Network Working Group   L. Yong
Internet Draft   Huawei
Category: Informational   M. Toy
                                                              Comcast
                                                            A. Isaac
                                                            Bloomberg
                                                           V. Manral
                                                       Ionos Networks
                                                          L. Dunbar
                                                             Huawei

Expires: February 2016   August 4, 2015

              Use Cases for Data Center Network Virtualization Overlays

                       draft-ietf-nvo3-use-case-06

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 to different applications.

Status of this Memo

Copyright Notice

Table of Contents

[1](#). **Introduction**

   Server Virtualization has changed the Information Technology (IT)
   industry in terms of the efficiency, cost, and speed of providing a
   new applications and/or services. 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 DC is to decouple the communication among tenant systems
   from DC physical infrastructure networks and to allow one physical
   network infrastructure to provide:

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

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

   o Flexible Virtual Machines (VM) and/or workload placement
      including the ability to move them from server to server without
      requiring VM address and configuration change and the ability
      doing 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 can interconnect with another physical network, i.e.,
   not the physical network that the NVO3 network is over. For example:
   1) DCs that migrate toward NVO3 solution will be done in steps; 2)
   many DC applications serve to Internet cloud users who are on
   physical networks; 3) some applications are CPU bound such as Big
   Data analytics and may not run on virtualized resources.

   This document describes general NVO3 use cases that apply to various
   data centers. Three types of the use cases described here are:

   o Basic NVO3 virtual networks in a DC (Section 2). All Tenant
      Systems (TS) in virtual networks are located within one DC. The
      individual virtual networks can be either Layer 2 (L2) or Layer 3
      (L3). The number of virtual networks supported by NVO3 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.

   o Virtual networks that span across multiple Data Centers and/or to
      customer premises, i.e., a virtual network that connects some
      tenant systems in a DC interconnects another virtual or physical
      network outside the data center. An enterprise customer may use a
      traditional carrier VPN or an IPsec tunnel over Internet to
      communicate its systems in the DC. This is described in Section 3.

   o DC applications or services that may use NVO3 (Section 4). Three
      scenarios are described: 1) use NVO3 and other network
      technologies to build a tenant network; 2) construct several
      virtual networks as a tenant network; 3) apply 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

   NAT: Network Address Translation

   Note that a virtual network in this document is a virtual network in
   DC that is implemented with NVO3 technology.

2. Basic Virtual Networks in a Data Center

   A virtual network in a DC enables a communication 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 a 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
   global unique or local 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 Network Virtualization Authority (NVA)[NVO3ARCH].

   DC operators can construct many virtual networks that have no
   communication in between at all. In this case, each virtual network
   can have its own address spaces such as MAC and IP. DC operators can
   also construct multiple virtual networks in a way so that the
   policies are enforced when the TSs in one virtual network
   communicate with the TSs in other virtual networks. This is referred
   to as Distributed Gateway [NVO3ARCH].

   A Tenant System can be configured with one or multiple addresses and
   participate in multiple virtual networks, i.e., use the same or
   different address in different virtual networks. For examples, a
   Tenant System can be a NAT GW or a firewall and connect to more than
   one virtual network.

   Network Virtualization Overlay in this context means that a virtual
   network is implemented with an overlay technology, i.e., tenant
   traffic is encapsulated at its local NVE and carried by a tunnel
   over DC IP network to another NVE where the packet is decapsulated
   prior to sending to a target tenant system. This architecture
   decouples tenant system address space and configuration from the
   infrastructure's, which brings a great flexibility for VM placement
   and mobility. The technology results the transit nodes in the
   infrastructure not aware of the existence of the virtual networks.
   One tunnel may carry the traffic belonging to different virtual
   networks; a virtual network identifier is used for traffic
   demultiplexing.

   A virtual network may be an L2 or L3 domain. The TSs attached to an
   NVE can belong to different virtual networks that are either in L2
   or L3. A virtual network can carry unicast traffic and/or
   broadcast/multicast/unknown traffic from/to tenant systems. There
   are several ways to transport virtual network BUM traffic
   [NVO3MCAST].

  It is worth to mention two distinct cases regarding to NVE location.
  The first is that TSs and an NVE are co-located on a same end device,
  which means that the NVE can be aware of the TS state at any time
  via internal API. The second is that TSs and an NVE reside on
  different devices that connect via a wire; in this case, a protocol
  is necessary for NVE to know TS 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 reachbility update mechanisms need be fast enough so that the
  updates do not cause any service interruption. The capability of

supporting many TSs in a virtual network and many more virtual
networks in a DC is critical for NVO3 solution.

If a virtual network spans across multiple DC sites, one design is
to allow the network seamlessly to 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 of 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

For customers (an enterprise or individuals) who utilize DC
provider's compute and storage resources to run their applications,
they 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 the connectivity to all
the resources designated for a customer and allows the customer to
access their 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 Internet or WANs. Two use cases are
described here.

3.1. DC Virtual Network Access via Internet

A customer can connect to a DC virtual network via Internet in a
secure way. Figure 1 illustrates this case. A virtual network is
configured on NVE1 and NVE2 and two NVEs are connected via an IP
tunnel in the Data Center. A set of tenant systems are attached to
NVE1 on a server. The 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. A customer can
access their systems, i.e., TS1 or TSn, in the DC via Internet by
using IPsec tunnel [RFC4301]. The IPsec tunnel is configured between
the vGW and the customer gateway at customer site. Either static
route or iBGP may be used for routes update. The vGW provides IPsec
functionality such as authentication scheme and encryption; iBGP
protocol is carried within the IPsec tunnel. Some vGW features are
listed below:

o Some vGW functions such as firewall and load balancer can be
   performed by locally attached network appliance devices.

o The virtual network in DC may use different address space than
   external users, then vGW needs to provide the NAT function.

   o More than one IPsec tunnels can be configured for the redundancy.

   o vGW can be implemented on a server or VM. In this case, IP
      tunnels or IPsec tunnels can be used over DC infrastructure.

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


```
     Server+---------------+
           |   TS1 TSn     |
           |    |...|       |
           |  +-+---+-+     |              Customer Site
           |  |  NVE1 |     |               +-----+
           |  +---+---+     |               | CGW |
           +------+--------+                +--+--+
                  |                            *
            L3 Tunnel                          *
                  |                            *
     DC GW +------+---------+           .--.   .--.
           |  +---+---+     |          (     '*    '.--.
           |  |  NVE2 |     |       .-.'    *           )
           |  +---+---+     |      (      *  Internet    )
           |  +---+---+.    |        ( *              /
           |  |  vGW  | * * * * * * * * '_'          '_'
           |  +-------+ |   | IPsec        \../ \.--/'
           |   +--------+   | Tunnel
           +---------------+
```
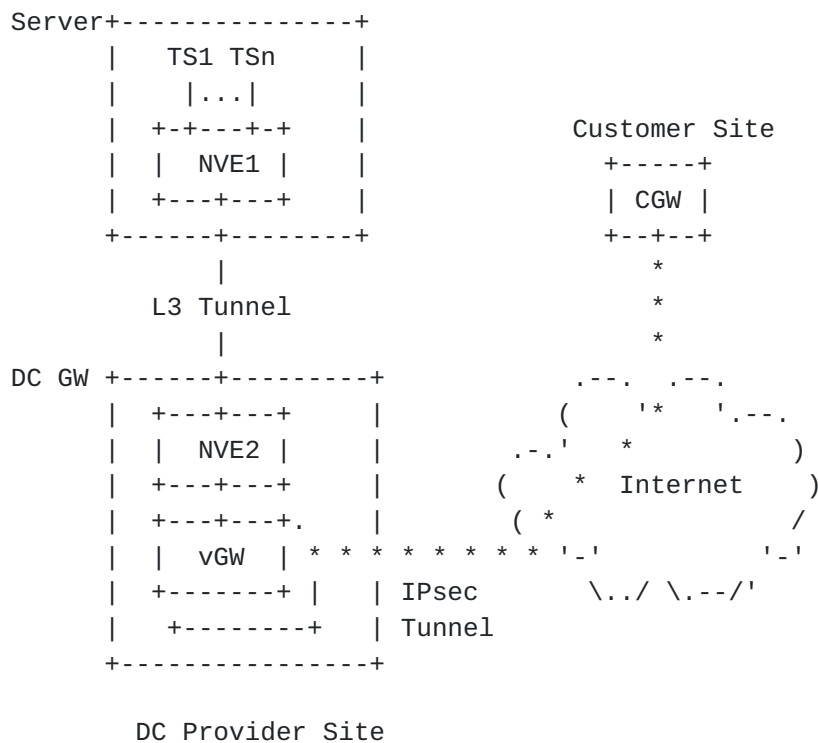
             DC Provider Site

           Figure 1 - DC Virtual Network Access via Internet

. **DC VN and SP WAN VPN Interconnection**

   In this case, an Enterprise customer wants to use Service Provider
   (SP) WAN VPN [RFC4364] [RFC7432] to interconnect its sites and a
   virtual network in DC site. Service Provider constructs a VPN for
   the enterprise customer. Each enterprise site peers with a SP PE.
   The DC Provider and VPN Service Provider can build a DC virtual
   network (VN) and VPN independently and interconnects the VN and VPN
   via a local link or a tunnel between DC GW and WAN PE devices. The
   control plan interconnection options between the VN and VPN are
   described in RFC4364 [RFC4364]. In Option A with VRF-LITE [VRF-LITE],
   both ASBRs, i.e., DC GW and SP PE, maintain a routing/forwarding

   table, and perform the table lookup in forwarding. In Option B, DC
   ASBR and SP ASBR do not maintain the forwarding table, it only
   maintains the VN and VPN identifier mapping, and swap the
   identifiers on the packet in the forwarding process. Both option A
   and B requires tunnel termination. In option C, the VN and VPN use
   the same identifier, and Both ASBRs perform the tunnel stitching,
   i.e., change the tunnel end points. Each option has pros/cons (see
   RFC4364) and has been deployed in SP networks depending on the
   applications. The BGP protocols can be used in these options for
   route distribution. Note that if the DC is the SP 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.

   This configuration allows the enterprise networks communicating to
   the tenant systems attached to the VN in a DC provider site without
   interfering with DC provider underlying physical networks and other
   virtual networks in the DC. The enterprise can use its own address
   space in the VN. The DC provider can manage which VM and storage
   attaching to the VN. The enterprise customer manages what
   applications to run on the VMs in the VN without the knowledge of
   VMs location 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 awareness, i.e., no impact on the
   enterprise 'live' applications running these resources. Such
   advanced technologies bring DC providers great benefits in offering
   cloud applications but add some requirements for NVO3 [RFC7364] as
   well.

4. DC Applications Using NVO3

   NVO3 technology brings DC operators the flexibility in designing and
   deploying different applications in an end-to-end virtualization
   overlay environment, where the operators no longer need to worry
   about the constraints of the DC physical network configuration when
   creating VMs and configuring a virtual network. DC provider may use
   NVO3 in various ways and also use it in the conjunction with other
   physical networks in DC for a reason. This section just highlights
   some use cases.

4.1. Supporting Multiple Technologies and Applications

   Most likely servers deployed in a large data center are rolled in at
   different times and may have different capacities/features. Some
   servers may be virtualized, some may not; some may be equipped with
   virtual switches, some may not. For the ones equipped with

   hypervisor based virtual switches, some may support VxLAN [RFC7348]
   encapsulation, some may support NVGRE encapsulation [NVGRE], and
   some may not support any types of encapsulation. To construct a
   tenant network among these servers and the ToR switches, operators
   can construct one NVO3 virtual network and one traditional VLAN
   network; or two virtual networks that one uses VxLAN encapsulation
   and another uses NVGRE.

   In these cases, a gateway device or virtual GW is used to
   participate in two virtual networks. It performs the packet
   encapsulation/decapsulation translation and may also perform address
   translation, and etc.

   A data center may be also constructed with multi-tier zones. 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)
   with service applications such as payment and booking, 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,
   the communication between the zones must pass through the security
   GW/firewall. One virtual network can be configured in each zone and
   a GW is used to interconnect two virtual networks, i.e., two zones.
   If individual zones use 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 tenant networks.
   The inter-subnet policies may be placed at some designated gateway
   devices only. Such design requires the inter-subnet traffic to be
   sent to one of the gateways first for the policy checking, which may
   cause traffic hairpin at the gateway in a DC. It is desirable that
   an NVE can hold some policies and be able to forward inter-subnet
   traffic directly. To reduce NVE burden, the hybrid design may be
   deployed, i.e., an NVE can perform forwarding for the selected
   inter-subnets and the designated GW performs for the rest. For
   example, each NVE performs inter-subnet forwarding for a tenant, and
   the designated GW is used for inter-subnet traffic from/to the
   different tenant networks.

   A tenant network may span across multiple Data Centers that are in
   difference 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 DC GW
   device. Figure 2 illustrates this configuration.

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

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

```
NVE5/DCGW+------------+                    +-----------+ NVE6/DCGW
        | +-----+    | '''''''''''''' |    +-----+ |
        | |L3VNI+----+'    L3VNx      '+---+L3VNI| |
        | +--+--+    | '''''''''''''' |    +--+--+ |
        |    |VIRB   |                |  VIRB|    |
        | +--+---+   |                |  +---+--+ |
        | |L2VNIs|   |                |  |L2VNIs| |
        | +--+---+   |                |  +---+--+ |
        +----+-------+                +------+----+
         ''''|''''''''''                ''''''|'''''''
         '      L2VNa      '            '      L2VNz     '
   NVE1/S ''/''''''''\'' NVE2/S    NVE3/S'''/''''''\'' NVE4/S
      +-----+---+  +----+----+        +------+--+ +----+----+
      | +--+--+ |  | +--+--+ |        | +---+-+ | | +--+--+ |
      | |L2VNI| |  | |L2VNI| |        | |L2VNI| | | |L2VNI| |
      | ++---++ |  | ++---++ |        | ++---++ | | ++---++ |
      +--+---+--+  +--+---+--+        +--+---+--+ +--+---+--+
       |...|        |...|             |...|        |...|


        Tenant Systems              Tenant Systems

          DC Site A                   DC Site Z
```
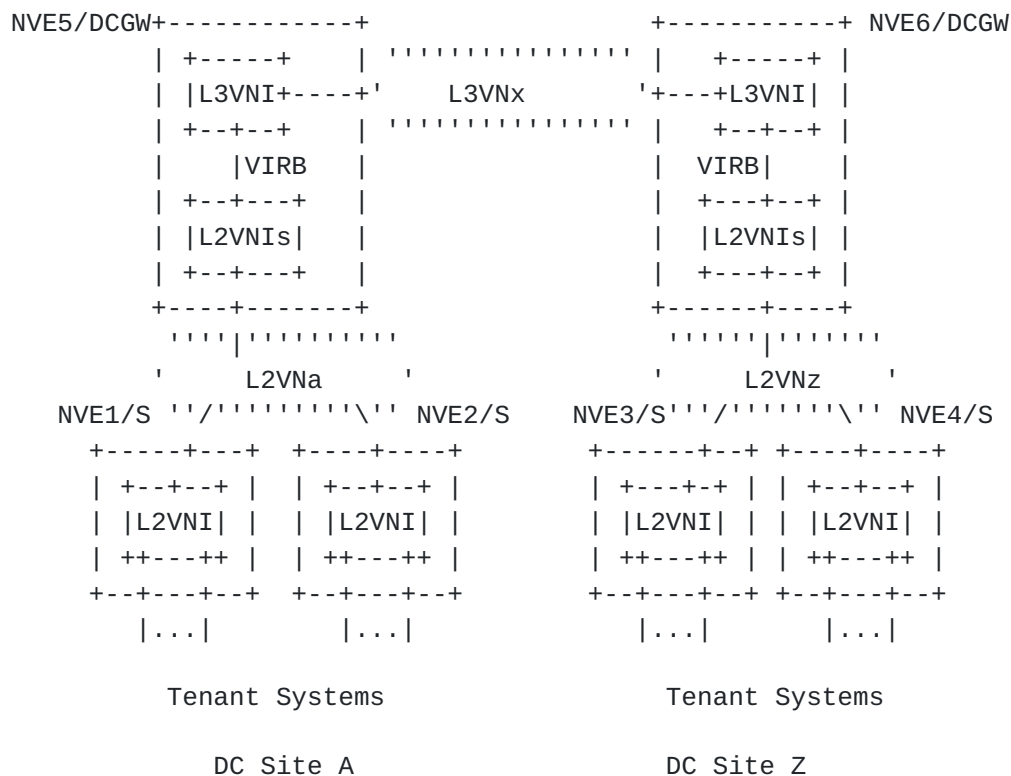
Figure 2 - Tenant Virtual Network with Bridging/Routing

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 DC infrastructure and offer
   a virtual DC service to enterprise customers. A vDC at DC Provider
   site provides the same capability as a physical DC at the customer
   site. A customer manages what and how applications to run in its
   vDC. DC Provider can further offer different network service
   functions to a vDC. The network service functions may include
   firewall, DNS, load balancer, gateway, and etc.

   Figure 3 below illustrates one scenario. For the simple
   illustration, it only shows the L3 VN or L2 VN in abstraction. In
   this example, DC Provider operators create several L2 VNs (L2VNx,
   L2VNy, L2VNz) to group the tenant systems together per application
   basis, create one L3 VN, e.g., VNa for the internal routing. A
   network function, firewall and gateway, runs on a VM or server that
   connects to the L3VNa and is used for inbound and outbound traffic
   process. A load balancer (LB) is used in L2 VNx. A VPN is also built
   between the gateway and enterprise router. Enterprise customer runs
   Web/Mail/Voice applications on VMs at the provider DC site that can
   spread out on many servers; the users at 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.

   Enterprise customer decides which applications are accessed by
   intranet only and which by both intranet and extranet and configures
   the proper security policy and gateway function at firewall/gateway.
   Furthermore an enterprise customer may want multi-zones in a vDC
   (See section 4.1) for the security and/or set different QoS levels
   for the different applications.

   The vDC use case requires the NVO3 solution to provide the DC
   operators 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
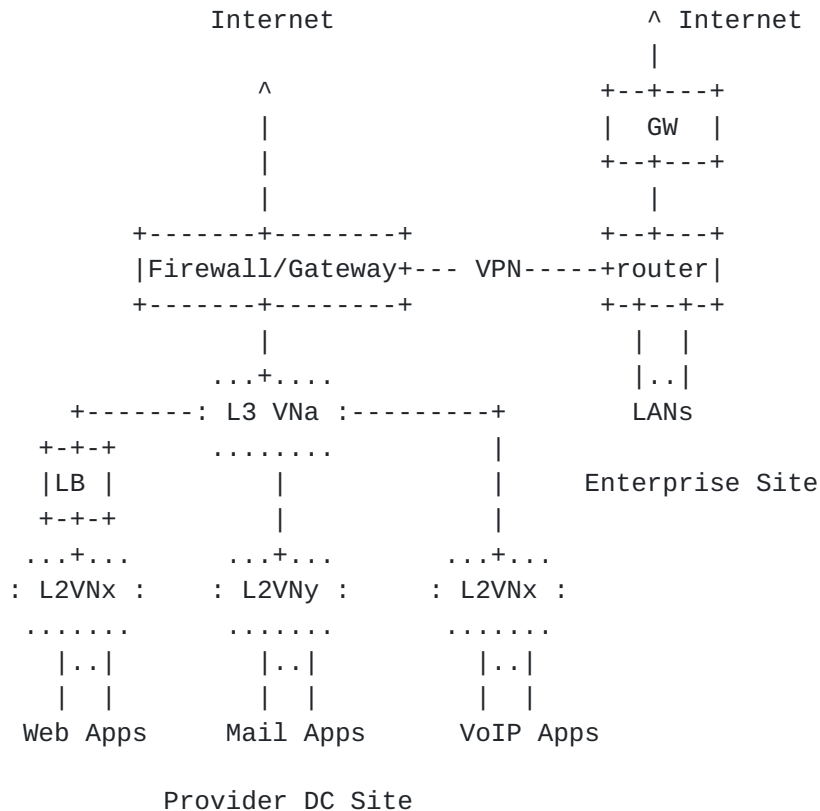   in the vDC via the vDC topology.

```
                     Internet                      ^ Internet
                                                   |
                         ^                     +--+---+
                         |                     |  GW  |
                         |                     +--+---+
                         |                        |
              +-------+--------+               +--+---+
              |Firewall/Gateway+--- VPN-----+router|
              +-------+--------+               +-+--+-+
                      |                        |  |
                  ...+....                     |..|
              +-------: L3 VNa :----------+      LANs
            +-+-+      ........            |
            |LB |        |                 |    Enterprise Site
            +-+-+        |                 |
          ...+...     ...+...       ...+...
          : L2VNx :   : L2VNy :     : L2VNx :

          .......     .......       .......
            |..|        |..|          |..|
            |  |        |  |          |  |
          Web Apps   Mail Apps    VoIP Apps


              Provider DC Site
```

Figure 3 - Virtual Data Center (vDC)

5. Summary

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

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

NVO3 uses tunnel technique so that two NVEs appear as one hop to
each other in a virtual network. Many tunneling technologies can
serve this function. The tunneling may in turn be tunneled over
other intermediate tunnels over the Internet or other WANs.

A DC virtual network 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
centralized controller to manage tenant system reachbility in a
virtual network, other prefer to use distributed 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 a
secured virtual network, which means one tenant's traffic isolated
from other tenant's traffic and non-tenant's traffic; 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
infrastructure network. For example, a tenant application attempts
to generate a large volume of traffic to overload DC underlying
network. The NVO3 solution has to address these issues.

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.

   [NVGRE]   Sridharan, M., "NVGRE: Network Virtualization using Generic
             Routing Encapsulation", draft-sridharan-virtualization-
             nvgre-07, 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.

   [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

   [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

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