INTERNET-DRAFT
Intended Status: Standards track
Expires: August 25, 2013

Luyuan Fang
David Ward
Rex Fernando
Cisco
Maria Napierala
AT&T
Nabil Bitar
Verizon
Dhananjaya Rao
Cisco
Bruno Rijsman
Juniper
Ning So
TATA Communications

February 25, 2013

BGP IP VPN Virtual PE
draft-fang-l3vpn-virtual-pe-01

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 plane, the data plane, and the provisioning management process. The vPE solutions supports both Software Defined Networking (SDN) approach by allowing physical decoupling of the control and the forwarding plane of a vPE, as well as a distributed routing approach. The solution allows vPE to be co-resident with the application virtual machines (VMs) on a single end device, such as a server, as well as on a Top-of-Rack switch (ToR), or in any network or compute device. The ability to provide end-to-end native BGP IP VPN connections between a Data Center (DC) (or other types of service network) applications and the enterprise IP VPN sites is highly desirable to both Service Providers and Enterprises.

Copyright and License Notice

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 IP VPNs solutions. It 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 infrastructure.

With the advent of compute capabilities and the proliferation of virtualization in Data Center servers, multi-tenant data centers have become a reality. As applications and appliances are increasingly being virtualized, supporting virtual edge devices, such as virtual IP VPN PE routers, becomes feasible and a natural part of the overall virtualization solutions. And there is strong desire from Service Providers to extend their existing BGP IP VPN deployment into Data Centers to provide Virtual Private Cloud (VPC) services.

The virtual Provider Edge (vPE) solution described in this document allows extending the PE functionality of BGP/MPLS IP VPN to the end devices, such as servers where the applications reside, or to the first hop routing/switching device, such as a Top of the Rack switch (ToR) in a Data Center.

The vPE solutions support both Software Defined Network (SDN) approach by allowing physical decoupling of the control and the forwarding plane of a vPE, and distributed routing approach in the same fashion as IP VPN is done with the physical PEs.

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
3GPP	3rd Generation Partnership Project (3GPP)
AS	Autonomous System
ASBR	Autonomous System Border Router
BGP	Border Gateway Protocol
CE	Customer Edge
ED	End device: where Guest OS, Host OS/Hypervisor,

	applications, VMs, and virtual router may reside
Forwarder	L3VPN forwarding function
GRE	Generic Routing Encapsulation
Hypervisor	Virtual Machine Manager
I2RS	Interface to Routing Systems
IaaS	Infrastructure as a Service
LDP	Label Distribution Protocol
LTE	Long Term Evolution
MP-BGP	Multi-Protocol Border Gateway Protocol
PCEF	Policy Charging and Enforcement Function
P	Provider backbone router
QoS	Quality of Service
RR	Route Reflector
RT	Route Target
RTC	RT Constraint
SDN	Software Defined Network
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 Reflector1.2 Scope of the document
WAN	Wide Area Network

1.2 Motivation and requirements

The recent rapid adoption of Cloud Services by enterprises and the phenomenal growth of mobile IP applications accelerate the needs to extend the BGP IP VPN capability into cloud service end devices. For examples, enterprise customers' want to extend the existing IP VPN services in the WAN into the new cloud services supported by various Data Center (DC) technologies; Large enterprise have existing L3VPN deployment are extending them into their Data Centers; Mobile providers adopting IP VPN into their 3GPP Mobile infrastructure are looking to extend the IP VPNs to their end devices of the call processing center. In general, Service Providers intend to use the vPE solutions for cloud service development regardless with or without the inter-connection to existing enterprise BGP IP VPNs.

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 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 Data Center 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 plane and forwarding for network virtualization and abstraction.

5) MUST support of control plane through SDN controller, as well as through traditional distributed MP-BGP approach.

6) MUST support VM mobility

7) SHOULD support orchestration/provisioning

8) SHOULD support service chaining

The architecture and protocols defined in BGP/MPLS IP VPN [RFC4364] provide the foundation for virtual PE 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 the backbone network. It is collection of "sites". A site can be considered as a set of IP systems maintain IP inter-connectivity without connecting through the backbone. The typical use of L3VPM has been to inter-connect different sites of an Enterprise networks through Service Provider's BGP IP VPNs in the WAN.

A virtual PE (vPE) is a BGP 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 vPE 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 in a Data Center. When a vPE-F is residing in a server, its connection to a co-resident VM is as the PE-CE relationship in the regular BGP IP VPNs, but

without routing protocols running between the virtual PE and CE because the connection is internal to the device.

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, vPE-F may be in a server as 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.

Alternatively, vPE control plane can reside in the same physical device where the vPE-F resides. In this case, it is similar as the traditional implemention VPN PE, 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 of application VMs. For example, 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 Data Center, not as co-resident with the application VMs.

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

2.2.2 vPE control plane topology

Option 2a. vPE control plane is physically decoupled from vPE forwarder, 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 vPE forwarding plane.

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

2.2.3 Data Center orchestration models

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

Option 3b. Pull model: It is a bottom-up approach, pull 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 vPE solution with vPE-F in the end device co-resident with applications VMs, while vPE-C is physically decoupled and residing on the controller.

The Data Center (e.g. a DC) is connected to the IP/MPLS core via the Gatways/ASBRs. The IP VPN , e.g. VPN RED, in the Data Center has one terminating point at the vPE-F on the end device in the Data Center, inter-connecting the the IP VPN in the WAN which belong to the same client, 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 Data Center fabric/intermediate underlay devices in the Data Center do not participate IP VPNs, their function is the same as P routers in MPLS back bone, they do not maintain the IP VPN states, not 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 are both reside on the end-device,
vPE functions the same as it is in a physical PE. MP-BGP is used for
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  |   |  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 function same as physical PE, MP-BGP is used for

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 Server view of vPE

An end device shown in Figure 4 is a virtualized server which hosts multiple VMs, the virtual PE is co-resident in the server. 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 with each other 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. Because now the vPE and CE are co-resident in the server, the connection between them is internal implementation to 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

3. Control Plane

3.1 vPE Control Plane (vPE-C)

The vPE control plane MAY use SDN controller approach or use distributed MP-BGP.

3.1.1 SDN approach

This approach is used when vPE control plane and data plane are physically decoupled. The control plane directing the data flow may reside elsewhere, such a SDN controller. This requires standard interface to routing system (I2RS). The Interface to Routing System (IRS) is work in progress in IETF [I-D.ward-irs-framework], [I-D.rfernando-irs-fw-req].

Though MP-BGP is often the de facto preferred choice between vPE and gateway-PE, using extensible signaling messaging protocols MAY often be more practical in Data Center environment, such technologies have been proposed for this segment of signaling [I-D.ietf-l3vpn-end-system], and more protocols are available (to add details later).

3.1.2 Distributed control plane

vPE participates in overlay L3VPN control protocol: MP-BGP [RFC4364].

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

When 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 data centers can pass the scale of those in SP backbone VPN networks. There are may be tens of thousands of end devices in a single Data Center.

Use of Router Reflector (RR) is necessary in large scale L3VPN networks to avoid full iBGP mesh among all vPEs and PEs. The L3 VPN routes can be partitioned to a set of RRs, the partition techniques are detailed in [RFC4364].

When RR is residing in a physical device, e.g., a server, which is partitioned to support multi-functions and applications VMs, the RR becomes virtualized RR (vRR). Since RR's performs control plane only, a physical 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 be in Gateway PE, or in an end device.

3.4 Use of RT constraint

The Route Target Constraint (RT Constraint, RTC) [RFC4684] is a powerful tool for VPN selective L3VPN route distribution. With RT Constraint, only the BGP receiver (e.g, PE/vPE/RR/vRR/ASBRs, etc.) with the particular L3VPNs will receive the route update for the corresponding VPNs. It is critical to use RT constraint to support large scale L3VPN development.

4. Forwarding Plane

4.1 Virtual Interface

Virtual Interface (VI) is an interface in an end device which is used for connecting the vPE to the application VMs in the end device. The latter cab be treated as CEs in the regular L3VPN's view.

4.2 VPN Forwarder (vPE-F)

VPN Forwarder is the forwarding component of a vPE where the MPLS VPN labels are pushed/popped..

The VPN forwarder location options:

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

2) in an external device which the end device connect to, for example, a Top of the Rack (ToR) in a data center.

Multiple factors should be considered for the location of the VPN forwarder, including device capability, overall solution economics, QoS/firewall/NAT placement, optimal forwarding, latency and performance, operation 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 for BGP/MPLS L3VPN.

1. MPLS Encapsulation with Label Distribution Protocol [LDP], [RFC3032].

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

3. Other types of encapsulation. For example, VXLAN [I-D.mahalingam-dutt-dcops-vxlan], NVGRE [I-D.sridharan-virtualization-nvgre], and other modified version of these or other existing protocols.

The most common BGP/MPLS L3VPNs deployment in SP networks are using MPLS forwarding. This requires MPLS, 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 network against attacks.

However, the Data Center environment, such as a data center, is different than Service Provider VPN networks or large enterprise backbones. MPLS deployment MAY or MAY not be feasible or desirable. Two major challenges for MPLS deployment 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 more practical in most data center, and computing environment. Note that when IP encapsulations are used, the associated security considerations must be analyzed carefully.

In addition, there are new encapsulation proposals for Data Center/Data center currently as work in progress in IETF, including several UDP based encapsulations proposals and some TCP based proposal. These overlay encapsulations can be suitable alternatives for a vPE, considering the availability and leverage of support in virtual and physical devices.

4.4 Optimal forwarding

As reported by many large cloud service operators, the traffic pattern in their data centers were dominated by East-West across subnet traffic (between the end device hosting different applications in different subnets) than North-South traffic (going in and out the DC to the WAN) or switched traffic within subnets. This is a primary reason that many large scale new design has moved away from traditional L2 design to L3, especially for overlay networks.

When forwarding the traffic within the same VPN, the vPE should be able to provide direct communication among the VMs/application senders/receivers without the need of going through gateway devices. If it is on the same end device, the traffic should not need to leave the same device. If it is on different end device, optimal routing should be applied.

When multiple VPNs need to be accessed to accomplish the task the

user requested (this is common too), the end device virtual interfaces CAN directly access multiple VPNs via use of extranet VPN techniques without the need of Gateway facilitation. This is done through the use of BGP L3VPN policy control mechanisms to support this function. In addition, ECMP is a build in layer 3 mechanism, it is used for load sharing.

Optimal use of available bandwidth can be achieved by virtue of using ECMP in the underlay, as long as the encapsulation include certain entropy in the header (e.g. VXLAN).

5. Addressing

5.1 IPv4 and IPv6 support

Both IPv4 and IPv6 MUST be supported in the virtual PE solution.

This may present challenging to older devices, but may not be issues to newer forwarding devices and servers. A server is replaced much more frequently than a network router/switch in the infrastructure network, newer equipment should be capable of IPv6 support.

5.2 Address space separation

The addresses used for IP VPN overlay in the Data Center, such as a Data Center, SHOULD be in separate address blocks than the ones used the underlay infrastructure of the Data Center. This practice is to protect the Data Center infrastructure being attacked if the attacker gain access of the tenant VPNs.

Similarity, the addresses used for the Data Center, e.g., a Data Center, SHOULD be separated from the WAN backbone addresses space.

6.0 Inter-connection considerations

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

There are deployment scenarios that IP VPN may not be supported in every segment of the networks to provide end-to-end IP VPN connectivity, an IP VPN 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 overall solution, a clear demarcation should be preserved at the inter-connecting points. The problems encountered in one domain should not impact the other domains.

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

As described earlier in this document, the IP VPN vPE should ideally be located as close to the "customer" edge devices. For cases, where 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 best suited BGP solution is to use IP VPNs [RFC4364] for all sites with PE and/or vPE solutions. This is a straightforward case.

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

One can consider to use combined vPE and vCE solution to solved the problem. Use IP VPN for all sites with IP connectivity, and use a physical or virtual CE (vCE, may reside on the end device) to aggregate the L2 sites which, for example, are in a single container in a data center. The CE/vCE can be considered as inter-connecting point, where the L2 network are terminated and the corresponding routes for connectivity of the L2 network are inserted into L3VPN VRF. The L2 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 assumptions:

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

- WAN ASBR/PEs are provisioned by existing WAN provisioning systems

- If a single ASBR/PE connecting WAN on one side, and connecting DC network on the other side, this ASBR/PE is the demarcation point between the two networks

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

- Managing IP VPNs in the WAN is not in scope except the inter-connection point.

7.2 Management/Orchestration system interfaces

The Management/Orstration system CAN be used to communicate with both the Data Center Gateway, 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 are current under definition in IETF Interface to Routing Systems (I2RS)) initiative. [I-D.ward-irs-framework], [I-D.rfernando-irs-fw-req].

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), 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 light-weight and familiar by the computing communities. Candidate examples include ReSTful web services, JSON [RFC4627], XMPP [RFC6120], and XML. [I-D.ward-irs-framework].

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 Data Center.

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

3) MAY 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: It is a top down approach - 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 the existing IP VPN enterprise deployment. When extending existing WAN IP VPN solution 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, bind interface state to tenant.

Provisioning for vPE solution:

1) Provisioning process

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

2) Auto state correlation

4) Inter-connection options:

Inter-AS options defined in [RFC4364] may or may not be sufficient for a given inter-connecting scenario. BGP IP VPN inter-connection with 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 NGN, including the particular interfaces on core router and connecting interfaces on the DC PE.

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

Dynamic BGP session 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 could 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: It is a bottom-up approach - 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 deployment. 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 which implicitly includes tenant context.

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 WAN IP VPN provisioning systems.

3) Cloud/DC ASBR detects new VLAN, send Radius Access-Request

4) Radius Access-Accept with VRF and other policies

Auto accounting correlation and auto state correlation is supported.

7. Security Considerations

vPE solution presented a virtualized IP VPN PE model. There are potential implications to IP VPN control plane, forwarding plane, and management plane. Security considerations are currently under study, will be included in the future revisions.

8. IANA Considerations

None.

9. References

9.1 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.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[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.

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

9.2 Informative References

[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., "BGP-signaled end-system IP/VPNs", draft-ietf-l3vpn-end-system-00, October 2012.

[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-01, Oct. 2012.

[I-D.ward-irs-framework] Atlas, A., Nadeau, T., Ward. D., "Interface to the Routing System Framework", draft-ward-irs-framework-00, July 2012.

[I-D.rfernando-irs-fw-req] Fernando, R., Medved, J., Ward, D., Atlas, A., Rijsman, B., "IRS Framework Requirements", draft-rfernando-irs-framework-requirement-00, Oct. 2012.

[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-01, Feb. 2013.

[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-00, Feb. 2013.

[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-02, Aug. 2012.

[I-D.sridharan-virtualization-nvgre]: SridharanNetwork, M., et al., "Virtualization using Generic Routing Encapsulation", draft-sridharan-virtualization-nvgre-01.txt, July 2012.

Authors' Addresses

Luyuan Fang
Cisco
111 Wood Ave. South
Iselin, NJ 08830

Email: lufang@cisco.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