Abstract

This document discusses two categories of inter-connections of BGP/MPLS IP VPN and Data Center (DC) overlay networks. In the first category, DC overlay virtual network is built with BGP/MPLS IP VPN (IP VPN) technologies, and the inter-connection of IP VPN in the DC either to IP VPN in other DCs or to IP VPN in the WAN enables end-to-end IP VPN connectivity. In the second category, DC overlay network uses non IP VPN overlay technologies, and the inter-connection of any overlay virtual network in the DC to IP VPN in the WAN provides end user connectivity through stitching of different overlay technologies.

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
1 Introduction

With the growth of cloud services, the need of inter-connecting DC overlay networks and Enterprise BGP/MPLS IP VPNs in the Wide Area Network (WAN) arises.
Two categories of inter-connections of BGP/MPLS IP VPN [RFC4364] and Data Center (DC) overlay networks are discussed in this document. In the first category, DC overlay virtual network is built with BGP/MPLS IP VPN (IP VPN) technologies, and the inter-connection of IP VPN in the DC either to IP VPN in other DCs or to IP VPN in the WAN enables end-to-end IP VPN connectivity for Virtual Private Cloud (VPC) services. In the second category, DC overlay network uses non IP VPN overlay technologies, the inter-connection of any overlay virtual network in the DC to IP VPN in the WAN provides end user connectivity through stitching of different overlay technologies.

This document discusses use cases of the inter-connection of BGP/MPLS VPN to Data Centers, the general requirements, and the proposed solutions for the inter-connections.

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].

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS</td>
<td>Autonomous System</td>
</tr>
<tr>
<td>ASBR</td>
<td>Autonomous System Border Router</td>
</tr>
<tr>
<td>BGP</td>
<td>Border Gateway Protocol</td>
</tr>
<tr>
<td>CE</td>
<td>Customer Edge</td>
</tr>
<tr>
<td>GRE</td>
<td>Generic Routing Encapsulation</td>
</tr>
<tr>
<td>Hypervisor</td>
<td>Virtual Machine Manager</td>
</tr>
<tr>
<td>I2RS</td>
<td>Interface to Routing System</td>
</tr>
<tr>
<td>MP-BGP</td>
<td>Multi-Protocol Border Gateway Protocol</td>
</tr>
<tr>
<td>NVGRE</td>
<td>Network Virtualization using GRE</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RD</td>
<td>Route Distinguisher</td>
</tr>
<tr>
<td>RR</td>
<td>Route Reflector</td>
</tr>
<tr>
<td>RT</td>
<td>Route Target</td>
</tr>
<tr>
<td>RTC</td>
<td>RT Constraint</td>
</tr>
<tr>
<td>SDN</td>
<td>Software Defined Network</td>
</tr>
<tr>
<td>ToR</td>
<td>Top-of-Rack switch</td>
</tr>
<tr>
<td>vCE</td>
<td>virtual Customer Edge Router</td>
</tr>
<tr>
<td>VM</td>
<td>Virtual Machine</td>
</tr>
<tr>
<td>VN</td>
<td>Virtual Network</td>
</tr>
<tr>
<td>VPC</td>
<td>Virtual Private Cloud</td>
</tr>
<tr>
<td>vPE</td>
<td>virtual Provider Edge Router</td>
</tr>
<tr>
<td>VPN</td>
<td>Virtual Private Network</td>
</tr>
<tr>
<td>VXLAN</td>
<td>Virtual eXtensible Local Area Network</td>
</tr>
<tr>
<td>WAN</td>
<td>Wide Area Network</td>
</tr>
</tbody>
</table>
2. Use Cases
2.1 Case 1: End-to-end BGP IP VPN cloud inter-connection

BGP/MPLS IP VPN is a proven scalable overlay technology with extensive deployment. It is an excellent candidate for end-to-end (host-to-host) overlay technology for Cloud-Scale DC application support. In addition, many SPs are interested to extend the IP VPN capabilities into their DCs to provide end-to-end native BGP IP VPN services to their enterprise customers.

BGP IP VPN provides routing isolation, rich policy control, and QoS support. The technologies developed to extend BGP IP VPN into DC servers or ToR are work in progress in IETF, [I-D.fang-l3vpn-virtual-pe], and [I-D.ietf-l3vpn-end-system].

The WAN and DC may be managed by the same or different administrative domains.

2.2 Case 2: Hybrid cloud inter-connection

Inter-connecting network SPs Enterprise IP VPNs to Cloud/Content providers DCs for expanded services. The inter-connection between the SP BGP/MPLS IP VPNs and the cloud provider networks is needed to provide the service end-to-end. The inter-connection of different types of providers can be BGP/MPLS IP VPN to other VPN or overlay technologies which may be virtualized or non-virtualized.

3. Architecture reference models

The architecture reference models described below focus on the interconnection aspects. Although the intra-DC implementation is not within the scope of this discussion, the intra-DC technology has a direct impact to inter-DC connection. Therefore, various models are illustrated.

3.1 BGP/MPLS IP VPN Inter-AS model

The BGP/MPLS IP VPNs are implemented in both the WAN network and the Data Center. A customer VPN, for example VPNA in figure 1, consists of enterprise remote sites and VMs supporting applications in the DC. The IP VPN implementation is using vPE technology in DC. The two segments of the VPNs are inter-connected through ASBRs facing each other in the respective networks.
Figure 1. BGP/MPLS IP VPN Inter-Connection with ASBR in each network

One boarding ASBR can be shared for the inter-connection of the two networks, especially if the WAN and DC belong to the same provider. Figure 2 illustrates this shared ASBR model.

Figure 2. BGP/MPLS IP VPN Inter-Connection with share ASBR

3.2 BGP/MPLS IP VPN Gateway PE to DC vCE Model
A simple virtual CE (vCE) [I-D.fang-l3vpn-virtual-ce] model can be used to inter-connect client containers to the DC Gateway which function as PE. This model is used by SPs to provide managed services, when scale can meet the service requirement.

---

Figure 3. BGP/MPLS IP VPN GW/PE to vCEs without BGP/MPLS IP VPN in the DC

### 3.3 Hybrid inter-connection model

The BGP/MPLS IP VPNs are implemented in the WAN network, and non BGP/MPLS IP VPN Overlay are implemented in the DC. The connection of the two networks is outside of the technologies for Inter-AS connections for BGP IP VPNs. This model includes many variations depending on the specific technologies used in the DC overlay. Figure 4 provides a general view of this inter-connecting model with ASBR on the MPLS WAN side, and the DC GW on the DC side. It is also viable to use one shared ASBR/GW for the inter-connection, especially if the WAN and the DC belong to the same provider.
Figure 4. BGP/MPLS IP VPN Inter-Connection with non BGP/MPLS IP VPN Overlay in DC

4. Inter-connect IP VPN between DC and WAN

4.1 Existing Inter-AS options and DCI gap analysis

The inter-AS options described in [RFC4364] can be used for DC interconnection. Option A, B, and C must be supported.

4.1.1 Option A pros and cons

In Option A: back-to-back VRF. The PE-ASBR in one AS performs MPLS or IP VPN de-encapsulation and transmits packets to the peer PE-ASBR in the adjacent AS. The peer PE-ASBR performs MPLS or IP VPN encapsulation on the customer IPv4/IPv6 packets received, and transmits the packet through the IP backbone of the AS. VPN service providers exchange routes across a back-to-back VRF connection. Each VRF instance represents a separate VPN client, and it is configured on a separate PE-ASBR interface, allowing a PE-ASBR to communicate with its peer PE-ASBR as if the peer was a CE router.

Pros: This is the most secure option among options A, B, and C. And it is the simplest model from operation perspective. Each PE-ASBR is treating the other as a CE.

Cons: This option suffers from scaling limitations, because per Inter-AS VPN VRF and interface are needed on the PE-ASBR.

Option A has been commonly used in BGP/MPLS VPN Inter-Provider inter-
connections because of the security considerations and its clear operational demarcation.

DCI considerations: This is a simple way to connect DC and WAN if both sides are of small scale. Scale will be the major concern for DC inter-connect if large scale support is needed. Even if the DC scale is small, there are major concerns on receiving relevant routes (potentially a large number) from the WAN side, and Vice Versa.

4.1.2 Option B pros and cons

In Option B: EBGP redistribution of labeled VPN-IPv4/IPv6 routes between the neighboring ASes. ASes exchange VPN routing information (routes and labels) to establish connections. To control connections between ASes, the PE routers and EBGP border edge routers maintain a label forwarding information base (LFIB). The LFIB manages the labels and routes that the PE routers and EBGP border edge routers receive during the exchange of VPN information. The ASes exchange VPN routing information, such as, the destination network, the next hop field associated with the distributing router, a local MPLS label, and an RD. ASBRs are configured to change the next hop (next-hop-self) when sending VPN-IPv4 NLRIIs to the IBGP neighbors; the ASBRs must allocate a new label when they forward the NLRI to the IBGP neighbors.

Pros: It provides improved scalability when compared with option A, since it removes the needs of per Inter-AS VPN VRF and interface on the ASBR.

Cons: vanilla version of Option B is considered less secure in comparison with Option A, due to the dynamic routing information exchange that is involved. The ASBR scaling may still be an issue because ASBR must maintain all VPN routes.

Option B is commonly used within single provider or for inter-provider connections.

DCI considerations: Option B is one viable option to be used in DC inter-connection. However, it has the same scale concerns as other options because of the potentially large number of routes exchanged between the WAN and the DC.

4.1.3 Option C pros and cons

In option C: Multihop eBGP redistribution of labeled VPN-IPv4/Ipv6 routes between source and destination ASs, with eBGP redistribution of labeled IPv4/IPv6 routes from AS to neighboring AS. The ASBRs need only to exchange host routes (/32 or /128) to the PE routers involved in the VPN, with the labels needed to get there. A Label Switch Path
(LSP) is built from the ingress PE router in one AS to the egress PE in the other AS (using Loopback addresses). VPN traffic uses this LSP to reach the other AS. From data plane’s perspective, the ASBRs act as P routers, with no knowledge about the VPNs concerned. Between the two inter-connecting ASBRs, the VPN traffic is treated just as between two P routers, each VPN data packet is pre-pended with the VPN label and then with an egress-PE label. Option C can be further scaled by using route reflectors (RRs) in each AS.

Pros: It is the most scalable option among all three. ASBR is no longer a bottle neck for VPN routes scaling as in Option B.

Cons: Major security issues as IGP reachability must be exchanged between the inter-connecting ASes.

Option C has seen used within a single SP for inter-AS connections. Using RR for VPN routes exchange is the common approach.

DCI consideration: Option C SHOULD NOT be used for any DCI which is between two different providers for security reasons.

In this option, though ASBR is not longer the scaling bottleneck, the scaling issues still call for careful design, as the total numbers of VRFs, VPN interfaces, and the VPN routes being exchanged between the two ASes can be very large.

4.1.4 Use of RTC

RT constraint [RFC4684] function must be used to only distribute the IP VPN routes of a VPN from one AS to another under the condition that they both support that VPN in each of the AS. This is one most important function for scalable solution.

However, all IP VPN routes are exchanged between the two ASes (e.g. WAN and DC) as long as they have to support the same VPNs. The potential IP VPN routes distribution can still be very substantial in large WAN and DC deployment. Additional aggregation and abstraction mechanisms can be used to avoid large numbers of VPN routes being exchanges across the border between the interconnecting WAN and the DC in either directions.

5. Inter-connect IP VPN and non-IP VPN overlay networks

As one significant instance of the hybrid use-case described in section 2.2, a DC may support a multi-tenant virtualized service network using IP based DC overlay encapsulations such as VXLAN [I-D.mahalingam-dutt-dcops-vxlan] or NVGRE [I-D.sridharan-virtualization-nvgre]. Different deployment models may
be used within the DC depending on the DC provider’s functional and operational requirements.

When an IP DC overlay is terminated at the DC Gateway router and traffic directed into a BGP/MPLS IP VPN, the DC Gateway router performs MPLS encapsulation towards the WAN and IP overlay based forwarding within the DC.

The inter-connection mechanisms between the DC and the WAN may fall into two categories:

1. VRF Termination

The overlay based virtual network terminates into a BGP IP VPN VRF at the DC-WAN Gateway router. Both the internal routes of the DC as well as the external routes received from the WAN router can be installed in the VRF forwarding table at the DC Gateway router. The DC Gateway performs an IP lookup, appropriate MPLS or IP encapsulation, and forward traffic.

The DC Gateway router peers with the WAN router using one of the existing inter-AS mechanisms described above. The DC Gateway functions as an IP-VPN ASBR with local VRFs; for example, packets still undergo an IP forwarding lookup.

2. DC-VN and IP VPN Inter-working

In this case, the DC Gateway router performs a direct translation between VN-IDs and IP VPN labels while switching packets between the DC and WAN interfaces without performing an IP lookup. The forwarding table at the DC Gateway router is set up to do a VN-ID or label lookup and derive the output label or VN-ID. The DC Gateway Router acts as an Inter-AS Option B ASBR peering with other ASBRs.

6. Security Considerations

BGP/MPLS Inter-AS security threats and defense techniques described in RFC 4111 [RFC4111] are applicable for the WAN/DC inter-connections.

In addition, the protocols between the Gateway routers and the controller/orchestrator 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.

7. IANA Considerations
8. References

8.1 Normative References


8.2 Informative References


Authors’ Addresses