INTERNET-DRAFT
Intended Status: Informational
Expires: April 15, 2013

Luyuan Fang
Rex Fernando
Cisco
Maria Napierala
AT&T

October 15, 2012

BGP L3VPN Virtual PE Framework
draft-fang-l3vpn-virtual-pe-framework-00

Abstract

This document describes a framework for BGP/MPLS L3VPN with virtual PE solutions. It provides the information on control plane, data plane of the virtual PE solutions, and its interaction with other network elements. The solution supports further control and forwarding separation from traditional L3VPN solutions. It allows the L3VPN functions extended to application end systems for large scale and efficient IP application support.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

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

Copyright and License Notice

Table of Contents

1 Introduction

Network virtualization is to provide multiple individual network services by sharing common available network resources. Network virtualization is not a new concept, for example, BGP layer 3 Virtual Private Networks (L3 VPNs) have been widely deployed to provide network based, service provider provisioned VPNs. It provides routing isolation and allow address overlapping among different VPNs while forward traffic over common network infrastructure.

The recent development of server virtulization, provided the new opportunities for the virtual Provider Edge (vPE) solution on application end-system. The virtual PE solution can be powerful and attractive to service providers and enterprises in the fast growing cloud computing service and intelligent mobility environment.

1.1 Overview

L3 VPN Virtual Provider Edge may provide full or partial L3VPN PE functions or partial PE functions. The virtual PE has two components - control and forwarding. The control component can be a software instance resides in a physical device, most commonly seen in the end-system devices where multiple applications are supported, e.g., mobile application server in a wireless call center, or an end-system in a SP virtual Private Cloud (vPC) data center, or a host in an Financial back-office.

The architecture and protocols defined in IETF for BGP/MPLS L3VPN [RFC4364] is the foundation for virtual PE solution. Certain protocol extensions or integration may be needed to support the virtual PE solutions.

This document defines a framework for using standard protocols to build BGP L3 VPN with virtual PE solutions. The goal is to support large scale deployment and reduce operational complexity. The targeted environment can be virtualized wireless providers call

centers, large scale service providers data centers, large enterprise central and branch facilities, and service provider managed services.

1.2 Scope of the document

This focus of this document is BGP L3VPN virtual PE solutions.

It is assumed that the readers are familiar with BGP/MPLS L3VPN technologies, the base technology and operation will not be covered in this document.

The following network elements are in scope: BGP L3VPN vPE; the interaction of vPE with other network elements, including BGP L3VPN physical PE, physical BGP Route Reflector (RR) and virtual BGP Route Reflector (vRR), and Autonomic System Border Router (ASBR), external controller, provisioning/orchestration systems. Definitions of protocol extensions is out of scope of this document.

1.3 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].

AS	Autonomous Systems
ASBR	Autonomous Systems Border Router
BGP	Border Gateway Protocol
End-Device	A device where Guest OS and Host OS/Hypervisor may reside, aka End-system
GRE	Generic Routing Encapsulation
IaaS	Infrastructure as a Service
IRS	Interface to Routing System
LTE	Long Term Evolution
PCEF	Policy Charging and Enforcement Function
RR	Route Reflector
RT	Route Target
RTC	RT Constraint
ToR	Top-of-Rack switch

VM	Virtual Machine
Hypervisor	Virtual Machine Manager
VM	Virtual Machine
SDN	Software Defined Network
VI	Virtual Interface
vPC	virtual Private Cloud
vPE	virtual Provider Edge
VPN	Virtual Private Network
vRR	virtual Route Reflector

2. Virtual PE Architecture and Reference Model

2.1 Virtual PE

A virtual PE is a PE with control instance and a forwarding components reside in a shared physical device where multiple applications are supported. The control and forwarding components are decoupled, they may reside in the same or different physical devices.

A key motivation of using virtual PE solution is to place the L3VPN termination point as close as possible to where the services/applications reside; and to take the advantage of control and forwarding decoupling for better scalability and allow flexible routing policy control and fast provisioning. In many cases, the virtual PE is placed in the service end systems where the Virtual Machines running various applications.

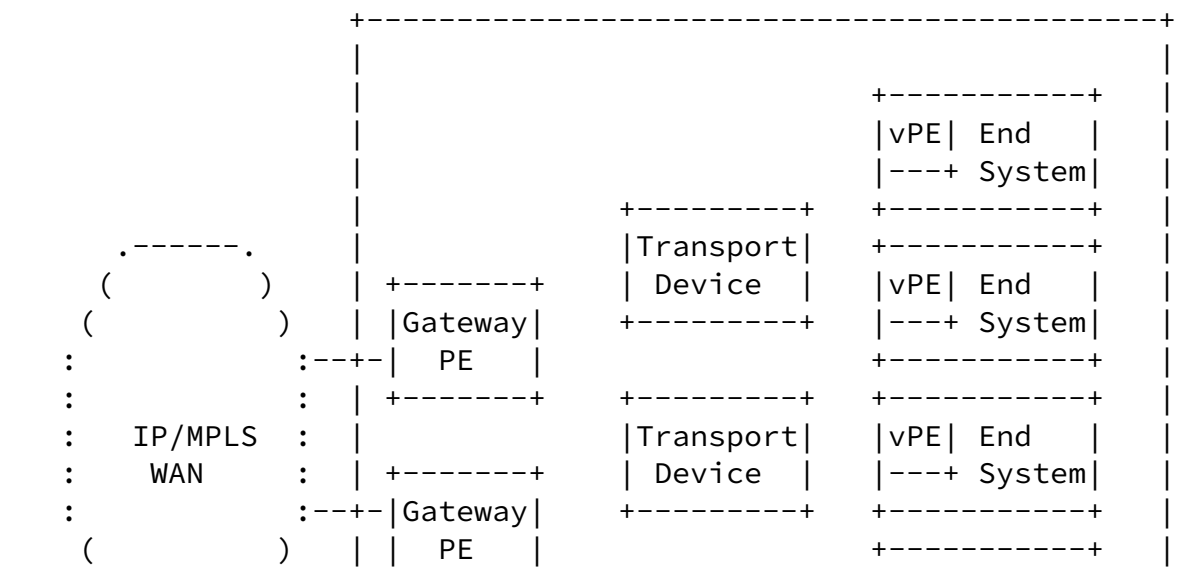
As defined in [RFC4364], a L3VPN is created by applying policies to form a subset of sites among all sites connected the backbone network. It is collection of "sites". A site can be considered as a set of IP systems maintain IP inter-connectivity without connecting through the backbone. The typical use of L3VPM has been to inter-connect different sites of an Enterprise networks through Service Provider's L3VPN in the WAN.

The recent rapid adoption of Cloud Services, and the phenomenal growth of mobile IP applications, accelerate the needs to extend the VPN capability to the end-systems. Enterprise customers requests Service Providers to extend the L3VPN services in the WAN into the new Cloud services supported by various Data Center technologies. Mobile providers who have already adopted L3VPN into the Mobile

infrastructure are looking to support service virtualization with L3VPN on the end-system of mobile applications.

2.2 Architecture reference model

Figure 1 illustrate the topology that vPE is reside inside the end-system where the applications are hosted.



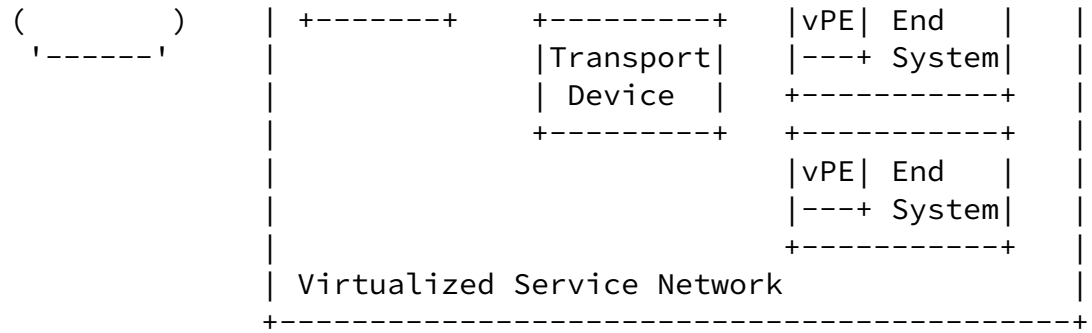
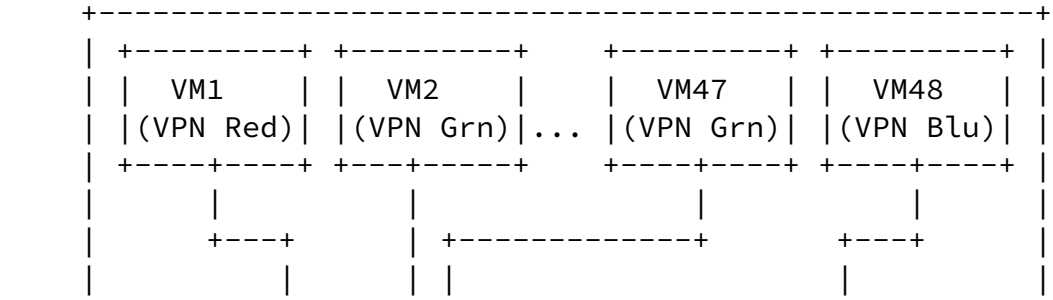


Figure 1. Virtualized Service network with vPE

The Virtualized Service Network in Figure 1 consists of WAN gateway PE devices, transport devices, and end-systems.

Examples of service network may be a network that supports cloud computing services, mobile call centers, and SP or enterprise private or public data centers.

The transport devices in the service network do not participate L3VPNs, they do not maintain the L3VPN states, they are not aware of the L3VPN in the network.



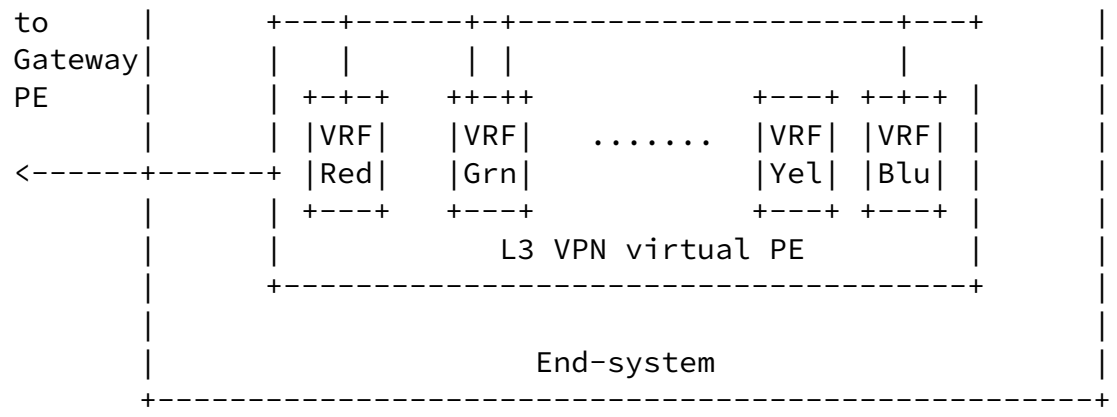


Figure 2. Virtual L3 VPN PE logical connections in the End-system

An end-system shown in Figure 2 is a virtualized server which hosting multiple VMs, the a virtual PE resides in the end-system. The vPE supports multiple VRFs, VRF Red, VRF Grn, VRF Yel, VRF Blu, etc. Each 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 RFC Grn, etc. Routing isolations apply between VPNs.

The vPE connectivity relationship between vPE to VM is similar like the PE to CE in a regular BGP L3VPNs.

VM1 and VM2 can not connect to each other in a simple intranet VPN topology as shown in the configuration. The L3VPN provides routing isolation among different VPNs.

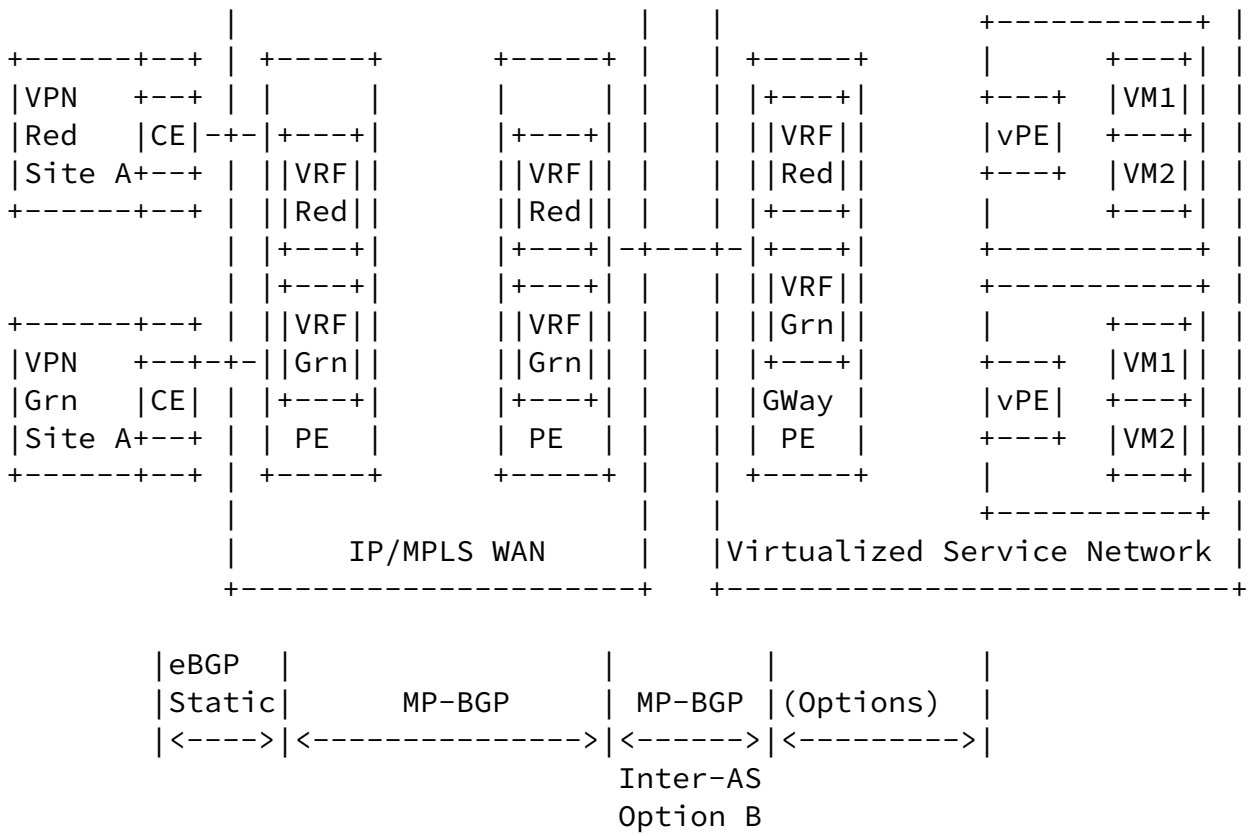


Figure 3. L3VPN Logical Connection protocols

Options for protocol between Gateway PE to vPE on the end-system:

1. MP-BGP.

2. XMPP or other extensible messaging protocols which are familiar by computing work force.

3. Network Controller

Options for protocols between IP/MPLS WAN network and Virtualized service network depending on the design. Inter-AS with Option B is an example.

3. Control Plane

L3VPN control protocol is MP-BGP as defined in [RFC4364].

When extending L3VPN into service network, supporting MP-BGP on the end-system may or may not be practical, alternatives are also

considered here.

3.1 Router server of vPE

A virtual PE consist the control plane element and the forwarding plane element. Since the proposed solution decoupled the two element, they may or may not reside on the same physical device.

The Route Server of L3VPN vPE is a software application that implements the BGP/MPLS L3VPN PE control plane functionality.

In the case other control/signaling/messaging protocol are used (see options listed in the next sub-section), the route server is also the server of the particular protocol(s), it interacts with VPN forwarder (see the section on forwarding).

3.2 Control plane options of vPE

a. MP-BGP

MP-BGP can be used in service network where both the end system implementation and operation work force has the knowledge and skills to support it. In this case, the design and deployment is very similar to what applies to regular L3VPN deployment.

b. Extensible messaging protocols

It is redeemed as a good alternative for bridging the gap between Gateway PE and the vPE where operators could not support BGP.

Although the modern end-systems may be able to support MP-BGP, the operational support of the computing and storage community often may not have the adequate BGP skills or willingness to step up to BGP operation. Since this is a common situation today, an light weight, extensible IP protocol is very welcome by the cloud service communities. One example of such protocol is to use XMPP to signal L3VPN connectivity between the Gateway PE the virtual PE on the end-systems. The technology solution is described in details in [I-D.ietf-l3vpn-end-system].

c. Controller

This is a SDN approach. In the virtual PE implementation, not only the service network infrastructure and the VPN overlay networks are decoupled, but also the vPE control plane and data plane are physically decoupled. The control plane directing the data flow may

reside elsewhere, such a centralized controller. This requires standard interface to routing system (IRS). The IRS work is in

progress in IETF, [ID.ward-irs-framework], [ID.rfernando-irs-framework-requirement].

3.3 Use of router reflector

Modern service networks can be very large in scale, the very large data centers can easily pass the scale of SP backbone VPN networks. The end-systems and therefore the vPEs may be several thousand in a single service network.

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 a device which is partitioned to support multi-functions and application, the RR become virtualized RR (vRR). Since RR's is control plane only, a physical or virtualized server with large scale of computing power and memory can be a perfect candidate as host of vRRs. The vRR can be in Gateway PE, or an end-system.

3.4 Use of RT constraint

The Route Target Constraint [RFC4684] is a powerful tool for VPN route filtering. With RT constraint, only the BGP receiver (a PE, vPE, RR, vRR, ASBRs, etc.) with the particular L3VPN routes will receive the route update. It is critical to use RT constraint particularly in large scale development.

4. Forwarding Plane

4.1 Virtual Interface

Virtual Interface (VI) is an interface in an end-system which is used for connecting the vPE to the VMs or other applications in the end-system. The later can be viewed as CEs in traditional L3VPN scenario.

4.2 VPN forwarder

VPN Forwarder is the forwarding component of a vPE solution.

The VPN forwarder location options:

1) within the end-system where the virtual interface is

2) in an external device of end-system which the end-system connect to, for example, a ToR.

Multiple factors should be considered for the location of the VPN forwarder, including device capability, overall economy, QoS/firewall/NAT placement, optimal forwarding, latency and performance, etc. There are design trade offs, it is worth the effort to study the traffic pattern and forwarding looking trend in your own unique service network as part of the exercise.

4.3 Encapsulation

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

1. MPLS Encapsulation, [RFC3032].

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

The most common BGP/MPLS L3VPNs deployment in SP networks are using MPLS forwarding. This requires MPLS to be deployed in the network. It is proven to scale and provide various security mechanism to against attacks.

The service network 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. Two major challenges for MPLS deployment in the new environment: 1) the capabilities for the end-system devices or transport/forwarding devices; 2) the workforce skill set.

Encapsulating MPLS in IP or GRE tunnel may be more practical in many cases. But bare in mind, IP or GRE tunnel does not provide the same level of security mechanism as MPLS forwarding.

There are new encapsulation proposals for service network/Data center as work in progress in IETF, including several UDP based encapsulations proposals and some TCP based proposals. These mechanism may be considered as alternative to MPLS and IP/GRE encap.

4.4 Optimal forwarding

As reported by many large cloud service operators, the traffic pattern in their data centers are dominated by East-West traffic (between the end-systems hosting different applications) than North-South traffic (going in and out the DC to the WAN). This is a key reason that many large scale new design has moved away from traditional L2 design to L3.

When forwarding the traffic within the same VPN, the end-system

should be able to access directly between the VMs/application sender/receivers without the need of going through gateway devices. If it is on the same end-system, the traffic should not need to leave the same device. If it is on different end-system, 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-system virtual interfaces should be able to directly access multiple VPNs in extranet VPN style without the need of Gateway facilitation. BGP L3VPN policy control is the tool to support this function.

5. Addressing

5.1 IPv4 and IPv6 support

Both IPv4 and IPv6 should 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 most likely support IPv6.

5.2 Address space separation

The addresses used for L3VPNs in the service network should be in separate address blocks than the ones used the underlay infrastructure of the service network. This practice is to protect the service network infrastructure being attacked if the attacker gain access of the tenant VPNs.

Similarity, the addresses used for the service network, e.g., a cloud service center of a SP, should be separated from the WAN backbone addresses space, for security reasons.

7. Security Considerations

To be added.

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.

[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.

9.2 Informative References

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

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

Authors' Addresses

Luyuan Fang
Cisco
111 Wood Ave. South

Iselin, NJ 08830
Email: lufang@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