

Internet Engineering Task Force (IETF)
Request for Comments: 8151
Category: Informational
ISSN: 2070-1721

L. Yong
L. Dunbar
Huawei
M. Toy
Verizon
A. Isaac
Juniper Networks
V. Manral
Nano Sec Co
May 2017

Use Cases for Data Center Network Virtualization Overlay Networks

Abstract

This document describes Network Virtualization over Layer 3 (NV03) use cases that can be deployed in various data centers and serve different data-center applications.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are a candidate for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at <http://www.rfc-editor.org/info/rfc8151>.

Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

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. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
1.1. Terminology	4
1.2. NV03 Background	5
2. DC with a Large Number of Virtual Networks	6
3. DC NV03 Virtual Network and External Network Interconnection	6
3.1. DC NV03 Virtual Network Access via the Internet	7
3.2. DC NV03 Virtual Network and SP WAN VPN Interconnection	8
4. DC Applications Using NV03	9
4.1. Supporting Multiple Technologies	9
4.2. DC Applications Spanning Multiple Physical Zones	10
4.3. Virtual Data Center (vDC)	10
5. Summary	12
6. Security Considerations	12
7. IANA Considerations	12
8. Informative References	13
Acknowledgements.....	14
Contributors	15
Authors' Addresses.....	16

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 limits in supporting cloud applications and multi-tenant networks [RFC7364]. The goal of data center Network Virtualization over Layer 3 (NV03) networks is to decouple the communication among tenant systems from DC physical infrastructure networks and to allow one physical network infrastructure to:

- o carry many NV03 virtual networks and isolate the traffic of different NV03 virtual networks on a physical network.
- o provide independent address space in individual NV03 virtual network such as Media Access Control (MAC) and IP.
- o Support flexible Virtual Machines (VMs) 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 those VMs.

These characteristics of NV03 virtual networks (VNs) help address the issues that cloud applications face in data centers [RFC7364].

Hosts in one NV03 VN may communicate with hosts in another NV03 VN that is carried by the same physical network, or different physical network, via a gateway. The use-case examples for the latter are as follows:

- 1) DCs that migrate toward an NV03 solution will be done in steps, where a portion of tenant systems in a VN are on virtualized servers while others exist on a LAN.
- 2) many DC applications serve Internet users who are on different physical networks;
- 3) some applications are CPU bound, such as Big Data analytics, and may not run on virtualized resources.

The inter-VN policies are usually enforced by the gateway.

This document describes general NV03 VN use cases that apply to various data centers. The use cases described here represent the DC provider's interests and vision for their cloud services. The

document groups the use cases into three categories from simple to sophisticated in terms of implementation. However, the implementation details of these use cases are outside the scope of this document. These three categories are described below:

- o Basic NV03 VNs (Section 2). All Tenant Systems (TSs) 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 NV03 VNs in a DC is much larger than the number that traditional VLAN-based virtual networks [IEEE802.1Q] can support.
- o A virtual network that spans across multiple DCs and/or to customer premises where NV03 virtual networks are constructed and interconnect other virtual or physical networks outside the DC. An enterprise customer may use a traditional carrier-grade VPN or an IPsec tunnel over the Internet to communicate with its systems in the DC. This is described in Section 3.
- o DC applications or services require an advanced network that contains several NV03 virtual networks that are interconnected by gateways. Three scenarios are described in Section 4:
 - (1) supporting multiple technologies;
 - (2) constructing several virtual networks as a tenant network; and
 - (3) applying NV03 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 terminology defined in [RFC7365] and [RFC4364]. Some additional terms used in the document are listed here.

ASBR:	Autonomous System Border Router.
DC:	Data Center.
DMZ:	Demilitarized Zone. A computer or small subnetwork between a more-trusted internal network, such as a corporate private LAN, and an untrusted or less-trusted external network, such as the public Internet.
DNS:	Domain Name Service [RFC1035].

DC Operator: An entity that is responsible for constructing and managing all resources in DCs, including, but not limited to, computing, storage, networking, etc.

DC Provider: An entity that uses its DC infrastructure to offer services to its customers.

NAT: Network Address Translation [RFC3022].

vGW: virtual GateWay. A gateway component used for an NV03 virtual network to interconnect with another virtual/physical network.

NV03: Network Virtualization over Layer 3. A virtual network that is implemented based on the NV03 architecture.

PE: Provider Edge.

SP: Service Provider.

TS: A Tenant System, which can be instantiated on a physical server or virtual machine (VM).

VRF-LITE: Virtual Routing and Forwarding - LITE [VRF-LITE].

VN: Virtual Network

VoIP: Voice over IP

WAN VPN: Wide Area Network Virtual Private Network [RFC4364] [RFC7432].

1.2. NV03 Background

An NV03 virtual network is in a DC that is implemented based on the NV03 architecture [RFC8014]. This architecture is often referred to as an overlay architecture. The traffic carried by an NV03 virtual network is encapsulated at a Network Virtualization Edge (NVE) [RFC8014] and carried by a tunnel to another NVE where the traffic is decapsulated and sent to a destination Tenant System (TS). The NV03 architecture decouples NV03 virtual networks from the DC physical network configuration. The architecture uses common tunnels to carry NV03 traffic that belongs to multiple NV03 virtual networks.

An NV03 virtual network may be an L2 or L3 domain. The network provides switching (L2) or routing (L3) capability to support host (i.e., TS) communications. An NV03 virtual network may be required to carry unicast traffic and/or multicast or broadcast/unknown-

unicast (for L2 only) traffic to/from TSs. There are several ways to transport NV03 virtual network Broadcast, Unknown Unicast, and Multicast (BUM) traffic [NV03MCAST].

An NV03 virtual network provides communications among TSs in a DC. A TS can be a physical server/device or a VM on a server end-device [RFC7365].

2. DC with a Large Number of Virtual Networks

A DC provider often uses NV03 virtual networks for internal applications where each application runs on many VMs or physical servers and the provider requires applications to be segregated from each other. A DC may run a larger number of NV03 virtual networks to support many applications concurrently, where a traditional VLAN solution based on IEEE 802.1Q is limited to 4094 VLANs.

Applications running on VMs may require a different quantity of computing resources, which may result in a computing-resource shortage on some servers and other servers being nearly idle. A shortage of computing resources may impact application performance. DC operators desire VM or workload movement for resource-usage optimization. VM dynamic placement and mobility results in frequent changes of the binding between a TS and an NVE. The TS reachability update mechanisms should take significantly less time than the typical retransmission Timeout window of a reliable transport protocol such as TCP and Stream Control Transmission Protocol (SCTP), so that endpoints' transport connections won't be impacted by a TS becoming bound to a different NVE. The capability of supporting many TSs in a virtual network and many virtual networks in a DC is critical for an NV03 solution.

When NV03 virtual networks segregate VMs belonging to different applications, DC operators can independently assign MAC and/or IP address space to each virtual network. This addressing is more flexible than requiring all hosts in all NV03 virtual networks to share one address space. In contrast, typical use of IEEE 802.1Q VLANs requires a single common MAC address space.

3. DC NV03 Virtual Network and External Network Interconnection

Many customers (enterprises 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 NV03 virtual network that provides connectivity to all the resources designated for a customer, and it allows the customer

to access the resources via a virtual GateWay (vGW). WAN connectivity to the vGW can be provided by VPN technologies such as IPsec VPNs [RFC4301] and BGP/MPLS IP VPNs [RFC4364].

If a virtual network spans multiple DC sites, one design using NV03 is to allow the network to seamlessly span 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 NV03 virtual network across multiple DC sites.

Two use cases are described in the following sections.

3.1. DC NV03 Virtual Network Access via the Internet

A customer can connect to an NV03 virtual network via the Internet in a secure way. Figure 1 illustrates an example of this case. The NV03 virtual network has an instance at NVE1 and NVE2, and the two NVEs are connected via an IP tunnel in the DC. A set of TSs are attached to NVE1 on a server. NVE2 resides on a DC Gateway device. NVE2 terminates the tunnel and uses the VN Identifier (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 TS_n, 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 Internal Border Gateway Protocol (IBGP) may be used for prefix advertisement. The vGW provides IPsec functionality such as authentication scheme and encryption; IBGP traffic is carried within the IPsec tunnel. Some vGW features are listed below:

- o The vGW maintains the TS/NVE mappings and advertises the TS prefix to the customer via static route or IBGP.
- o Some vGW functions such as the firewall and load-balancer (LB) can be performed by locally attached network appliance devices.
- o If the NV03 virtual network uses different address space than external users, then the vGW needs to provide the NAT function.
- o More than one IPsec tunnel can be configured for redundancy.

- o 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.
- o DC operators need to construct a vGW for each customer.

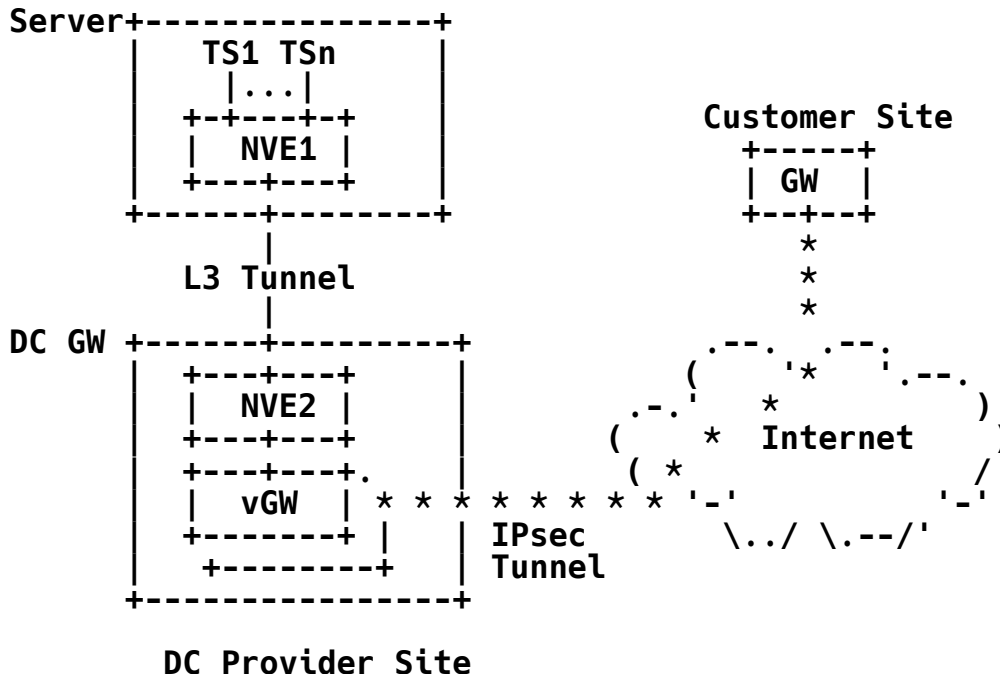


Figure 1: DC Virtual Network Access via the Internet

3.2. DC NV03 Virtual 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 NV03 virtual network in a DC site. The SP constructs a VPN for the enterprise customer. Each enterprise site peers with an SP PE. The DC provider and VPN SP can build an NV03 virtual 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]. Using the option "a" specified in [RFC4364] with VRF-LITE [VRF-LITE], both ASBRs, i.e., DC GW and SP PE, maintain a routing/forwarding table (VRF). Using the option "b" specified in [RFC4364], the DC ASBR and SP ASBR do not maintain the VRF table; they only maintain the NV03 virtual network and VPN identifier mappings, i.e., label mapping, and swap the label on the packets in the forwarding process. Both option "a" and option "b" allow the use of NV03 VNs and VPNs using their own identifiers, and two identifiers are mapped at the DC GW. With the option "c" in [RFC4364], the VN

and VPN use the same identifier and both ASBRs perform the tunnel stitching, i.e., tunnel segment mapping. Each option has pros and cons [RFC4364] and has been deployed in SP networks depending on the application requirements. BGP is used in these options for route distribution between DCs and SP WANs. Note that if the DC is the SP's DC, the DC GW and SP PE can be merged into one device that performs the interworking of the VN and VPN within an Autonomous System.

These solutions allow the enterprise networks to communicate with the tenant systems attached to the NV03 virtual network in the DC without interfering with the DC provider's underlying physical networks and other NV03 virtual networks in the DC. The enterprise can use its own address space in the NV03 virtual network. The DC provider can manage which VM and storage elements attach to the NV03 virtual 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 information.)

Furthermore, in this use case, the DC operator can move the VMs assigned to the enterprise from one server 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 NV03 [RFC7364] as well.

4. DC Applications Using NV03

NV03 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 NV03 in various ways, in conjunction with other physical networks and/or virtual networks in the DC. This section highlights some use cases for this goal.

4.1. Supporting Multiple Technologies

Servers deployed in a large DC are often installed at different times, and they 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 a standardized NV03 encapsulation, some may not support any encapsulation, and some may support a documented encapsulation protocol (e.g., Virtual eXtensible Local Area Network (VXLAN) [RFC7348] and Network Virtualization using Generic Routing

Encapsulation (NVGRE) [RFC7637]) or proprietary encapsulations. To construct a tenant network among these servers and the Top-of-Rack (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 has high CPU and memory consumption, which only makes sense to run on standalone servers; other software of the tenant may be good to run on VMs. However, provider DC infrastructure is configured to use NV03 to connect VMs and VLANs [IEEE802.1Q] to physical servers. The tenant network requires interworking between NV03 and traditional VLAN.

4.2. DC Applications Spanning Multiple Physical Zones

A DC can be partitioned into multiple physical zones, with each zone having different access permissions and running 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 one or more security functions in a physical DMZ zone. Each zone can be implemented by one NV03 virtual network and the security functions in DMZ zone can be used to between two NV03 virtual networks, i.e., two zones. If network functions (NFs), especially the security functions in the physical DMZ, can't process encapsulated NV03 traffic, the NV03 tunnels have to be terminated for the NF to perform its processing on the application traffic.

4.3. Virtual Data Center (vDC)

An enterprise DC 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 NV03 technology, a DC provider can construct a vDC over its physical DC infrastructure and offer a vDC 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 a firewall, DNS, LB, gateway, etc.

Figure 2 illustrates one such scenario at the service-abstraction level. In this example, the vDC contains several L2 VNs (L2VN_x, L2VN_y, L2VN_z) to group the tenant systems together on a per-application basis, and one L3 VN (L3VN_a) for the internal routing. A network firewall and gateway runs on a VM or server that connects to L3VN_a and is used for inbound and outbound traffic processing. An LB is used in L2VN_x. 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 DC provider site.

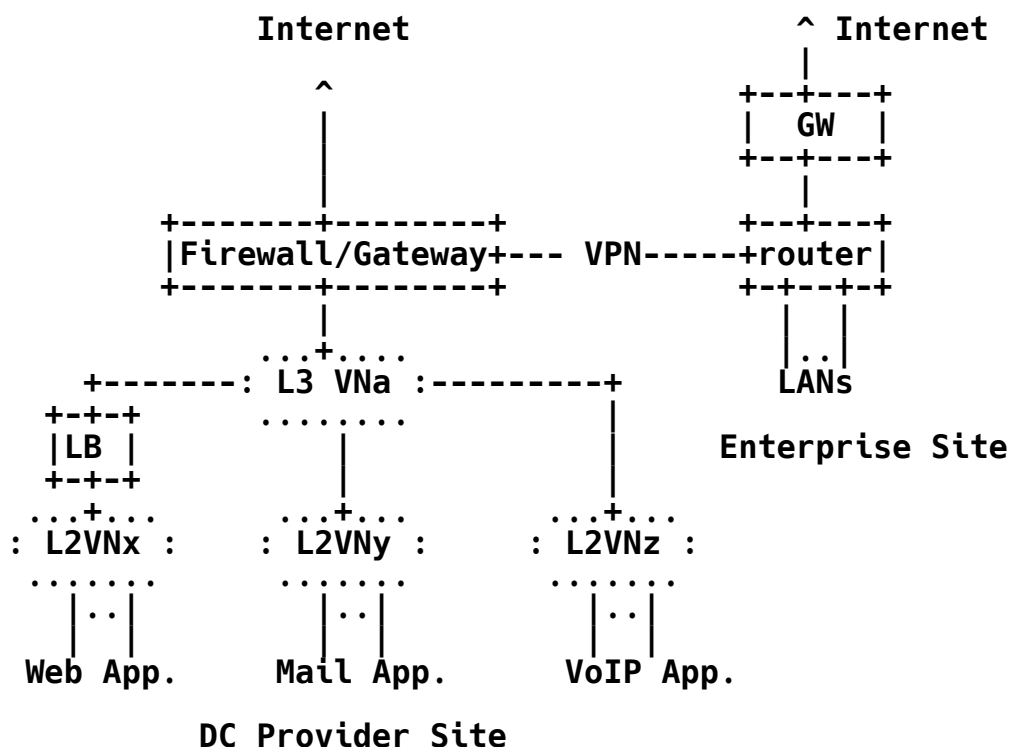


Figure 2: Virtual Data Center Abstraction View

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 it 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 NV03 solution to provide DC operators with an easy and quick way to create an NV03 virtual network and NVEs for any vDC design, to allocate TSs and assign TSs to the corresponding NV03 virtual 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 NV03 use cases in DCs. The combination of these cases will give operators the flexibility and capability to design more sophisticated support for various cloud applications.

DC services may vary, NV03 virtual networks make it possible to scale a large number of virtual networks in a DC and ensure the network infrastructure not impacted by the number of VMs and dynamic workload changes in a DC.

NV03 uses tunnel techniques to deliver NV03 traffic over DC physical infrastructure network. A tunnel encapsulation protocol is necessary. An NV03 tunnel may, in turn, be tunneled over other intermediate tunnels over the Internet or other WANs.

An NV03 virtual 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 and is not leaked to the underlay networks. Tenants are vulnerable to observation and data modification/injection by the operator of the underlay and should only use operators they trust. DC operators also need to prevent a tenant application attacking their underlay DC networks; further, they need to protect 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. This can be done by limiting the bandwidth of such communications.

7. IANA Considerations

This document does not require any IANA actions.

8. Informative References

- [IEEE802.1Q] IEEE, "IEEE Standard for Local and metropolitan area networks -- Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks", IEEE Std 802.1Q-2011, DOI 10.1109/IEEESTD.2011.6009146.
- [NV03MCAST] Ghanwani, A., Dunbar, L., McBride, M., Bannai, V., and R. Krishnan, "A Framework for Multicast in Network Virtualization Overlays", Work in Progress, draft-ietf-nvo3-mcast-framework-07, May 2016.
- [RFC1035] Mockapetris, P., "Domain names - implementation and specification", STD 13, RFC 1035, DOI 10.17487/RFC1035, November 1987, <<http://www.rfc-editor.org/info/rfc1035>>.
- [RFC3022] Srisuresh, P. and K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)", RFC 3022, DOI 10.17487/RFC3022, January 2001, <<http://www.rfc-editor.org/info/rfc3022>>.
- [RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, DOI 10.17487/RFC4301, December 2005, <<http://www.rfc-editor.org/info/rfc4301>>.
- [RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February 2006, <<http://www.rfc-editor.org/info/rfc4364>>.
- [RFC7348] Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M., and C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC 7348, DOI 10.17487/RFC7348, August 2014, <<http://www.rfc-editor.org/info/rfc7348>>.
- [RFC7364] Narten, T., Ed., Gray, E., Ed., Black, D., Fang, L., Kreeger, L., and M. Napierala, "Problem Statement: Overlays for Network Virtualization", RFC 7364, DOI 10.17487/RFC7364, October 2014, <<http://www.rfc-editor.org/info/rfc7364>>.

- [RFC7365] Lasserre, M., Balus, F., Morin, T., Bitar, N., and Y. Rekhter, "Framework for Data Center (DC) Network Virtualization", RFC 7365, DOI 10.17487/RFC7365, October 2014, <<http://www.rfc-editor.org/info/rfc7365>>.
- [RFC7432] Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A., Uttaro, J., Drake, J., and W. Henderickx, "BGP MPLS-Based Ethernet VPN", RFC 7432, DOI 10.17487/RFC7432, February 2015, <<http://www.rfc-editor.org/info/rfc7432>>.
- [RFC7637] Garg, P., Ed., and Y. Wang, Ed., "NVGRE: Network Virtualization Using Generic Routing Encapsulation", RFC 7637, DOI 10.17487/RFC7637, September 2015, <<http://www.rfc-editor.org/info/rfc7637>>.
- [RFC8014] Black, D., Hudson, J., Kreeger, L., Lasserre, M., and T. Narten, "An Architecture for Data-Center Network Virtualization over Layer 3 (NV03)", RFC 8014, DOI 10.17487/RFC8014, December 2016, <<http://www.rfc-editor.org/info/rfc8014>>.
- [VRF-LITE] Cisco, "Configuring VRF-lite", <<http://www.cisco.com/c/en/us/td/docs/switches/lan/catalyst4500/12-2/31sg/configuration/guide/conf/vrf.pdf>>.

Acknowledgements

The authors would 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, Matthew Bocci, and Alia Atlas for the reviews, comments, and suggestions.

Contributors

David Black
Dell EMC
176 South Street
Hopkinton, MA 01748
United States of America

Email: David.Black@dell.com

Vinay Bannai
PayPal
2211 N. First Street
San Jose, CA 95131
United States of America

Phone: +1-408-967-7784
Email: vbannai@paypal.com

Ram Krishnan
Brocade Communications
San Jose, CA 95134
United States of America

Phone: +1-408-406-7890
Email: ramk@brocade.com

Kieran Milne
Juniper Networks
1133 Innovation Way
Sunnyvale, CA 94089
United States of America

Phone: +1-408-745-2000
Email: kmilne@juniper.net

Authors' Addresses

Lucy Yong
Huawei Technologies
Phone: +1-918-808-1918

Email: lucy.yong@huawei.com

Linda Dunbar
Huawei Technologies,
5340 Legacy Drive
Plano, TX 75025
United States of America

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

Mehmet Toy
Verizon

Email: mehmet.toy@verizon.com

Aldrin Isaac
Juniper Networks
1133 Innovation Way
Sunnyvale, CA 94089
United States of America

Email: aldrin.isaac@gmail.com

Vishwas Manral
Nano Sec Co
3350 Thomas Rd.
Santa Clara, CA
United States of America

Email: vishwas@nanosec.io