Network Working Group
Internet Draft
Category: Informational

L. Yong
Huawei
M. Toy
Comcast
A. Isaac
Bloomberg
V. Manral
Ionos Networks
L. Dunbar
Huawei

Expires: April 2016 October 16, 2015

# Use Cases for Data Center Network Virtualization Overlays

draft-ietf-nvo3-use-case-07

Abstract

This document describes Data Center (DC) Network Virtualization over Layer 3 (NVO3) use cases that can be deployed in various data centers and serve different 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 Network Virtualization Overlays in the DC is to decouple the communication among tenant systems from DC physical infrastructure networks and to allow one physical network infrastructure to provide:

- Multi-tenant virtual networks and traffic isolation among the virtual networks over the same physical network.

- Independent address spaces in individual virtual 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 and configuration changes, and the ability to perform a "hot move" with no disruption to the live application running on VMs.

These characteristics of NVO3 help address the issues that cloud applications face in Data Centers [RFC7364].

An NVO3 network may interconnect with another NVO3 virtual network, or another physical network (i.e., not the physical network that the NVO3 network is 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 use cases that apply to various data centers. Three types of the use cases described in this document are:

- Basic NVO3 virtual networks in a DC (Section 2). All Tenant Systems (TS) in the virtual network are located within the same DC. The individual virtual networks can be either Layer 2 (L2) or Layer 3 (L3). The number of NVO3 virtual 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.

- Virtual networks that span across multiple Data Centers and/or to customer premises, i.e., an NVO3 virtual network where some tenant systems in a DC attach to interconnects 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 virtual networks that are interconnected by the gateways. Three scenarios are described in Section 4: 1) using NVO3 and other network technologies to build a tenant network; 2) constructing several virtual networks as a tenant network; 3) applying NVO3 to a virtualized DC (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]

NAT: Network Address Translation [RFC1631]

Note that a virtual network in this document refers to an NVO3 virtual network in a DC [RFC7365].

## 2. Basic Virtual Networks in a Data Center

A virtual network in a DC enables communications among Tenant Systems (TS). A TS can be a physical server/device or a virtual

machine (VM) on a server, i.e., end-device [RFC7365]. A Network Virtual Edge (NVE) can be co-located with a TS, i.e., on the same end-device, or reside on a different device, e.g., a top of rack switch (ToR). A virtual network has a virtual network identifier (can be globally unique or locally significant at NVEs).

Tenant Systems attached to the same NVE may belong to the same or different virtual networks. An NVE provides tenant traffic forwarding/encapsulation and obtains tenant systems reachability information from a Network Virtualization Authority (NVA)[NVO3ARCH]. DC operators can construct multiple separate virtual networks, and provide each with own address space.

Network Virtualization Overlay in this context means that a virtual network is implemented with an overlay technology, i.e., within a DC that has IP infrastructure, tenant traffic is encapsulated at its local NVE and carried by a tunnel to another NVE where the packet is decapsulated and sent to a target tenant system. This architecture decouples tenant system address space and configuration from the infrastructure's, which provides great flexibility for VM placement and mobility. It also means that the transit nodes in the infrastructure are not aware of the existence of the virtual networks and tenant systems attached to the virtual networks. The tunneled packets are carried as regular IP packets and are sent to NVEs. One tunnel may carry the traffic belonging to multiple virtual networks; a virtual network identifier is used for traffic demultiplexing. A tunnel encapsulation protocol is necessary for NVE to encapsulate the packets from Tenant Systems and encode other information on the tunneled packets to support NVO3 implementation.

A virtual network implemented by NVO3 may be an L2 or L3 domain. The virtual network can carry unicast traffic and/or multicast, broadcast/unknown (for L2 only) traffic from/to tenant systems. There are several ways to transport virtual 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 virtual 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 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 the intra DC and inter DC tunnels can be stitched together to form a 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 Virtual 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 virtual 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 a DC virtual network and the network at customer site(s) via the Internet or WANs. Two use cases are described here.

### 3.1. DC Virtual Network Access via the Internet

A customer can connect to a DC virtual network via the Internet in a secure way. Figure 1 illustrates this case. The DC virtual 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 virtual network in the DC 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.

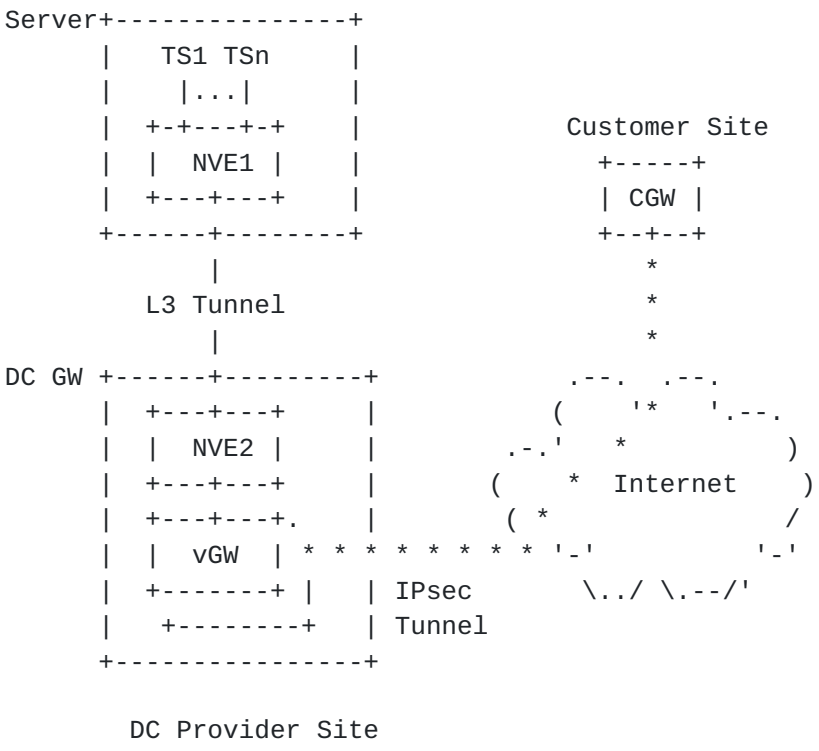



Figure 1 - DC Virtual Network Access via the Internet

## 3.2. DC VN 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 a virtual 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 a DC virtual

network (VN) and 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 VN 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 VN 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 VN in a DC without interfering with the DC provider's underlying physical networks and other virtual networks. The enterprise can use its own address space in the VN. The DC provider can manage which VM and storage elements attach to the VN. The enterprise customer manages which applications run on the VMs in the VN 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. Operators no longer need to worry about the constraints of the DC physical network configuration when creating VMs and configuring a virtual network. 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 and Applications

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.

A data center may be also constructed with multi-tier zones, where each zone has different access permissions and runs different applications. For example, the 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. In this case, communications between the zones must pass through the security GW/firewall. One virtual network can be configured in each zone and a GW can be used to interconnect two virtual networks, i.e., two zones. If the virtual network in individual zones uses the different implementations, the GW needs to support these implementations as well.

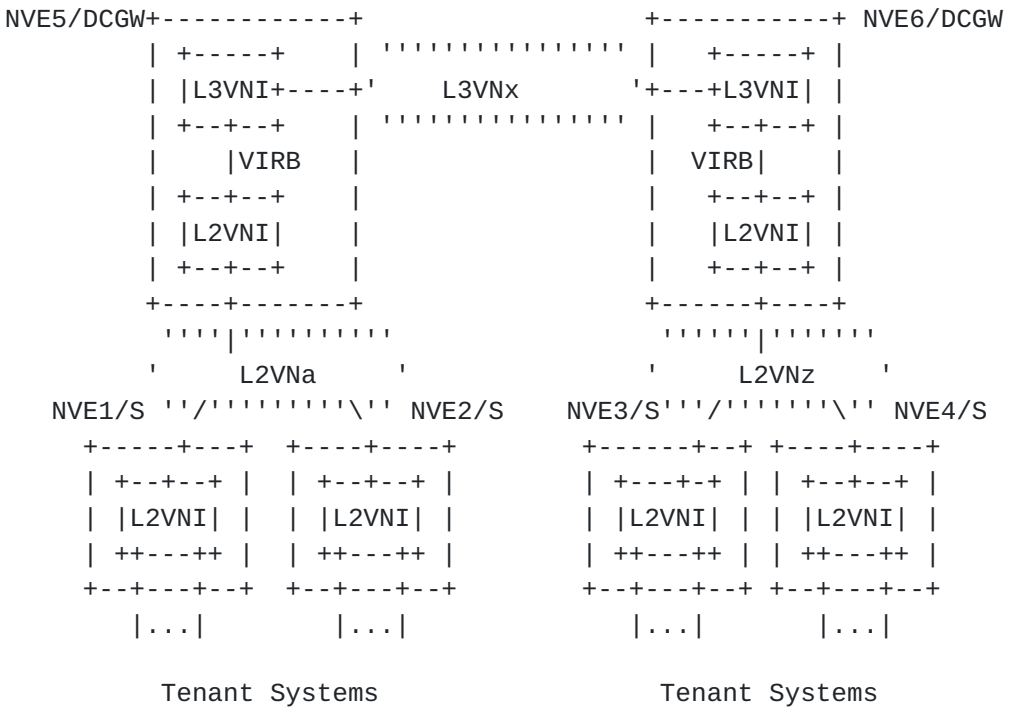
### 4.2. Tenant Network with Multiple Subnets

A tenant network may contain multiple subnets. The DC physical network needs to support the connectivity for many such tenant networks. In some cases, the inter-subnet policies can be placed at designated gateway devices. Such a design requires the inter-subnet traffic to be sent to one of the gateway devices first for the policy checking, which may cause traffic to "hairpin" at the gateway in a DC. It is desirable for an NVE to be able to hold some policies and be able to forward the inter-subnet traffic directly. To reduce the burden on the NVE, a hybrid design may be deployed, i.e., an NVE can perform forwarding for selected inter-subnets while the designated GW performs forwarding for the rest. For example, each NVE performs inter-subnet forwarding for intra-DC traffic while the designated GW is used for traffic to/from a remote DC.

A tenant network may span across multiple Data Centers that are at different locations. DC operators may configure an L2 VN within each DC and an L3 VN between DCs for a tenant network. For this configuration, the virtual L2/L3 gateway can be implemented on the DC GW device. Figure 2 illustrates this configuration.

Figure 2 depicts two DC sites. Site A constructs one L2 VN, say L2VNa, on NVE1, NVE2, and NVE5. NVE1 and NVE2 reside on the servers which host multiple tenant systems. NVE5 resides on the DC GW device. Site Z has similar configuration, with L2VNz on NVE3, NVE4, and NVE6. An L3 VN, L3VNx, is configured on NVE5 at Site A and the NVE6 at Site Z. An internal Virtual Interface of Routing and Bridging (VIRB) is used between the L2VNI and L3VNI on NVE5 and NVE6, respectively. The L2VNI requires the MAC/NVE mapping table and the L3VNI requires the IP prefix/NVE mapping table. A packet arriving at NVE5 from L2VNa will be decapsulated, converted into an IP packet, and then encapsulated and sent to Site Z. A packet to NVE5 from L3VNx will be decapsulated, converted into a MAC frame, and then encapsulated and sent within Site A. The ARP protocol [RFC826] can be used to get the MAC address for an IP address in the L2VNa. The policies can be checked at the VIRB.

Note that L2VNa, L2VNz, and L3VNx in Figure 2 are NVO3 virtual networks.



DC Site A DC Site Z

Figure 2 - Tenant Virtual Network with Bridging/Routing

### 4.3. Virtualized 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 virtualized DC over its physical DC infrastructure and offer a virtual DC service to enterprise customers. A vDC at the DC Provider site provides the same capability as a physical DC at the customer site. A customer manages their own applications running in their 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 3 below illustrates one such scenario. For simplicity, it only shows the L3 VN or L2 VN in abstraction. In this example, the DC Provider operators create 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. The Enterprise customer runs Web/Mail/Voice applications on VMs at the provider DC site which may be spread across many servers. The users at the Enterprise site access the applications running in the provider DC site via the VPN; Internet users access these applications via the gateway/firewall at the provider DC.

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 4.1) for the security and/or the ability to set different QoS levels for the different applications.

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

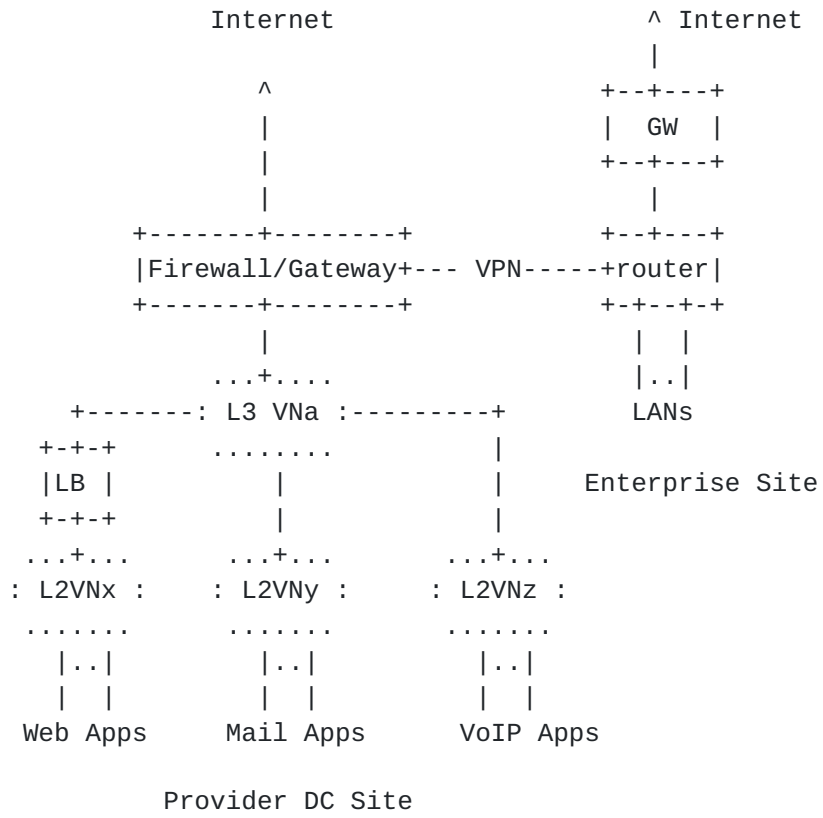



Figure 3 - Virtual Data Center (vDC)

## 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 virtual networks are just a portion of such services.

NVO3 uses tunnel techniques to deliver VN traffic over an IP 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 virtual network in a DC may be accessed by external users in a secure way. Many existing technologies can help achieve this.

NVO3 implementations may vary. Some DC operators prefer to use a centralized controller to manage tenant system reachability in a virtual network, while other operators prefer to use distribution protocols to advertise the tenant system location, i.e., NVE location. When a tenant network spans across multiple DCs and WANs, each network administration domain may use different methods to distribute the tenant system locations. Both control plane and data plane interworking are necessary.

## 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 non-tenants' traffic. DC operators also need to prevent against a tenant application attacking their underlying DC network through the tenant's virtual 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. References

### 8.1. Normative References

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

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

### 8.2. Informative References

[IEEE 802.1Q] 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-01, work in progress.

[NVO3ARCH] Black, D., et al, "An Architecture for Overlay Networks (NVO3)", draft-ietf-nvo3-arch-02, work in progress.

[NVO3MCAST] Ghanwani, A., "Framework of Supporting Applications Specific Multicast in NVO3", draft-ghanwani-nvo3-app-mcast-framework-02, work in progress.

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

[RFC1631] Egevang, K., Francis, P., "The IP network Address Translator (NAT)", RFC1631, May 1994.

[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., ific Multicast in etc "VXLAN: A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC7348 August 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, and Eric Gray for the review, comments, and suggestions.

## Authors' Addresses

Lucy Yong
Huawei Technologies

Phone: +1-918-808-1918
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

Ionas Networks

Email: vishwas@ionosnetworks.com

Linda Dunbar
Huawei Technologies,
5340 Legacy Dr.
Plano, TX 75025 US

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