St,
San Jose, CA 95131
Phone: +1-408-967-7784
Email: vbannai@paypal.com

Ram Krishnan
Brocade Communications
San Jose, CA 95134
Phone: +1-408-406-7890
Email: ramk@brocade.com

Kieran Milne
Juniper Networks
1133 Innovation Way
Sunnyvale, CA 94089
Phone: +1-408-745-2000
Email: kmilne@juniper.net

Acknowledgements

Authors like to thank Sue Hares, Young Lee, David Black, Pedro Marques, Mike McBride, David McDysan, Randy Bush, Uma Chunduri, Eric Gray, David Allan, Joe Touch, Olufemi Komolafe, and Matthew Bocci for the review, comments, and suggestions.

Authors' Addresses

Lucy Yong
Huawei Technologies

Phone: +1-918-808-1918
Email: lucy.yong@huawei.com

Linda Dunbar
Huawei Technologies,

5340 Legacy Dr.
Plano, TX 75025 US

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

Mehmet Toy
Verizon
Phone : +1-856-792-2801
E-mail : mtoy054@yahoo.com

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

Vishwas Manral

Email: vishwas@ionosnetworks.com