INTERNET-DRAFT   Luyuan Fang
Intended Status: Standards track   David Ward
Expires: April 18, 2014   Rex Fernando
                                                             Cisco
Ning So   Maria Napierala
TATA Communications   AT&T
Jim Guichard   Nabil Bitar
Cisco   Verizon
Wen Wang   Dhananjaya Rao
CenturyLink   Cisco
Manuel Paul   Bruno Rijsman
Deutsche Telekom   Juniper

                                                  October 18, 2013

                      BGP/MPLS IP VPN Virtual PE
                    draft-fang-l3vpn-virtual-pe-04

Abstract

   This document describes the architecture solutions for BGP/MPLS IP
   Virtual Private Networks (VPNs) with virtual Provider Edge (vPE)
   routers. It provides a functional description of the vPE control,
   forwarding, and management. The proposed vPE solutions support both
   the Software Defined Networks (SDN) approach which allows physical
   decoupling of the control and the forwarding, and the traditional
   distributed routing approach. A vPE can reside in any network or
   compute devices, such as a server as co-resident with the application
   virtual machines (VMs), or a Top-of-Rack (ToR) switch in a Data
   Center (DC) network.

The list of current Internet-Drafts can be accessed at
http://www.ietf.org/1id-abstracts.html

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

Table of Contents

1 Introduction

Network virtualization enables multiple isolated individual networks over a shared common network infrastructure. BGP/MPLS IP Virtual Private Networks (IP VPNs) [RFC4364] have been widely deployed to provide network based Layer 3 VPNs solutions. [RFC4364] provides routing isolation among different customer VPNs and allow address overlapping among these VPNs through the implementation of per VPN Virtual Routing and Forwarding instances (VRFs) at a Service Provider Edge (PE) routers, while forwarding customer traffic over a common IP/MPLS network.

With the advent of compute capabilities and the proliferation of virtualization in Data Center servers, multi-tenant Data Centers are becoming the norm. As applications and appliances are increasingly being virtualized, support for virtual edge devices, such as virtual IP VPN PE routers, becomes feasible and desirable for Service Providers who want to extend their existing IP VPN deployments into Data Centers to provide end-to-end Virtual Private Cloud (VPC) services. Virtual PE work is also one of early effort for Network Functions Virtualization (NFV). In general, scalability, agility, and cost efficiency are primary motivations for vPE solutions.

The virtual Provider Edge (vPE) solution described in this document allows for the extension of the PE functionality of IP VPN to an end device, such as a server where the applications reside, or to a first hop routing/switching device, such as a Top of the Rack (ToR) switch in a DC.

The vPE solutions support both the Software Defined Networks (SDN) approach, which allows physical decoupling of the control and the forwarding, and the traditional distributed routing approach.

1.1 Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Term	Definition
ASBR	Autonomous System Border Router
BGP	Border Gateway Protocol
CE	Customer Edge
Forwarder	IP VPN forwarding function
GRE	Generic Routing Encapsulation
Hypervisor	Virtual Machine Manager

```
     I2RS              Interface to Routing Systems
     LDP               Label Distribution Protocol
     MP-BGP            Multi-Protocol Border Gateway Protocol
     PCEF              Policy Charging and Enforcement Function
     QoS               Quality of Service
     RR                Route Reflector
     RT                Route Target
     RTC               RT Constraint
     SDN               Software Defined Networks
     ToR               Top-of-Rack switch
     VI                Virtual Interface
     vCE               virtual Customer Edge Router
     VM                Virtual Machine
     vPC               virtual Private Cloud
     vPE               virtual Provider Edge Router
     vPE-C             virtual Provider Edge Control plane
     vPE-F             virtual Provider Edge Forwarder
     VPN               Virtual Private Network
     vRR               virtual Route Reflector
     WAN               Wide Area Network
```

   End device: where Guest OS, Host OS/Hypervisor, applications, VMs,
   and virtual router may reside.

1.2 Requirements

   The following are key requirements for vPE solutions.

   1) MUST support end device multi-tenancy, per tenant routing
   isolation and traffic separation.

   2) MUST support large scale IP VPNs in the Data Center, upto tens of
   thousands of end devices and millions of VMs in the single Data
   Center.

   3) MUST support end-to-end IP VPN connectivity, e.g. IP VPN can start
   from a DC end device, connect to a corresponding IP VPN in the WAN,
   and terminate in another Data Center end device.

   4) MUST allow physical decoupling of IP VPN PE control and forwarding
   for network virtualization and abstraction.

   5) MUST support the control plane with both SDN controller approach,
   and the traditional distributed control plane approach with MP-BGP
   protocol.

   6) MUST support VM mobility.

7) MUST support orchestration/auto-provisioning deployment model.

8) SHOULD be capable to support service chaining as part of the
   solution [I-D.rfernando-l3vpn-service-chaining],
   [I-D.bitar-i2rs-service-chaining].

The architecture and protocols defined in BGP/MPLS IP VPN [RFC4364]
provide the foundation for vPE extension. Certain protocol extensions
may be needed to support the virtual PE solutions.

2. Virtual PE Architecture

2.1 Virtual PE definitions

As defined in [RFC4364], an IP VPN is created by applying policies to
form a subset of sites among all sites connected to the backbone
networks. It is a collection of "sites". A site can be considered as
a set of IP systems maintaining IP inter-connectivity without direct
connecting through the backbone. The typical use of L3VPM has been to
inter-connect different sites of an Enterprise networks through a
Service Provider's BGP IP VPNs in the WAN.

A virtual PE (vPE) is a BGP/MPLS IP VPN PE software instance which
may reside in any network or computing devices. The control and
forwarding components of the vPE can be decoupled, they may reside in
the same physical device, or most often in different physical
devices.

A virtualized Provider Edge Forwarder (vPE-F) is the forwarding
element of a vPE. vPE-F can reside in an end device, such as a server
in a Data Center where multiple application Virtual Machines (VMs)
are supported, or a Top-of-Rack switch (ToR) which is the first hop
switch from the Data Center edge. When a vPE-F is residing in a
server, its connection to a co-resident VM is as the same as the PE-
CE relationship in the regular BGP IP VPNs, but without routing
protocols or static routing between the virtual PE and CE because the
connection is internal to the device.

The vPE Control plane (vPE-C) is the control element of a vPE. When
using the approach where control plane is decoupled from the physical
topology, the vPE-F may be in a server and co-resident with
application VMs, while one vPE-C can be in a separate device, such as
an SDN Controller where control plane elements and orchestration
functions are located. Alternatively, the vPE-C can reside in the
same physical device as the vPE-F. In this case, it is similar to the
traditional implementation of VPN PEs where, distributed MP-BGP is
used for IP VPN information exchange, though the vPE is not a
dedicated physical entity as it is in a physical PE implementation.

[2.2](#) vPE Architecture and Design options

[2.2.1](#) vPE-F host location

   Option 1a. vPE-F is on an end device as co-resident with application
   VMs. For example, the vPE-F is on a server in a Data Center.

   Option 1b. vPE-F forwarder is on a ToR or other first hop devices in
   a DC, not as co-resident with the application VMs.

   Option 1c. vPE-F is on any network or compute devices in any types of
   networks.

[2.2.2](#) vPE control plane topology

   Option 2a. vPE control plane is physically decoupled from the vPE-F.
   The control plane may be located in a controller in a separate device
   (a stand alone device or can be in the gateway as well) from the vPE
   forwarding plane.

   Option 2b. vPE control plane is supported through dynamic routing
   protocols and located in the same physical device as the vPE-F.

[2.2.3](#) Data Center orchestration models

   Option 3a. Push model: It is a top down approach, push IP VPN
   provisioning state from a network management system or other
   centrally controlled provisioning system to the IP VPN network
   elements.

   Option 3b. Pull model: It is a bottom-up approach, pull state
   information from network elements to network management/AAA based
   upon data plane or control plane activity.

[2.3](#) vPE Architecture reference models

[2.3.1](#) vPE-F in an end-device and vPE-C in the controller

   Figure 1 illustrates the reference model for a vPE solution with the
   vPE-F in the end device co-resident with applications VMs, while the
   vPE-C is physically decoupled and residing on a controller.

   The Data Center is connected to the IP/MPLS core via the
   Gatways/ASBRs. The IP VPN , e.g. VPN RED, has a single termination
   point within the DC at one of the VPE-F, and is inter-connected in
   the WAN to other member sites which belong to the same client, and
   the remote ends of VPN RED can be a PE which has VPN RED attached to
   it, or another vPE in a different Data Center.

Note that the DC fabrics/intermediate underlay devices in the DC do
not participate IP VPNs, their function is the same as provider
backbone routers in the IP/MPLS back bone and they do not maintain
the IP VPN states, nor IP VPN aware.

```
                         ,-----.
                        (       ')
                     .--(.       '.---.
                    (     '     '      )
                    (      IP/MPLS WAN    )
                     (.                 .)
                      (     (       .)
        WAN            ''--' '-''---'
        ----------------|----------|-----------------------
        Service/DC      |          |
        Network     +-------+   +-------+
                    |Gateway|---|Gateway|   *
                    | /ASBR |   | /ASBR |      *
                    +-------+   +-------+        *
                        |           |      +-------------+
                        |     ,---. |      |Controller   |
                      .---. (       '.---. |(vPE-C and   |
                     (     '      '       ') |orchestrator)|
                     (       Data Center     ) +-------------+
                      (.        Fabric       )         *
                       (      (        ).--'           *
                      /    ''--' '-''--'      \        *
                    /      /            \      \     *
          +-------+   +-------+   +-------+   +-------+
          | vPE-F |   | vPE-F |   | vPE-F |   | vPE-F |
          +---+---+   +---+---+   +---+---+   +---+---+
          |VM |VM |   |VM |VM |   |VM |VM |   |VM |VM |
          +---+---+   +---+---+   +---+---+   +---+---+
          |VM |VM |   |VM |VM |   |VM |VM |   |VM |VM |
          +---+---+   +---+---+   +---+---+   +---+---+

          End Device  End Device  End Device  End Device
```
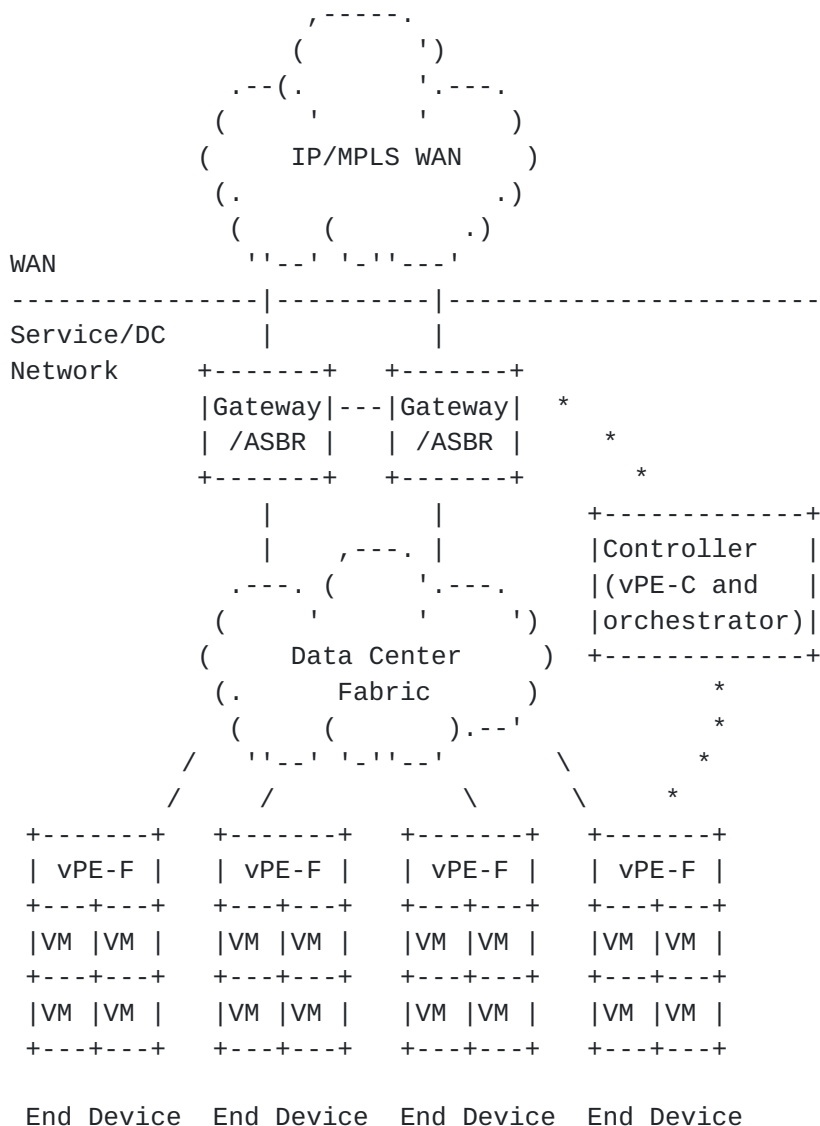
        Figure 1. Virtualized Data Center with vPE at
     the end device and vPE-C and vPE-F physically decoupled

Note:

a) *** represents Controller logical connections to the all
Gateway/ASBRs and to all vPE-F.

b) ToR is assumed included in the Data Center cloud.

2.3.2 vPE-F and vPE-C on the same end-device

In this option, vPE-F and vPE-C functionality are both resident in
the end-device. The vPE functions the same as it is in a physical PE.
MP-BGP is used for the VPN control plane. Virtual or physical Route
Reflectors (RR) (not shown in the diagram) can be used to assist
scaling.

```
                            ,-----.
                           (       ')
                        .--(.       '.---.
                       (      '       '    )
                       (      IP/MPLS WAN    )
                        (.                 .)
                         (      (       .)
            WAN           ''--' '-''---'
            ----------------|----------|----------------------
            Service/DC      |          |
            Network     +-------+   +-------+
                        |Gateway|---|Gateway|
                        | /ASBR |   | /ASBR | *
                        +-------+   +-------+   *
                            |           |         * MP-BGP
                            |    ,---. |              *
                         .---. (       '.---.           *
                        (      '       '    ')           *
                        (      Data Center    )            *
                         (.      Fabric        )            *
                          (      (       ).--'             *
                        /    ''--' '-''--'        \        *
                       /    /            \     \      *
              +-------+   +-------+   +-------+   +-------+
              |  vPE  |   |  vPE  |   |  vPE  |   |  vPE  |
              +---+---+   +---+---+   +---+---+   +---+---+
              |VM |VM |   |VM |VM |   |VM |VM |   |VM |VM |
              +---+---+   +---+---+   +---+---+   +---+---+
              |VM |VM |   |VM |VM |   |VM |VM |   |VM |VM |
              +---+---+   +---+---+   +---+---+   +---+---+

              End Device  End Device  End Device  End Device

              Figure 2. Virtualized Data Center with vPE at
              the end device, VPN control signal uses MP-BGP
```

Note:

a) *** represents the logical connections using MP-BGP among the
Gateway/ASBRs and to the vPEs on the end devices.

   b) ToR is assumed included in the Data Center cloud.

[2.3.3](#) vPE-F and vPE-C are on the ToR

   In this option, vPE functionality is the same as a physical PE. MP-
   BGP is used for the VPN control plane. Virtual or physical Route
   Reflector (RR) (not shown in the diagram) can be used to assist
   scaling.

```
                           ,-----.
                          (       ')
                        .--(.        '.---.
                       (      '     '      )
                       (      IP/MPLS WAN    )
                        (.                 .)
                         (     (        .)
          WAN             ''--' '-''---'
          ----------------|----------|---------------------
          Service/DC      |          |
          Network     +-------+   +-------+
                      |Gateway|---|Gateway|
                      | /ASBR |   | /ASBR | *
                      +-------+   +-------+   *
                          |           |        *  MP-BGP
                          |     ,---. |          *
                        .---. (       '.---.       *
                       (      '     '     ')        *
                       (      Data Center    )       *
                        (.      Fabric       )      *
                         (      (        ).--'      *
                        /''--' '-/'--'     \  *
                  +---+---+  +---+---+  +---+---+
                  |vPE|   |  |vPE|   |  |vPE|   |
                  +---+   |  +---+   |  +---+   |
                  | ToR   |  | ToR   |  | ToR   |
                  +-------+  +-------+  +-------+
                      /    \   /    \   /    \
              +-------+   +-------+   +-------+   +-------+
              |  vPE  |   |  vPE  |   |  vPE  |   |  vPE  |
              +---+---+   +---+---+   +---+---+   +---+---+
              |VM |VM |   |VM |VM |   |VM |VM |   |VM |VM |
              +---+---+   +---+---+   +---+---+   +---+---+
              |VM |VM |   |VM |VM |   |VM |VM |   |VM |VM |
              +---+---+   +---+---+   +---+---+   +---+---+
              End Device  End Device  End Device  End Device
```
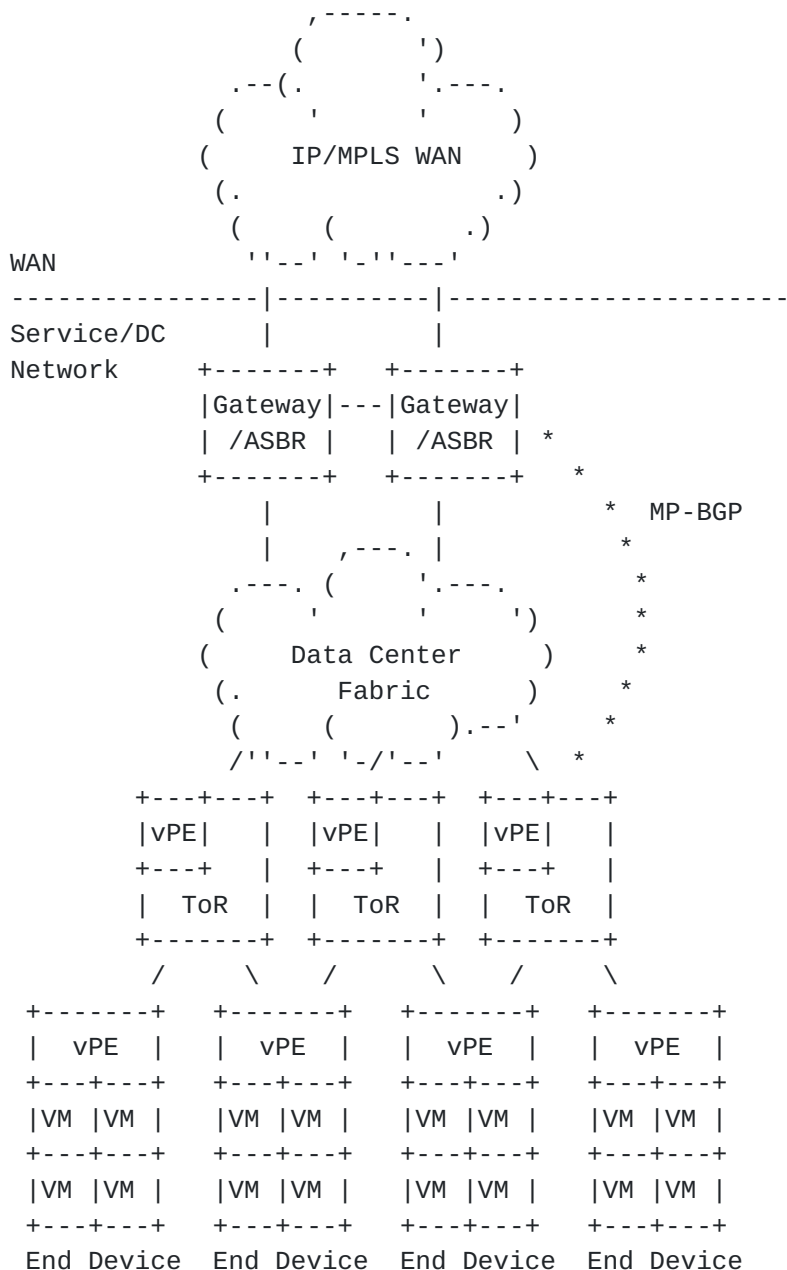
          Figure 3. Virtualized Data Center with vPE at
          the ToP, VPN control signal uses MP-BGP

Note: *** represents the logical connections using MP-BGP among the
Gateway/ASBRs and to the vPEs on the ToRs.

2.3.4 vPE-F on the ToR and vPE-C on the controller

In this option, the L3VPN termination is at the ToR, but the control
plane decoupled from the data plane and resided in a controller,
which can be on a stand alone device, or can be placed at the
Gateway/ASBR.

2.3.5 The server view of a vPE

An end device shown in Figure 4 is a virtualized server that hosts
multiple VMs. The virtual PE is co-resident in the server with
application VMs. The vPE supports multiple VRFs, VRF Red, VRF Grn,
VRF Yel, VRF Blu, etc. Each application VM is associated to a
particular VRF as a member of the particular VPN. For example, VM1 is
associated to VRF Red, VM2 and VM47 are associated to VRF Grn, etc.
Routing isolation applies between VPNs for multi-tenancy support. For
example, VM1 and VM2 cannot communicate directly in a simple intranet
L3VPN topology as shown in the configuration.

The vPE connectivity relationship between vPE and the application VM
is similar to the PE-to-CE relationship in a regular BGP IP VPNs.
However, as the vPE and CE functions are co-resident in the same
server, the connection between them is an internal implementation of
the server.

```
          +-----------------------------------------------------+
          | +---------+ +---------+    +---------+ +---------+ |
          | |  VM1    | |  VM2    |    | VM47    | | VM48    | |
          | |(VPN Red)| |(VPN Grn)|... |(VPN Grn)| |(VPN Blu)| |
          | +----+----+ +---+-----+    +----+----+ +----+----+ |
          |      |         |                |           |       |
          |    +---+       | +------------+        +---+        |
          |      |         | |  | |                  |          |
   to     |    +---+------+-+------------------+---+          |
   Gateway|    |   |    | |                        |  |          |
   PE     |    | +-+-+  ++-++            +---+ +-+-+ |          |
          |    | |VRF|  |VRF|  .......   |VRF| |VRF| |          |
   <------+------+ |Red|  |Grn|            |Yel| |Blu| |          |
          |    | +---+  +---+            +---+ +---+ |          |
          |    |           L3 VPN virtual PE        |  |          |
          |    +------------------------------------+          |
          |                                                     |
          |                   End Device                        |
          +-----------------------------------------------------+
```
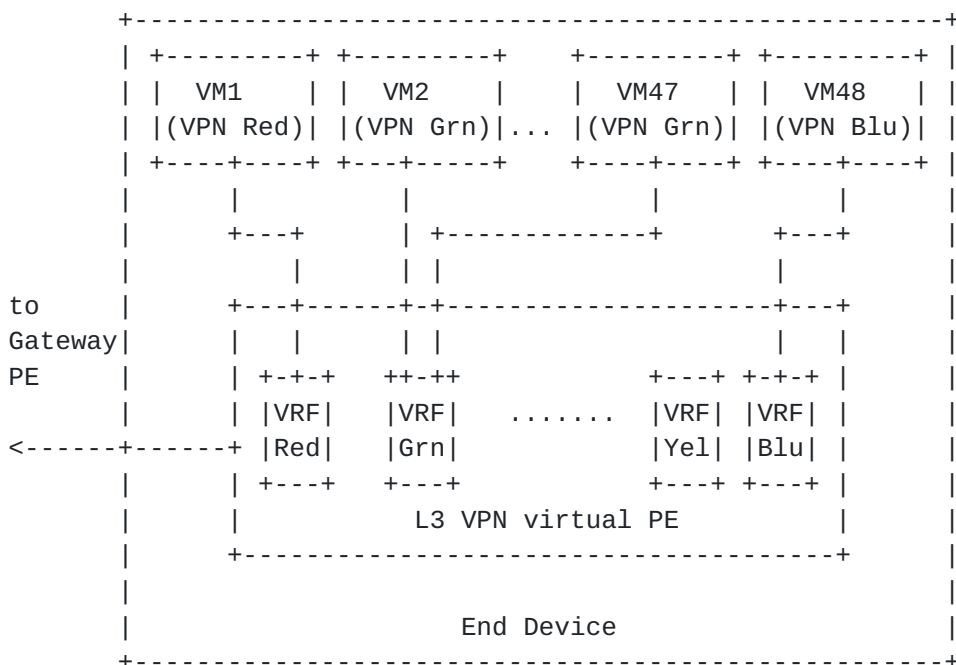
Figure 4. Server View of vPE to VM relationship

   An application VM may send packets to a vPE forwarder that need to be
   bridged, either locally to another VM, or to a remote destination. In
   this case, the vPE contains a virtual bridge instance to which the
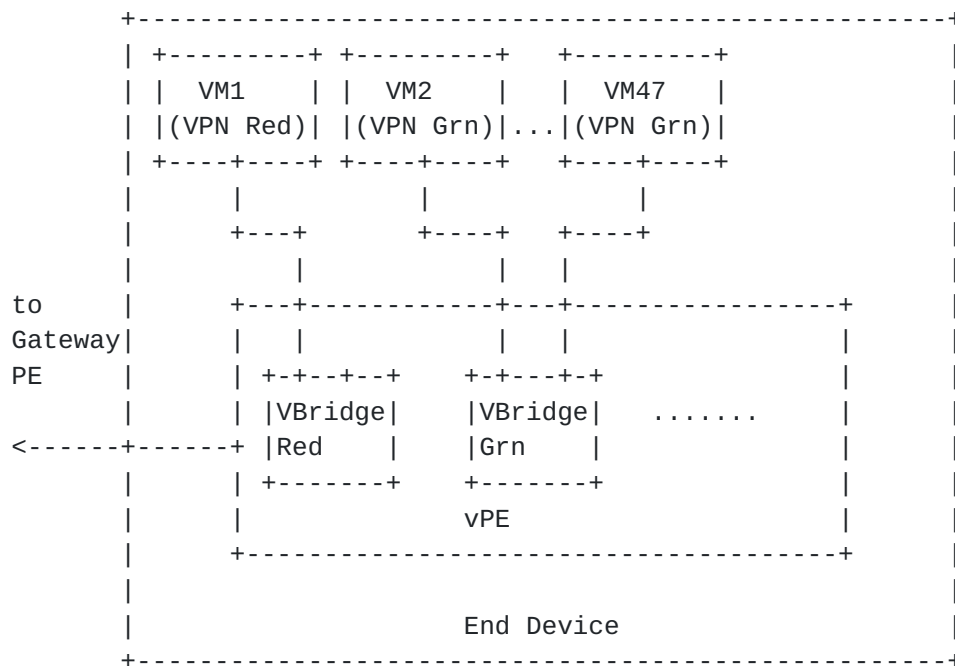   application VMs (CEs) are attached.

```
            +------------------------------------------------------+
            | +---------+ +---------+   +---------+                 |
            | |  VM1    | |  VM2    |   |  VM47   |                 |
            | |(VPN Red)| |(VPN Grn)|...|(VPN Grn)|                 |
            | +----+----+ +----+----+   +----+----+                 |
            |      |           |             |                      |
            |    +---+      +----+   +----+                         |
            |      |        |    |   |                              |
     to     |    +---+-----------+---+----------------+             |
    Gateway|     |    |           |    |              |             |
    PE      |    | +-+--+--+   +-+----+-+             |             |
            |    | |VBridge|   |VBridge|   .......     |             |
    <------+------+ |Red    |   |Grn    |              |             |
            |    | +-------+   +-------+              |             |
            |    |               vPE                 |             |
            |    +------------------------------------+             |
            |                                                       |
            |               End Device                             |
            +------------------------------------------------------+
```

Figure 4. Bridging Service at vPE

3. Control Plane

3.1 vPE Control Plane (vPE-C)

3.1.1 The SDN approach

   This approach is appropriate when the vPE control and data planes are
   physically decoupled. The control plane directing the data flow may
   reside elsewhere, e.g. in a SDN controller. This approach requires a
   standard interface to the routing system. The Interface to Routing
   System (I2RS) is work in progress in IETF as described in
   [I-D.ietf-i2rs-architecture], [I-D.ietf-i2rs-problem-statement].

   Although MP-BGP is often the de facto preferred choice between vPE
   and gateway-PE/ASBR, the use of extensible signaling messaging
   protocols MAY often be more practical in a Data Center environment.
   One such proposal that uses this approach is detailed in
   [I-D.ietf-l3vpn-end-system].

3.1.2 Distributed control plane

In the distributed control plane approach, the vPE participates in the overlay L3VPN control protocol: MP-BGP [RFC4364].

When the vPE function is on a ToR, it participates the underlay routing through IGP protocols: ISIS or OSPF.

When the vPE function is on a server, it functions as a host attached to a server.

3.3 Use of router reflector

Modern Data Centers can be very large in scale. For example, the number of VPNs routes in a very large DC can surpass the scale of those in a Service Provider backbone VPN networks. There may be tens of thousands of end devices in a single DC.

Use of Router Reflector (RR) is necessary in large-scale IP VPN networks to avoid a full iBGP mesh among all vPEs and PEs. The IP VPN routes can be partitioned to a set of RRs, the partitioning techniques are detailed in [RFC4364].

When a RR software instance is residing in a physical device, e.g., a server, which is partitioned to support multi-functions and application VMs, the RR becomes a virtualized RR (vRR). Since RR performs control functions only, a dedicated or virtualized server with large scale of computing power and memory can be a good candidate as host of vRRs. The vRR can also reside in a Gateway PE/ASBR, or in an end device.

3.4 Use of Constrained Route Distribution [RFC4684]

The Constrained Route Distribution [RFC4684] is a powerful tool for selective IP VPN route distribution. With RTC, only the BGP receivers (e.g, PE/vPE/RR/vRR/ASBRs, etc.) with the particular IP VPNs attached will receive the route update for the corresponding VPNs. It is critical to use constrained route distribution to support large-scale IP VPN developments.

4. Forwarding Plane

4.1 Virtual Interface

A Virtual Interface (VI) is an interface within an end device that is used for connection of the vPE to the application VMs in the same end device. Such application VMs are treated as CEs in the regular IP VPN's view.

4.2 Virtual Provider Edge Forwarder (vPE-F)

The Virtual Provider Edge Forwarder (vPE-F) is the forwarding
component of a vPE where the tenant identifiers (for example, MPLS
VPN labels) are pushed/popped.

The vPE-F location options include:

1) Within the end device where the virtual interface and application
VMs are located.

2) In an external device such as a Top of the Rack switch (ToR) in a
DC into which the end device connects.

Multiple factors should be considered for the location of the vPE-F,
including device capabilities, overall solution economics,
QoS/firewall/NAT placement, optimal forwarding, latency and
performance, operational impact, etc. There are design tradeoffs, it
is worth the effort to study the traffic pattern and forwarding
looking trend in your own unique Data Center as part of the exercise.

4.3 Encapsulation

There are two existing standardized encapsulation/forwarding options
typically used for BGP/MPLS L3VPN.

   1. MPLS label stack encoding with Label Distribution Protocol
(LDP), [RFC3032][RFC5036].

   2. Encapsulating MPLS packets in IP or Generic Routing
Encapsulation (GRE), [RFC4023], [RFC4797].

   3. Other types of encapsulation are possible. For example, VXLAN
[I-D.mahalingam-dutt-dcops-vxlan], and NVGRE
[I-D.sridharan-virtualization-nvgre] are work in progress in IETF.

The most common BGP/MPLS IP VPN deployments in SP networks use MPLS
forwarding. This requires that an MPLS transport, e.g., Label
Switched Protocol (LDP) [RFC5036] to be deployed in the network. It
is proven to scale, and it comes with various security mechanisms to
protect the network against attacks.

However, the DC environment is different than the SP VPN networks or
large Enterprise backbones. MPLS deployment may or may not be
feasible or desirable. Two major challenges for MPLS deployments
exist in this new environment: 1) the capabilities of the end devices
and the transport/forwarding devices; 2) the workforce skill set.

   Encapsulating MPLS in IP or GRE tunnel [RFC4023] may often be a more
   practical approach for DC and computing environment. Note that when
   IP encapsulation is used, the associated security properties must be
   analyzed carefully.

   In addition, there are new encapsulation proposals for DCs currently
   as work in progress within IETF, including several UDP based encap
   proposals and some TCP based proposal. These overlay encapsulations
   can be suitable alternatives for the vPE solutions.

4.4 Optimal forwarding

   Many large cloud service providers have reported the DC traffic is
   now dominated by East-West across subnet traffic (between the end
   device hosting different applications in different subnets) rather
   than North-South traffic (going in/out of the Data Center and to/from
   the WAN) or switched traffic within subnets. This is the primary
   reason that newer DC design has moved away from traditional Layer-2
   design to Layer-3, especially for the overlay networks.

   When forwarding the traffic within the same VPN, the vPE SHOULD be
   capable to provide direct communication among the VMs/application
   senders/receivers without the need of going through Gateway devices.
   If the senders and the receivers are on the same end device, the
   traffic SHOULD NOT need to leave the device. If they are on different
   end devices, optimal routing SHOULD be applied.

   Extranet IP VPN techniques can be used for multiple VPNs access
   without the need of Gateway facilitation. This is done through the
   use of IP VPN policy control mechanisms.

   In addition, ECMP is a built in IP mechanism for load sharing.
   Optimal use of available bandwidth can be achieved by virtue of using
   ECMP in the underlay, as long as the encapsulation includes certain
   entropy in the header, VXLAN is such an example.

4.5 Routing and Bridging Services

   A VPN forwarder (vPE-F) may support both IP forwarding as well as
   Layer 2 bridging for traffic from attached end hosts. This traffic
   may be between end hosts attached to the same VPN forwarder or to
   different VPN forwarders.

   In both cases, forwarding at a VPN forwarder takes place based on the
   IP or MAC entries provisioned by the vPE controller.

   When the vPE is providing Layer 3 service to the attached CEs, the
   VPN forwarder has a VPN VRF instance with IP routes installed for

both locally attached end-hosts and ones reachable via other VPN
forwarders. The vPE may perform IP routing for all IP packets in this
mode.

When the vPE provides Layer 2 service to the attached end-hosts, the
VPN forwarder has an E-VPN instance with appropriate MAC entries.

The vPE may support an Integrated Routing and Bridging service, in
which case the relevant VPN forwarders will have both MAC and IP
table entries installed, and will appropriately route or switch
incoming packets.

The vPE controller performs the necessary provisioning functions to
support various services, as defined by an user.

5. Addressing

5.1 IPv4 and IPv6 support

IPv4 and IPv6 MUST be supported in the vPE solution.

This may present a challenge for older devices, but this normally is
not an issue for the newer generation of forwarding devices and
servers. Note that a server is replaced much more frequently than a
network router/switch, and newer equipment SHOULD be capable of IPv6
support.

5.2 Address space separation

The addresses used for the IP VPN overlay in a DC, SHOULD be taken
from separate address blocks outside the ones used for the underlay
infrastructure of the DC. This practice is to protect the DC
infrastructure from being attacked if the attacker gains access to
the tenant VPNs.

Similarity, the addresses used for the DC SHOULD be separated from
the WAN backbone addresses space.

6.0 Inter-connection considerations

The inter-connection considerations in this section are focused on
intra-DC inter-connections.

There are deployment scenarios where BGP/MPLS IP VPN may not be
supported in every segment of the networks to provide end-to-end IP
VPN connectivity. A vPE may be reachable only via an intermediate
inter-connecting network; interconnection may be needed in these
cases.

When multiple technologies are employed in the solution, a clear
demarcation should be preserved at the inter-connecting points. The
problems encountered in one domain SHOULD NOT impact other domains.

From an IP VPN point of view: An IP VPN vPE that implements [RFC4364]
is a component of the IP VPN network only. An IP VPN VRF on a
physical PE or vPE contains IP routes only, including routes learnt
over the locally attached network.

The IP VPN vPE should ideally be located as close to the "customer"
edge devices as possible. When this is not possible, simple existing
"IP VPN CE connectivity" mechanisms should be used, such as static,
or direct VM attachments such as described in the vCE
[I-D.fang-l3vpn-virtual-ce] option below.

Consider the following scenarios when BGP MPLS VPN technology is
considered as whole or partial deployment:

Scenario 1: All VPN sites (CEs/VMs) support IP connectivity. The most
suited BGP solution is to use IP VPNs [RFC4364] for all sites with PE
and/or vPE solutions.

Scenario 2: Legacy Layer 2 connectivity must be supported in certain
sites/CEs/VMs, and the rest of the sites/CEs/VMs need only Layer 3
connectivity.

One can consider using a combined vPE and vCE
[I-D.fang-l3vpn-virtual-ce] solution to solved the problem. Use IP
VPN for all sites with IP connectivity, and a physical or virtual CE
(vCE, may reside on the end device) to aggregate the Layer 2 sites
which for example, are in a single container in a Data Center. The
CE/vCE can be considered as inter-connecting points, where the Layer
2 network is terminated and the corresponding routes for connectivity
of the L2 network are inserted into IP VPN VRFs. The Layer 2 aspect
is transparent to the L3VPN in this case.

Reducing operation complicity and maintaining the robustness of the
solution are the primary reasons for the recommendations.

7. Management, Control, and Orchestration

7.1 Assumptions

The discussion in this section is based on the following set of
assumptions:

- The WAN and the inter-connecting Data Center, MAY be under control
of separate administrative domains

- WAN Gateways/ASBRs/PEs are provisioned by existing WAN provisioning
systems

- If a single Gateway/ASBR/PE connecting to the WAN on one side, and
connecting to the Data Center network on the other side, then this
Gateway/ASBR/PE is the demarcation point between the two networks.

- vPEs and VMs are provisioned by Data Center Orchestration systems.

- Managing IP VPNs in the WAN is not within the scope of this
document except the inter-connection points.

7.2 Management/Orchestration system interfaces

The Management/Orchestration system CAN be used to communicate with
both the DC Gateway/ASBR, and the end devices.

The Management/Orchestration system MUST support standard,
programmatic interface for full-duplex, streaming state transfer in
and out of the routing system at the Gateway.

The programmatic interface is currently under definition in IETF
Interface to Routing Systems (I2RS)) initiative.
[I-D.ietf-i2rs-architecture], and [I-D.ietf-i2rs-problem-statement].

Standard data modeling languages will be defined/identified in I2RS.
YANG - A Data Modeling Language for the Network Configuration
Protocol (NETCONF) [RFC6020] is a promising candidate currently under
investigation.

To support remote access between applications running on an end
device (e.g., a server) and routers in the network (e.g. the DC
Gateway), a standard mechanism is expected to be identified and
defined in I2RS to provide the transfer syntax, as defined by a
protocol, for communication between the application and the
network/routing systems. The protocol(s) SHOULD be lightweight and
familiar by the computing communities. Candidate examples include
ReSTful web services, JSON [RFC4627], NETCONF [RFC6241], XMPP
[RFC6120], and XML. [I-D.ietf-i2rs-architecture].

7.3 Service VM Management

Service VM Management SHOULD be hypervisor agnostic, e.g. On demand
service VMs turning-up SHOULD be supported.

7.4 Orchestration and IP VPN inter-provisioning

The orchestration system

1) MUST support IP VPN service activation in virtualized DC.

2) MUST support automated cross-provisioning accounting correlation
between the WAN IP VPN and Data Center for the same tenant.

3) MUST support automated cross provisioning state correlation
between WAN IP VPN and Data Center for the same tenant

There are two primary approaches for IP VPN provisioning - push and
pull, both CAN be used for provisioning/orchestration.

7.4.1 vPE Push model

Push model: push IP VPN provisioning from management/orchestration
systems to the IP VPN network elements.

This approach supports service activation and it is commonly used in
existing IP VPN Enterprise deployments. When extending existing WAN
IP VPN solutions into the a Data Center, it MUST support off-line
accounting correlation between the WAN IP VPN and the cloud/DC IP VPN
for the tenant. The systems SHOULD be able to bind interface
accounting to particular tenant. It MAY requires offline state
correlation as well, for example, binding of interface state to
tenant.

Provisioning the vPE solution:

1) Provisioning process

    a. The WAN provisioning system periodically provides to the DC
       orchestration system the VPN tenant and RT context.
    b. DC orchestration system configures vPE on a per request basis

2) Auto state correlation

3) Inter-connection options:

    Inter-AS options defined in [RFC4364] may or may not be sufficient
    for a given inter-connection scenario. BGP IP VPN inter-connection
    with the Data Center is discussed in
    [I-D.fang-l3vpn-data-center-interconnect].

    This model requires offline accounting correlation

    1) Cloud/DC orchestration configures vPE

    2) Orchestration initiates WAN IP VPN provisioning; passes
    connection IDs (e.g., of VLAN/VXLAN) and tenant context to WAN IP

VPN provisioning systems.

3) WAN IP VPN provisioning system provisions PE VRF and policies as in typical Enterprise IP VPN provisioning processes.

4) Cloud/DC Orchestration system or WAN IP VPN provisioning system MUST have the knowledge of the connection topology between the DC and WAN, including the particular interfaces on core router and connecting interfaces on the DC PE and/or vPE.

In short, this approach requires off-line accounting correlation and state correlation, and requires per WAN Service Provider integration.

Dynamic BGP sessions between PE/vPE and vCE MAY be used to automate the PE provisioning in the PE-vCE model, that will remove the needs for PE configuration. Caution: This is only under the assumption that the DC provisioning system is trusted and can support dynamic establishment of PE-vCE BGP neighbor relationships, for example, the WAN network and the cloud/DC belong to the same Service Provider.

7.4.2 vPE Pull model

Pull model: pull from network elements to network management/AAA based upon data plane or control plane activity. It supports service activation. This approach is often used in broadband deployments. Dynamic accounting correlation and dynamic state correlation are supported. For example, session based accounting is implicitly includes tenant context state correlation, as well as session-based state that implicitly includes tenant context. Note that the pull model is less common for vPE deployment solutions.

Provisioning process:

1) Cloud/DC orchestration configures vPE

2) Orchestration primes WAN IP VPN provisioning/AAA for new service, passes connection IDs (e.g., VLAN/VXLAN) and tenant context.

3) Cloud/DC ASBR detects new VLAN and sends Radius Access-Request (or Diameter Base Protocol request message [RFC6733]).

4) Radius Access-Accept (or Diameter Answer) with VRF and other policies

     Auto accounting correlation and auto state correlation is
     supported.

8. **Security Considerations**

     As vPE is an extended BGP/MPLS IP VPN solution, security threats
     and defense techniques described in RFC 4111 [RFC4111] for IP VPN
     generally apply.

     When the SDN approach is used, the protocols between the vPE agent
     and the vPE-C in the controller MUST be mutually authenticated.
     Given the potentially very large scale and the dynamic nature in
     the cloud/DC environment, the choice of key management mechanisms
     need to be further studied.

9. **IANA Considerations**

     None.

10. **References**

10.1 **Normative References**

   [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
              Requirement Levels", BCP 14, RFC 2119, March 1997.

   [RFC3032] Rosen, E., Tappan, D., Fedorkow, G., Rekhter, Y.,
              Farinacci, D., Li, T., and A. Conta, "MPLS Label Stack
              Encoding", RFC 3032, January 2001.

   [RFC4023] Worster, T., Rekhter, Y., and E. Rosen, Ed.,
              "Encapsulating MPLS in IP or Generic Routing Encapsulation
              (GRE)", RFC 4023, March 2005.

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

   [RFC4684] Marques, P., Bonica, R., Fang, L., Martini, L., Raszuk,
              R., Patel, K., and J. Guichard, "Constrained Route
              Distribution for Border Gateway Protocol/MultiProtocol
              Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual
              Private Networks (VPNs)", RFC 4684, November 2006.

   [RFC5036] Andersson, L., Ed., Minei, I., Ed., and B. Thomas, Ed.,
              "LDP Specification", RFC 5036, October 2007.

   [RFC6020] Bjorklund, M., Ed., "YANG - A Data Modeling Language for

the Network Configuration Protocol (NETCONF)", RFC 6020,
October 2010.

[RFC6120]   Saint-Andre, P., "Extensible Messaging and Presence
            Protocol (XMPP): Core", RFC 6120, March 2011.

[RFC6241]   Enns, R., Ed., Bjorklund, M., Ed., Schoenwaelder, J., Ed.,
            and A. Bierman, Ed., "Network Configuration Protocol
            (NETCONF)", RFC 6241, June 2011.

[RFC6733]   Fajardo, V., Ed., Arkko, J., Loughney, J., and G. Zorn,
            Ed., "Diameter Base Protocol", RFC 6733, October 2012.

10.2 Informative References

[RFC4111]   Fang, L., Ed., "Security Framework for Provider-
            Provisioned Virtual Private Networks (PPVPNs)", RFC 4111,
            July 2005.

[RFC4627]   Crockford, D., "The application/json Media Type for
            JavaScript Object Notation (JSON)", RFC 4627, July 2006.

[RFC4797]   Rekhter, Y., Bonica, R., and E. Rosen, "Use of Provider
            Edge to Provider Edge (PE-PE) Generic Routing
            Encapsulation (GRE) or IP in BGP/MPLS IP Virtual Private
            Networks", RFC 4797, January 2007.

[I-D.ietf-l3vpn-end-system] Marques, P., Fang, L., Pan, P., Shukla,
            A., Napierala, M., Bitar, N., "BGP-signaled end-system
            IP/VPNs", draft-ietf-l3vpn-end-system, work in progress.

[I-D.fang-l3vpn-end-system-req] Napierala, M., and Fang, L.,
            "Requirements for Extending BGP/MPLS VPNs to End-Systems",
            draft-fang-l3vpn-end-system-requirements, work in
            progress.

[I-D.rfernando-l3vpn-service-chaining] Fernando, R., Rao, D., Fang,
            L., Napierala, M., So, N., draft-rfernando-l3vpn-service-
            chaining, work in progress.

[I-D.fang-l3vpn-virtual-ce] Fang, L., Evans, J., Ward, D., Fernando,
            R., Mullooly, J., So, N., Bitar., N., Napierala, M., "BGP
            IP VPN Virtual PE", draft-fang-l3vpn-virtual-ce, work in
            progress.

[I-D.ietf-i2rs-architecture] Atlas, A., Halpern, J., Hares, S., Ward,

             D., and Nadeau, T., "An Architecture for the Interface to
             the Routing System", draft-ietf-i2rs-architecture, work in
             progress.

   [I-D.ietf-i2rs-problem-statement] Atlas, A., Nadeau, T., and Ward,
             D., "Interface to the Routing System Problem Statement",
             draft-ietf-i2rs-problem-statement, work in progress.

   [I-D.bitar-i2rs-service-chaining] Bitar, N., Geron, G., Fang, L.,
             Krishnan, R., Leymann, N., Shah, H., Chakrabarti, S.,
             Haddad, W., draft-bitar-i2rs-service-chaining, work in
             progress.

   [I-D.fang-l3vpn-data-center-interconnect] Fang, L., Fernando, R.,
             Rao, D., Boutros, S., "BGP IP VPN Data Center
             Interconnect", draft-fang-l3vpn-data-center-interconnect,
             work in progress.

   [I-D.mahalingam-dutt-dcops-vxlan] Mahalingam, M, Dutt, D.., et al.,
             "A Framework for Overlaying Virtualized Layer 2 Networks
             over Layer 3 Networks" draft-mahalingam-dutt-dcops-vxlan,
             work in progress.

   [I-D.sridharan-virtualization-nvgre] SridharanNetwork, M., et al.,
             "Virtualization using Generic Routing Encapsulation",
             draft-sridharan-virtualization-nvgre, work in progress.

Authors' Addresses

   Luyuan Fang
   Cisco
   111 Wood Ave. South
   Iselin, NJ 08830
   Email: luyuanf@gmail.com

   David Ward
   Cisco
   170 W Tasman Dr
   San Jose, CA 95134
   Email: wardd@cisco.com

   Rex Fernando
   Cisco
   170 W Tasman Dr
   San Jose, CA
   Email: rex@cisco.com

Maria Napierala
AT&T
200 Laurel Avenue
Middletown, NJ 07748
Email: mnapierala@att.com

Nabil Bitar
Verizon
40 Sylvan Road
Waltham, MA 02145
Email: nabil.bitar@verizon.com

Dhananjaya Rao
Cisco
170 W Tasman Dr
San Jose, CA
Email: dhrao@cisco.com

Bruno Rijsman
Juniper Networks
10 Technology Park Drive
Westford, MA 01886
Email: brijsman@juniper.net

Ning So
Tata Communications
Plano, TX 75082, USA
Email: ning.so@tatacommunications.com

Jim Guichard
Cisco
Boxborough, MA 01719
Email: jguichar@cisco.com

Wen Wang
CenturyLink
2355 Dulles Corner Blvd.
Herndon, VA 20171
Email:Wen.Wang@CenturyLink.com

Manuel Paul
Deutsche Telekom
Winterfeldtstr. 21-27
10781 Berlin, Germany
Email: manuel.paul@telekom.de