BGP Configuration Guide

Configuring BGP for Access to Private IP Services (BGP/MPLS VPN Networks)

This Configuration Guide provides an overview of Private IP Services (BGP/MPLS VPN networks), compares methods for conveying customer routes to and from the provider network, and provides details on how to configure ProCurve Secure Routers as customer edge routers using BGP on the access link. For detailed information regarding specific command syntax, refer to the SROS Command Line Interface Reference Guide on your ProCurve SROS Documentation CD.

This guide consists of the following sections:

• Understanding Private IP Services on page 2
• Private IP Services: BGP/MPLS VPN Networks on page 2
• Connecting to the Private IP Service on page 4
  – Option 1: Static Routes on page 5
  – Option 2: RIP V2 on page 5
  – Option 3: OSPF on page 6
  – Option 4: BGP on page 6
• Configuration Examples on page 7
  – Example 1: Basic BGP Connectivity on page 7
  – Example 2: BGP Connectivity with Ingress Prefix Filters on page 9
• Appendix A: BGP Decision Logic in SROS J.03.01 on page 10
• Acronyms on page 11
Understanding Private IP Services

Provider-based WANs have long been servicing enterprise applications (where many branch offices connect to a corporate site) and retail applications (where thousands of stores connect to one or more data centers).

In the past, this type of application was often served by a Frame Relay/ATM network. As shown in Figure 1, the provider’s service was Layer 2, connecting and delivering PVCs to each customer site (where the customer router terminated each PVC). At the corporate or data center sites, the WAN router terminated PVCs from every remote. Each PVC was often implemented as a separate point-to-point sub-interface on the router. As the number of sites grew, router resources and PVC management became an issue. Also, since PVCs represent a recurring cost, adding mesh connectivity between sites with the Layer 2 solution became very expensive after reaching a certain size. For example, since a full mesh requires \( \frac{n(n-1)}{2} \) PVCs, four sites require six PVCs while eight sites require 28 PVCs.

A popular solution to this problem has been network-based Layer 3 VPNs (delivered as a service often called Private IP, PIP, Layer 3 VPN, IP VPN, or BGP/MPLS VPN Networks). This document briefly describes this service and discusses how customers can communicate routes to/from it. Details are then given on how to configure a ProCurve Secure Router as the customer router communicating with this service using BGP.

**Private IP Services: BGP/MPLS VPN Networks**

RFC2547 and updates to this RFC define the popular concept of network-based Layer 3 VPNs. While more complex for the provider to implement, complexity for the customer decreases significantly. In the end it is less complex for the provider because it aids in the IP convergence of their core network, eliminating multiple or overlay networks.

From the customer perspective, the fundamental difference between a Layer 2 and a Layer 3 implementation is as follows:

- With a Layer 2 WAN, the customer router views the provider as one or more Layer 2 pipes (PVCs), each connecting to one other customer site.
- With a Layer 3 WAN, the customer router sees the provider as a peer router on a point-to-point link. That peer router can have arbitrary connectivity to any of the customer’s other sites.
The customer router still connects to the provider network using familiar Layer 1 and 2 technologies (ADSL with PPPoE and T1 with Frame Relay are quite common), but this connectivity terminates at the local provider edge where the customer’s IP traffic enters the provider’s IP network. The customer packets are delivered by the network based on the provider’s knowledge of the customer’s IP networks. How the provider gets this knowledge is key to this document and is discussed later. But first, let’s look a little closer at the provider network.

**Provider Network Example**

In RFC2547 (BGP/MPLS VPNs March 1999), the customer router is referred to as a Customer Edge (CE) router. Internal to the provider’s Layer 3 network, are two types of routers: Provider Edge (PE) and Provider (P) routers. As shown in Figure 2, PE routers are located at the edge of the provider network and connect to CE and P routers. P routers only connect to PE and other P routers. They do not connect to CE routers.

---

The following points are vital to the internal operation of an RFC2547 network. Consider the red customer in Figure 3 on page 4. Red has several geographically dispersed sites connected by a Layer 3 WAN:

- Each PE router that connects to a red site must have knowledge of the networks at that site (at least those that are known as reachable across the WAN).
- Information about those networks is propagated via BGP1 directly to all other red-connecting PEs. This BGP is entirely separate from the BGP that might run on the CE router.
- P routers have no knowledge of this BGP communication and are not aware of any customer information. They only know the location of all P and PE routers. P and PE routers often use an internal gateway protocol (such as OSPF) to understand the physical connectivity of the provider’s internal network.

---

1 RFC2547 defines additions to BGP necessary to convey details about the VPNs within the network. These details are beyond the scope of this document.
• **Red**’s packets are kept entirely separate from the packets of other customers. This is where the term VPN is earned.
  
  – On the access link from CE to PE, this separation is performed by separate physical interfaces or by separate logical (virtual) interfaces (where multiple customers might share an access link such as VCs or VLANs).
  
  – Internal to the PE router, separation is performed by VRFs (Virtual Routing and Forwarding Instances). A VRF can be thought of as a logically separate route table within the PE that holds a single customer’s routes and cannot intermingle with another customer’s VRF in the same router.

• While moving from PE router to PE router across P routers, separation is performed using MPLS Label Switch Paths (LSPs). LSPs provide a dynamic PVC-like connection through the network that can quickly recalculate should a failure occur within the provider cloud.
  
  – PE routers are MPLS label edge routers in that they are the beginning and ending of MPLS LSPs used to switch customer packets through the provider network.
  
  – P routers are MPLS label switch routers. They simply switch customer packets on the LSPs.

**Figure 3. Customer Separation within a Private IP Provider**

**Connecting to the Private IP Service**

Using the basics of a Layer 3 VPN service, how does a PE learn about a connected customer’s networks in order to propagate that information to the PEs connecting the customer’s other sites? And, how does that PE convey customer network information learned from distant PEs to the connected CE? Is a routing protocol required on the CE-PE link?

Many people are under the impression that the CE router MUST use BGP to communicate this information to the PE. This is absolutely not true. However, there are reasons why BGP is sometimes the best option.

The following pages explore the pros and cons of several options. The ultimate decision depends on the requirements of the customer and the provider.
Option 1: Static Routes
The PE router has static routes configured representing the networks of a connected customer site. The next-hop gateway is typically the connected CE. The CE router may have discrete static routes or a static default route with the next hop gateway set to the PE.

**Pros**
- Simple to configure and operate.
- No protocol bandwidth overhead.
- No protocol processing overhead. Scales well on PE routers with many customers.

**Cons**
- Except for local link failure (Layer 1 or 2), there is no dynamic detection of a failure in the provider or customer network (whether used for basic detection of liveliness or for resolving routes).
- Alternate paths and dynamic detection are not provided.
- Requires redistribution to/from BGP at the PE router. This increases the chance of errors in the PE configuration.
- Customer site changes that alter the network information representing that site require the provider to make manual modifications on the local PE router. This requires provider interaction, takes time, and increases the chance of errors in the PE configuration.

Option 2: RIP V2
CE and PE routers exchange customer network information over the CE-PE link using RIP V2. Many people forget that this protocol runs only on the CE-PE link. Unlike Layer 2 WANs, the hub router is not concerned with periodic updates from each remote due to concentration of PVCs. In Layer 3 WANS, the periodic updates are local to each CE-PE link.

**Pros**
- Simple to configure and operate.
- Low protocol bandwidth overhead; one packet in each direction every 30 seconds (compared to hello packets used by other protocols that are typically sent every 10 seconds).
- Low protocol processing overhead. Scales reasonably well on PE routers with many customers.
- Perfectly suitable for indicating basic liveliness in both directions.
- Can be suitable for resolving mutually exclusive paths such as dial backup.
- Customer site changes are propagated dynamically to the PE router.

**Cons**
- Not recommended for non-stub applications where there are multiple paths active at one time between customer and provider, or directly between customer sites.

**Note:** A backup scheme where either a primary or backup link is active at one time can be considered a stub. May have to consider how routing information is managed in the primary and backup network during failure.
- Requires redistribution to/from BGP at the PE router. Increases the chance of errors in the PE configuration.
Option 3: OSPF
CE and PE routers exchange customer network information over the CE-PE link using OSPF.

**Pros**
- Suitable for indicating liveliness in both directions.
- Protocol bandwidth overhead is low with the proper design.
- Suitable for complex non-stub applications where there are multiple paths active at one time.
- Customer site changes are propagated dynamically to the PE router.

**Cons**
- Not simple to configure. OSPF settings must be coordinated with the provider.
- Protocol bandwidth can be large during topology changes, depending on the network design and number of routes to propagate.
- Protocol processing overhead can be significant, especially on PE routers with many customers. (Some providers refuse OSPF for this reason.)
- Requires redistribution to/from BGP at the PE router. Increases the chance of errors in the PE configuration.
- Widely agreed that this protocol is overkill for the application.

Option 4: BGP
CE and PE routers exchange customer network information over the CE-PE link using BGP.

**Pros**
- Medium protocol bandwidth overhead.
- Medium protocol processing overhead. Scales well on PE routers with many customers.
- Suitable for indicating liveliness in both directions.
- Suitable for complex applications where there are numerous paths active at one time.
- Customer site changes are propagated dynamically to the PE router.
- Is the native protocol of the provider router (no concern over protocol redistribution).
- Policy-based, providing robust policy control over what is sent and received.
- Though manual configuration could be viewed as a negative, the tight policy controls allowed by the protocol help customers protect themselves from provider misconfiguration and failure. With proper subnet design, the need to change BGP network settings over time can be eliminated.

**Cons**
- Not simple to configure. Certain settings must be coordinated with the provider.
- Policy controls can require significant manual configuration on each router. The more policy controls in use, the more configuration required.
Configuration Examples

This section provides configuration examples for deploying a ProCurve Secure Router as the customer-located CE router when connecting to a private IP network service (Layer 3 VPN), using BGP as the CE-PE routing protocol. These examples represent any of the ProCurve Secure Router 7000dl Series units. Slight modifications may be required for use on specific platforms.

Example 1: Basic BGP Connectivity

This example illustrates basic BGP connectivity with a Layer 3 provider over a single link. The provider’s BGP AS is 65000. The customer AS is 65001. BGP allows all sites to know when other sites are available.

The customer has assigned each customer site a /24 prefix from 10.10.x.0, where x represents the site number. In this example, the site number is 1. At each site, the /24 prefix is variably subnetted on local LANs. Only the /24 prefix is advertised to the provider network for propagation to other sites. The provider delivers each customer /24 prefix to the other customer sites.

Some configuration items not directly related to BGP connectivity are omitted.

Customer Site 1 CE Router:

```plaintext
! hostname "Site 1"
! ip routing
!
interface loop 1
   description Management Interface
   ip address 10.10.1.254 255.255.255.255
   no shutdown
!
interface eth 0/1
   description Local LAN
   ip address 10.10.1.193 255.255.255.224
   no shutdown
!
```

Figure 4. Single Connection to a Private IP Provider Network

![Figure 4. Single Connection to a Private IP Provider Network](image-url)
interface t1 1/1
    description NtwkT1 CktID 54321
    tdm-group 1 timeslots 1-24 speed 64
    no shutdown
!
interface fr 1 point-to-point
    description WANFR CktID 67890
    frame-relay lmi-type ansi
    no shutdown
    bind 1 t1 1/1 1 frame-relay 1
!
interface fr 1.17 point-to-point
    description PVC17
    frame-relay interface-dlci 17
    frame-relay bc 768000
    description Frame Relay 1.17 DLCI-17
    ip address 64.19.120.2 255.255.255.252
    bandwidth 768
!
!
router bgp 65001
    no auto-summary
    no synchronization
    bgp router-id 10.10.1.254
    network 10.10.1.0 mask 255.255.255.0
    neighbor 64.19.120.3
        no default-originate
        description BGP AS-65001 Peer-65000
        soft-reconfiguration inbound
        update-source loop 1
        remote-as 65000
!
!
ip route 10.10.1.0 255.255.255.0 null 0
! This is the /24 prefix allocated to this site. Since all local interfaces are
! variably subnetted from this address (longer prefixes), this exact prefix does not
! exist in the route table. A static route to null0 allows the prefix to exist so that
! the BGP network statement will inject the /24 route into BGP. Packets received will be
! routed to the interface with the more specific matching prefix. Any packets to a
! non-existing more specific prefix will be dropped by this static route.
!
!
o no ip http server

!
!
ip ftp source-interface loopback 1
! all mgt functions use mgt addr (lo 1).
ip radius source-interface loopback 1
ip tftp source-interface loopback 1
ip snntp source-interface loopback 1
logging forwarding source-interface loopback 1
!
line con 0
  no login
    line-timeout 60
!
line telnet 0 4
  login
    line-timeout 60
!
end

Example 2: BGP Connectivity with Ingress Prefix Filters

This configuration builds on Example 1 and adds ingress prefix filters to control which routes are accepted from the provider. The filters allow routes from a specific address range (those of other customer sites) to be accepted by the CE router. (Only the sections changed from Example 1 are shown here.)

router bgp 65001
  no auto-summary
  no synchronization
  bgp router-id 10.10.1.254
  network 10.10.1.0 mask 255.255.255.0
  neighbor 64.19.120.1
    no default-originate
    description BGP AS-65001 Peer-65000
    prefix-list Corp_Nets in ! To guard against provider errors that might breach corporate security,
    ! filter networks received from provider to only allowed networks.
    ! The following prefix list is used to limit the prefixes allowed from the provider.
    ! Routes from other site subnets (10.10.x.0 with a /24 prefix length) are allowed.
    !
    ip prefix-list Corp_Nets description "Customer VPN Connections"
    ip prefix-list Corp_Nets seq 20 permit 10.10.0.0/16 ge 24 le 24
Appendix A: BGP Decision Logic in SROS J.03.01

The following is the BGP decision logic used by ProCurve Secure Routers in release J.03.01 when comparing two BGP routes:

• Ignore route if next hop is unreachable.
• Prefer route with largest local preference.
• Prefer route with shortest AS_PATH.
• Prefer route injected by this router via network command, then redistribute (future) command, then aggregate (future) command.
• Prefer route with lowest origin type. Routes originally injected by the network command or aggregation (IGP) will be of lower origin than those originally injected by redistribution into BGP (incomplete).
• Prefer routes with lowest MED value.
• Prefer eBGP over iBGP.
  – Equivalent routes at this point are considered multipath candidates if in same AS.
• Compare IGP metric (administrative distance).
• Compare other, generic-route metrics.
• Compare incoming router ID.
• Compare neighbor IP address.
**Acronyms**

ADSL...............Asymmetric Digital Subscriber Line  
AS ....................Autonomous System  
ATM .................Asynchronous Transfer Mode  
BGP ..................Border Gateway Protocol  
CE ....................Customer Edge  
eBGP ..................External BGP  
iBGP ..................Internal BGP  
IGP ..................Interior Gateway Protocol  
IP .....................Internet Protocol  
LAN .................Local Area Network  
LSP ..................Label Switch Path  
MED ................Multi-Exit Discriminator  
MPLS ...............Multiprotocol Label Switching  
OSPF ...............Open Shortest Path First  
P .......................Provider  
PE .....................Provider Edge  
PIP ....................Private Internet Protocol  
PPPoE ..............Point-to-Point Protocol over Ethernet  
PVC ..................Permanent Virtual Circuit  
RIP ..................Routing Information Protocol  
SROS ...............Secure Router Operating System  
VC ....................Virtual Circuit  
VLAN ..............Virtual Local Area Network  
VPN ..................Virtual Private Network  
VRF ..................Virtual Routing and Forwarding  
WAN .................Wide Area Network

**References**

