Network Working Group
Request for Comments: 4364
Obsoletes: 2547
Category: Standards Track

E. Rosen
Cisco Systems, Inc.
Y. Rekhter
Juniper Networks, Inc.

February 2006

BGP/MPLS IP Virtual Private Networks (VPNs)

Status of This Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

Abstract

This document describes a method by which a Service Provider may use an IP backbone to provide IP Virtual Private Networks (VPNs) for its customers. This method uses a "peer model", in which the customers' edge routers (CE routers) send their routes to the Service Provider's edge routers (PE routers); there is no "overlay" visible to the customer's routing algorithm, and CE routers at different sites do not peer with each other. Data packets are tunneled through the backbone, so that the core routers do not need to know the VPN routes.

This document obsoletes RFC 2547.

Table of Contents

1. Introduction

This document describes a method by which a Service Provider may use an IP backbone to provide IP Virtual Private Networks (VPNs) for its customers. This method uses a "peer model", in which the customers' edge routers (CE routers) send their routes to the Service Provider's edge routers (PE routers). Border Gateway Protocol (BGP) [BGP, BGP-MP] is then used by the Service Provider to exchange the routes of a particular VPN among the PE routers that are attached to that VPN. This is done in a way that ensures that routes from different VPNs remain distinct and separate, even if two VPNs have an overlapping address space. The PE routers distribute, to the CE routers in a particular VPN, the routes from other the CE routers in that VPN. The CE routers do not peer with each other, hence there is no "overlay" visible to the VPN's routing algorithm. The term "IP" in "IP VPN" is used to indicate that the PE receives IP datagrams from the CE, examines their IP headers, and routes them accordingly.

Each route within a VPN is assigned a Multiprotocol Label Switching (MPLS) [MPLS-ARCH, MPLS-BGP, MPLS-ENCAPS] label; when BGP distributes a VPN route, it also distributes an MPLS label for that route. Before a customer data packet travels across the Service Provider's backbone, it is encapsulated with the MPLS label that corresponds, in the customer's VPN, to the route that is the best match to the packet's destination address. This MPLS packet is further encapsulated (e.g., with another MPLS label or with an IP or Generic Routing Encapsulation (GRE) tunnel header [MPLS-in-IP-GRE]) so that it gets tunneled across the backbone to the proper PE router. Thus, the backbone core routers do not need to know the VPN routes.

The primary goal of this method is to support the case in which a client obtains IP backbone services from a Service Provider or Service Providers with which it maintains contractual relationships. The client may be an enterprise, a group of enterprises that need an extranet, an Internet Service Provider, an application service provider, another VPN Service Provider that uses this same method to offer VPNs to clients of its own, etc. The method makes it very simple for the client to use the backbone services. It is also very scalable and flexible for the Service Provider, and allows the Service Provider to add value.

1.1. Virtual Private Networks

Consider a set of "sites" that are attached to a common network that we call "the backbone". Now apply some policy to create a number of subsets of that set, and impose the following rule: two sites may have IP interconnectivity over that backbone only if at least one of these subsets contains them both.

These subsets are Virtual Private Networks (VPNs). Two sites have IP connectivity over the common backbone only if there is some VPN that contains them both. Two sites that have no VPN in common have no connectivity over that backbone.

If all the sites in a VPN are owned by the same enterprise, the VPN may be thought of as a corporate "intranet". If the various sites in a VPN are owned by different enterprises, the VPN may be thought of as an "extranet". A site can be in more than one VPN; e.g., in an intranet and in several extranets. In general, when we use the term "VPN" we will not be distinguishing between intranets and extranets.

We refer to the owners of the sites as the "customers". We refer to the owners/operators of the backbone as the "Service Providers" (SPs). The customers obtain "VPN service" from the SPs.

A customer may be a single enterprise, a set of enterprises, an Internet Service Provider, an Application Service Provider, another SP that offers the same kind of VPN service to its own customers, etc.

The policies that determine whether a particular collection of sites is a VPN are the policies of the customers. Some customers will want the implementation of these policies to be entirely the responsibility of the SP. Other customers may want to share with the SP the responsibility for implementing these policies. This document specifies mechanisms that can be used to implement these policies. The mechanisms we describe are general enough to allow these policies to be implemented either by the SP alone or by a VPN customer together with the SP. Most of the discussion is focused on the former case, however.

The mechanisms discussed in this document allow the implementation of a wide range of policies. For example, within a given VPN, one can allow every site to have a direct route to every other site ("full mesh"). Alternatively, one can force traffic between certain pairs of sites to be routed via a third site. This can be useful, e.g., if it is desired that traffic between a pair of sites be passed through a firewall, and the firewall is located at the third site.

In this document, we restrict our discussion to the case in which the customer is explicitly purchasing VPN service from an SP, or from a set of SPs that have agreed to cooperate to provide the VPN service. That is, the customer is not merely purchasing internet access from an SP, and the VPN traffic does not pass through a random collection of interconnected SP networks.

We also restrict our discussion to the case in which the backbone provides an IP service to the customer, rather than, e.g., a layer 2 service such as Frame Relay, Asynchronous Transfer Mode (ATM), ethernet, High Level Data Link Control (HDLC), or Point-to-Point Protocol (PPP). The customer may attach to the backbone via one of these (or other) layer 2 services, but the layer 2 service is terminated at the "edge" of the backbone, where the customer's IP datagrams are removed from any layer 2 encapsulation.

In the rest of this introduction, we specify some properties that VPNs should have. The remainder of this document specifies a set of mechanisms that can be deployed to provide a VPN model that has all these properties. This section also introduces some of the technical terminology used in the remainder of the document.

1.2. Customer Edge and Provider Edge

Routers can be attached to each other, or to end systems, in a variety of different ways: PPP connections, ATM Virtual Circuits (VCs), Frame Relay VCs, ethernet interfaces, Virtual Local Area Networks (VLANs) on ethernet interfaces, GRE tunnels, Layer 2 Tunneling Protocol (L2TP) tunnels, IPsec tunnels, etc. We will use the term "attachment circuit" to refer generally to some such means of attaching to a router. An attachment circuit may be the sort of connection that is usually thought of as a "data link", or it may be a tunnel of some sort; what matters is that it be possible for two devices to be network layer peers over the attachment circuit.

Each VPN site must contain one or more Customer Edge (CE) devices. Each CE device is attached, via some sort of attachment circuit, to one or more Provider Edge (PE) routers.

Routers in the SP's network that do not attach to CE devices are known as "P routers".

CE devices can be hosts or routers. In a typical case, a site contains one or more routers, some of which are attached to PE routers. The site routers that attach to the PE routers would then be the CE devices, or "CE routers". However, there is nothing to prevent a non-routing host from attaching directly to a PE router, in which case the host would be a CE device.

Sometimes, what is physically attached to a PE router is a layer 2 switch. In this case, we do NOT say that the layer 2 switch is a CE device. Rather, the CE devices are the hosts and routers that communicate with the PE router through the layer 2 switch; the layer 2 infrastructure is transparent. If the layer 2 infrastructure provides a multipoint service, then multiple CE devices can be attached to the PE router over the same attachment circuit.

CE devices are logically part of a customer's VPN. PE and P routers are logically part of the SP's network.

The attachment circuit over which a packet travels when going from CE to PE is known as that packet's "ingress attachment circuit", and the PE as the packet's "ingress PE". The attachment circuit over which a packet travels when going from PE to CE is known as that packet's "egress attachment circuit", and the PE as the packet's "egress PE".

We will say that a PE router is attached to a particular VPN if it is attached to a CE device that is in a site of that VPN. Similarly, we will say that a PE router is attached to a particular site if it is attached to a CE device that is in that site.

When the CE device is a router, it is a routing peer of the PE(s) to which it is attached, but it is NOT a routing peer of CE routers at other sites. Routers at different sites do not directly exchange routing information with each other; in fact, they do not even need to know of each other at all. As a consequence, the customer has no backbone or "virtual backbone" to manage, and does not have to deal with any inter-site routing issues. In other words, in the scheme described in this document, a VPN is NOT an "overlay" on top of the SP's network.

With respect to the management of the edge devices, clear administrative boundaries are maintained between the SP and its customers. Customers are not required to access the PE or P routers for management purposes, nor is the SP required to access the CE devices for management purposes.

1.3. VPNs with Overlapping Address Spaces

If two VPNs have no sites in common, then they may have overlapping address spaces. That is, a given address might be used in VPN V1 as the address of system S1, but in VPN V2 as the address of a completely different system S2. This is a common situation when the VPNs each use an RFC 1918 private address space. Of course, within each VPN, each address must be unambiguous.

Even two VPNs that do have sites in common may have overlapping address spaces, as long as there is no need for any communication between systems with such addresses and systems in the common sites.

1.4. VPNs with Different Routes to the Same System

Although a site may be in multiple VPNs, it is not necessarily the case that the route to a given system at that site should be the same in all the VPNs. Suppose, for example, we have an intranet consisting of sites A, B, and C, and an extranet consisting of A, B, C, and the "foreign" site D. Suppose that at site A there is a server, and we want clients from B, C, or D to be able to use that server. Suppose also that at site B there is a firewall. We want all the traffic from site D to the server to pass through the firewall, so that traffic from the extranet can be access controlled. However, we don't want traffic from C to pass through the firewall on the way to the server, since this is intranet traffic.

It is possible to set up two routes to the server. One route, used by sites B and C, takes the traffic directly to site A. The second route, used by site D, takes the traffic instead to the firewall at site B. If the firewall allows the traffic to pass, it then appears to be traffic coming from site B, and follows the route to site A.

1.5. SP Backbone Routers

The SP's backbone consists of the PE routers, as well as other routers ("P routers") that do not attach to CE devices.

If every router in an SP's backbone had to maintain routing information for all the VPNs supported by the SP, there would be severe scalability problems; the number of sites that could be supported would be limited by the amount of routing information that could be held in a single router. It is important therefore that the routing information about a particular VPN only needs to be present in the PE routers that attach to that VPN. In particular, the P routers do not need to have ANY per-VPN routing information whatsoever. (This condition may need to be relaxed somewhat when multicast routing is considered. This is not considered further in this paper, but is examined in [VPN-MCAST].)

So just as the VPN owners do not have a backbone or "virtual backbone" to administer, the SPs themselves do not have a separate backbone or "virtual backbone" to administer for each VPN. Site-to-site routing in the backbone is optimal (within the constraints of the policies used to form the VPNs) and is not constrained in any way by an artificial "virtual topology" of tunnels.

Section 10 discusses some of the special issues that arise when the backbone spans several Service Providers.

1.6. Security

VPNs of the sort being discussed here, even without making use of cryptographic security measures, are intended to provide a level of security equivalent to that obtainable when a layer 2 backbone (e.g., Frame Relay) is used. That is, in the absence of misconfiguration or deliberate interconnection of different VPNs, it is not possible for systems in one VPN to gain access to systems in another VPN. Of course, the methods described herein do not by themselves encrypt the data for privacy, nor do they provide a way to determine whether data has been tampered with en route. If this is desired, cryptographic measures must be applied in addition. (See, e.g., [MPLS/BGP-IPsec].) Security is discussed in more detail in Section 13.

2. Sites and CEs

From the perspective of a particular backbone network, a set of IP systems may be regarded as a "site" if those systems have mutual IP interconnectivity that doesn't require use of the backbone. In general, a site will consist of a set of systems that are in geographic proximity. However, this is not universally true. If two geographic locations are connected via a leased line, over which Open Shortest Path First (OSPF) protocol [OSPFv2] is running, and if that line is the preferred way of communicating between the two locations, then the two locations can be regarded as a single site, even if each location has its own CE router. (This notion of "site" is topological, rather than geographical. If the leased line goes down, or otherwise ceases to be the preferred route, but the two geographic locations can continue to communicate by using the VPN backbone, then one site has become two.)

A CE device is always regarded as being in a single site (though as we shall see in Section 3.2, a site may consist of multiple "virtual sites"). A site, however, may belong to multiple VPNs.

A PE router may attach to CE devices from any number of different sites, whether those CE devices are in the same or in different VPNs. A CE device may, for robustness, attach to multiple PE routers, of the same or of different service providers. If the CE device is a router, the PE router and the CE router will appear as router adjacencies to each other.

While we speak mostly of "sites" as being the basic unit of interconnection, nothing here prevents a finer degree of granularity in the control of interconnectivity. For example, certain systems at

a site may be members of an intranet as well as members of one or more extranets, while other systems at the same site may be restricted to being members of the intranet only. However, this might require that the site have two attachment circuits to the backbone, one for the intranet and one for the extranet; it might further require that firewall functionality be applied on the extranet attachment circuit.

3. VRFs: Multiple Forwarding Tables in PEs

Each PE router maintains a number of separate forwarding tables. One of the forwarding tables is the "default forwarding table". The others are "VPN Routing and Forwarding tables", or "VRFs".

3.1. VRFs and Attachment Circuits

Every PE/CE attachment circuit is associated, by configuration, with one or more VRFs. An attachment circuit that is associated with a VRF is known as a "VRF attachment circuit".

In the simplest case and most typical case, a PE/CE attachment circuit is associated with exactly one VRF. When an IP packet is received over a particular attachment circuit, its destination IP address is looked up in the associated VRF. The result of that lookup determines how to route the packet. The VRF used by a packet's ingress PE for routing a particular packet is known as the packet's "ingress VRF". (There is also the notion of a packet's "egress VRF", located at the packet's egress PE; this is discussed in Section 5.)

If an IP packet arrives over an attachment circuit that is not associated with any VRF, the packet's destination address is looked up in the default forwarding table, and the packet is routed accordingly. Packets forwarded according to the default forwarding table include packets from neighboring P or PE routers, as well as packets from customer-facing attachment circuits that have not been associated with VRFs.

Intuitively, one can think of the default forwarding table as containing "public routes", and of the VRFs as containing "private routes". One can similarly think of VRF attachment circuits as being "private", and of non-VRF attachment circuits as being "public".

If a particular VRF attachment circuit connects site S to a PE router, then connectivity from S (via that attachment circuit) can be restricted by controlling the set of routes that gets entered in the corresponding VRF. The set of routes in that VRF should be limited to the set of routes leading to sites that have at least one VPN in

common with S. Then a packet sent from S over a VRF attachment circuit can only be routed by the PE to another site S' if S' is in one of the same VPNs as S. That is, communication (via PE routers) is prevented between any pair of VPN sites that have no VPN in common. Communication between VPN sites and non-VPN sites is prevented by keeping the routes to the VPN sites out of the default forwarding table.

If there are multiple attachment circuits leading from S to one or more PE routers, then there might be multiple VRFs that could be used to route traffic from S. To properly restrict S's connectivity, the same set of routes would have to exist in all the VRFs. Alternatively, one could impose different connectivity restrictions over different attachment circuit from S. In that case, some of the VRFs associated with attachment circuits from S would contain different sets of routes than some of the others.

We allow the case in which a single attachment circuit is associated with a set of VRFs, rather than with a single VRF. This can be useful if it is desired to divide a single VPN into several "sub-VPNs", each with different connectivity restrictions, where some characteristic of the customer packets is used to select from among the sub-VPNs. For simplicity though, we will usually speak of an attachment circuit as being associated with a single VRF.

3.2. Associating IP Packets with VRFs

When a PE router receives a packet from a CE device, it must determine the attachment circuit over which the packet arrived, as this determines in turn the VRF (or set of VRFs) that can be used for forwarding that packet. In general, to determine the attachment circuit over which a packet arrived, a PE router takes note of the physical interface over which the packet arrived, and possibly also takes note of some aspect of the packet's layer 2 header. For example, if a packet's ingress attachment circuit is a Frame Relay VC, the identity of the attachment circuit can be determined from the physical Frame Relay interface over which the packet arrived, together with the Data Link Connection Identifier (DLCI) field in the packet's Frame Relay header.

Although the PE's conclusion that a particular packet arrived on a particular attachment circuit may be partially determined by the packet's layer 2 header, it must be impossible for a customer, by writing the header fields, to fool the SP into thinking that a packet that was received over one attachment circuit really arrived over a different one. In the example above, although the attachment circuit is determined partially by inspection of the DLCI field in the Frame Relay header, this field cannot be set freely by the customer.

Rather, it must be set to a value specified by the SP, or else the packet cannot arrive at the PE router.

In some cases, a particular site may be divided by the customer into several "virtual sites". The SP may designate a particular set of VRFs to be used for routing packets from that site and may allow the customer to set some characteristic of the packet, which is then used for choosing a particular VRF from the set.

For example, each virtual site might be realized as a VLAN. The SP and the customer could agree that on packets arriving from a particular CE, certain VLAN values would be used to identify certain VRFs. Of course, packets from that CE would be discarded by the PE if they carry VLAN tag values that are not in the agreed-upon set. Another way to accomplish this is to use IP source addresses. In this case, the PE uses the IP source address in a packet received from the CE, along with the interface over which the packet is received, to assign the packet to a particular VRF. Again, the customer would only be able to select from among the particular set of VRFs that that customer is allowed to use.

If it is desired to have a particular host be in multiple virtual sites, then that host must determine, for each packet, which virtual site the packet is associated with. It can do this, e.g., by sending packets from different virtual sites on different VLANs, or out different network interfaces.

3.3. Populating the VRFs

With what set of routes are the VRFs populated?

As an example, let PE1, PE2, and PE3 be three PE routers, and let CE1, CE2, and CE3 be three CE routers. Suppose that PE1 learns, from CE1, the routes that are reachable at CE1's site. If PE2 and PE3 are attached, respectively, to CE2 and CE3, and there is some VPN V containing CE1, CE2, and CE3, then PE1 uses BGP to distribute to PE2 and PE3 the routes that it has learned from CE1. PE2 and PE3 use these routes to populate the VRFs that they associate, respectively, with the sites of CE2 and CE3. Routes from sites that are not in VPN V do not appear in these VRFs, which means that packets from CE2 or CE3 cannot be sent to sites that are not in VPN V.

When we speak of a PE "learning" routes from a CE, we are not presupposing any particular learning technique. The PE may learn routes by means of a dynamic routing algorithm, but it may also "learn" routes by having those routes configured (i.e., static routing). (In this case, to say that the PE "learned" the routes from the CE is perhaps to exercise a bit of poetic license.)

PEs also need to learn, from other PEs, the routes that belong to a given VPN. The procedures to be used for populating the VRFs with the proper sets of routes are specified in Section 4.

If there are multiple attachment circuits leading from a particular PE router to a particular site, they might all be mapped to the same forwarding table. But if policy dictates, they could be mapped to different forwarding tables. For instance, the policy might be that a particular attachment circuit from a site is used only for intranet traffic, while another attachment circuit from that site is used only for extranet traffic. (Perhaps, e.g., the CE attached to the extranet attachment circuit is a firewall, while the CE attached to the intranet attachment circuit is not.) In this case, the two attachment circuits would be associated with different VRFs.

Note that if two attachment circuits are associated with the same VRF, then packets that the PE receives over one of them will be able to reach exactly the same set of destinations as packets that the PE receives over the other. So two attachment circuits cannot be associated with the same VRF unless each CE is in the exact same set of VPNs as is the other.

If an attachment circuit leads to a site which is in multiple VPNs, the attachment circuit may still associated with a single VRF, in which case the VRF will contain routes from the full set of VPNs of which the site is a member.

4. VPN Route Distribution via BGP

PE routers use BGP to distribute VPN routes to each other (more accurately, to cause VPN routes to be distributed to each other).

We allow each VPN to have its own address space, which means that a given address may denote different systems in different VPNs. If two routes to the same IP address prefix are actually routes to different systems, it is important to ensure that BGP not treat them as comparable. Otherwise, BGP might choose to install only one of them, making the other system unreachable. Further, we must ensure that POLICY is used to determine which packets get sent on which routes; given that several such routes are installed by BGP, only one such must appear in any particular VRF.

We meet these goals by the use of a new address family, as specified below.

4.1. The VPN-IPv4 Address Family

The BGP Multiprotocol Extensions [BGP-MP] allow BGP to carry routes from multiple "address families". We introduce the notion of the "VPN-IPv4 address family". A VPN-IPv4 address is a 12-byte quantity, beginning with an 8-byte Route Distinguisher (RD) and ending with a 4-byte IPv4 address. If several VPNs use the same IPv4 address prefix, the PEs translate these into unique VPN-IPv4 address prefixes. This ensures that if the same address is used in several different VPNs, it is possible for BGP to carry several completely different routes to that address, one for each VPN.

Since VPN-IPv4 addresses and IPv4 addresses are different address families, BGP never treats them as comparable addresses.

An RD is simply a number, and it does not contain any inherent information; it does not identify the origin of the route or the set of VPNs to which the route is to be distributed. The purpose of the RD is solely to allow one to create distinct routes to a common IPv4 address prefix. Other means are used to determine where to redistribute the route (see Section 4.3).

The RD can also be used to create multiple different routes to the very same system. We have already discussed a situation in which the route to a particular server should be different for intranet traffic than for extranet traffic. This can be achieved by creating two different VPN-IPv4 routes that have the same IPv4 part, but different RDs. This allows BGP to install multiple different routes to the same system, and allows policy to be used (see Section 4.3.5) to decide which packets use which route.

The RDs are structured so that every Service Provider can administer its own "numbering space" (i.e., can make its own assignments of RDs), without conflicting with the RD assignments made by any other Service Provider. An RD consists of three fields: a 2-byte type field, an administrator field, and an assigned number field. The value of the type field determines the lengths of the other two fields, as well as the semantics of the administrator field. The administrator field identifies an assigned number authority, and the assigned number field contains a number that has been assigned, by the identified authority, for a particular purpose. For example, one could have an RD whose administrator field contains an Autonomous System number (ASN), and whose (4-byte) number field contains a number assigned by the SP to whom that ASN belongs (having been assigned to that SP by the appropriate authority).

RDs are given this structure in order to ensure that an SP that provides VPN backbone service can always create a unique RD when it

needs to do so. However, the structure is not meaningful to BGP; when BGP compares two such address prefixes, it ignores the structure entirely.

A PE needs to be configured such that routes that lead to a particular CE become associated with a particular RD. The configuration may cause all routes leading to the same CE to be associated with the same RD, or it may cause different routes to be associated with different RDs, even if they lead to the same CE.

4.2. Encoding of Route Distinguishers

As stated, a VPN-IPv4 address consists of an 8-byte Route Distinguisher followed by a 4-byte IPv4 address. The RDs are encoded as follows:

- Type Field: 2 bytes
- Value Field: 6 bytes

The interpretation of the Value field depends on the value of the type field. At the present time, three values of the type field are defined: 0, 1, and 2.

- Type 0: The Value field consists of two subfields:

  * Administrator subfield: 2 bytes
  * Assigned Number subfield: 4 bytes

  The Administrator subfield must contain an Autonomous System number. If this ASN is from the public ASN space, it must have been assigned by the appropriate authority (use of ASN values from the private ASN space is strongly discouraged). The Assigned Number subfield contains a number from a numbering space that is administered by the enterprise to which the ASN has been assigned by an appropriate authority.

- Type 1: The Value field consists of two subfields:

  * Administrator subfield: 4 bytes
  * Assigned Number subfield: 2 bytes

  The Administrator subfield must contain an IP address. If this IP address is from the public IP address space, it must have been assigned by an appropriate authority (use of addresses from the private IP address space is strongly discouraged). The Assigned Number subfield contains a number from a numbering space which is administered by the enterprise to which the IP address has been assigned.

- Type 2: The Value field consists of two subfields:

  * Administrator subfield: 4 bytes
  * Assigned Number subfield: 2 bytes

  The Administrator subfield must contain a 4-byte Autonomous System number [BGP-AS4]. If this ASN is from the public ASN space, it must have been assigned by the appropriate authority (use of ASN values from the private ASN space is strongly discouraged). The Assigned Number subfield contains a number from a numbering space which is administered by the enterprise to which the ASN has been assigned by an appropriate authority.

4.3. Controlling Route Distribution

In this section, we discuss the way in which the distribution of the VPN-IPv4 routes is controlled.

If a PE router is attached to a particular VPN (by being attached to a particular CE in that VPN), it learns some of that VPN's IP routes from the attached CE router. Routes learned from a CE routing peer over a particular attachment circuit may be installed in the VRF associated with that attachment circuit. Exactly which routes are installed in this manner is determined by the way in which the PE learns routes from the CE. In particular, when the PE and CE are routing protocol peers, this is determined by the decision process of the routing protocol; this is discussed in Section 7.

These routes are then converted to VPN-IP4 routes, and "exported" to BGP. If there is more than one route to a particular VPN-IP4 address prefix, BGP chooses the "best" one, using the BGP decision process. That route is then distributed by BGP to the set of other PEs that need to know about it. At these other PEs, BGP will again choose the best route for a particular VPN-IP4 address prefix. Then the chosen VPN-IP4 routes are converted back into IP routes, and "imported" into one or more VRFs. Whether they are actually installed in the VRFs depends on the decision process of the routing method used between the PE and those CEs that are associated with the VRF in question. Finally, any route installed in a VRF may be distributed to the associated CE routers.

4.3.1. The Route Target Attribute

Every VRF is associated with one or more Route Target (RT) attributes.

When a VPN-IPv4 route is created (from an IPv4 route that the PE has learned from a CE) by a PE router, it is associated with one or more

Route Target attributes. These are carried in BGP as attributes of the route.

Any route associated with Route Target T must be distributed to every PE router that has a VRF associated with Route Target T. When such a route is received by a PE router, it is eligible to be installed in those of the PE's VRFs that are associated with Route Target T. (Whether it actually gets installed depends upon the outcome of the BGP decision process, and upon the outcome of the decision process of the IGP (i.e., the intra-domain routing protocol) running on the PE/CE interface.)

A Route Target attribute can be thought of as identifying a set of sites. (Though it would be more precise to think of it as identifying a set of VRFs.) Associating a particular Route Target attribute with a route allows that route to be placed in the VRFs that are used for routing traffic that is received from the corresponding sites.

There is a set of Route Targets that a PE router attaches to a route received from site S; these may be called the "Export Targets". And there is a set of Route Targets that a PE router uses to determine whether a route received from another PE router could be placed in the VRF associated with site S; these may be called the "Import Targets". The two sets are distinct, and need not be the same. Note that a particular VPN-IPv4 route is only eligible for installation in a particular VRF if there is some Route Target that is both one of the route's Route Targets and one of the VRF's Import Targets.

The function performed by the Route Target attribute is similar to that performed by the BGP Communities attribute. However, the format of the latter is inadequate for present purposes, since it allows only a 2-byte numbering space. It is desirable to structure the format, similar to what we have described for RDs (see Section 4.2), so that a type field defines the length of an administrator field, and the remainder of the attribute is a number from the specified administrator's numbering space. This can be done using BGP Extended Communities. The Route Targets discussed herein are encoded as BGP Extended Community Route Targets [BGP-EXTCOMM]. They are structured similarly to the RDs.

When a BGP speaker has received more than one route to the same VPN-IPv4 prefix, the BGP rules for route preference are used to choose which VPN-IPv4 route is installed by BGP.

Note that a route can only have one RD, but it can have multiple Route Targets. In BGP, scalability is improved if one has a single route with multiple attributes, as opposed to multiple routes. One

could eliminate the Route Target attribute by creating more routes (i.e., using more RDs), but the scaling properties would be less favorable.

How does a PE determine which Route Target attributes to associate with a given route? There are a number of different possible ways. The PE might be configured to associate all routes that lead to a specified site with a specified Route Target. Or the PE might be configured to associate certain routes leading to a specified site with one Route Target, and certain with another.

If the PE and the CE are themselves BGP peers (see Section 7), then the SP may allow the customer, within limits, to specify how its routes are to be distributed. The SP and the customer would need to agree in advance on the set of RTs that are allowed to be attached to the customer's VPN routes. The CE could then attach one or more of those RTs to each IP route that it distributes to the PE. This gives the customer the freedom to specify in real time, within agreed-upon limits, its route distribution policies. If the CE is allowed to attach RTs to its routes, the PE MUST filter out all routes that contain RTs that the customer is not allowed to use. If the CE is not allowed to attach RTs to its routes, but does so anyway, the PE MUST remove the RT before converting the customer's route to a VPN-IPv4 route.

4.3.2. Route Distribution Among PEs by BGP

If two sites of a VPN attach to PEs that are in the same Autonomous System, the PEs can distribute VPN-IPv4 routes to each other by means of an IBGP connection between them. (The term "IBGP" refers to the set of protocols and procedures used when there is a BGP connection between two BGP speakers in the same Autonomous System. This is distinguished from "EBGP", the set of procedures used between two BGP speakers in different Autonomous Systems.) Alternatively, each can have an IBGP connection to a route reflector [BGP-RR].

When a PE router distributes a VPN-IPv4 route via BGP, it uses its own address as the "BGP next hop". This address is encoded as a VPN-IPv4 address with an RD of 0. ([BGP-MP] requires that the next hop address be in the same address family as the Network Layer Reachability Information (NLRI).) It also assigns and distributes an MPLS label. (Essentially, PE routers distribute not VPN-IPv4 routes, but Labeled VPN-IPv4 routes. Cf. [MPLS-BGP].) When the PE processes a received packet that has this label at the top of the stack, the PE will pop the stack, and process the packet appropriately.

The PE may distribute the exact set of routes that appears in the VRF, or it may perform summarization and distribute aggregates of those routes, or it may do some of one and some of the other.

Suppose that a PE has assigned label L to route R, and has distributed this label mapping via BGP. If R is an aggregate of a set of routes in the VRF, the PE will know that packets from the backbone that arrive with this label must have their destination addresses looked up in a VRF. When the PE looks up the label in its Label Information Base, it learns which VRF must be used. On the other hand, if R is not an aggregate, then when the PE looks up the label, it learns the egress attachment circuit, as well as the encapsulation header for the packet. In this case, no lookup in the VRF is done.

We would expect that the most common case would be the case where the route is NOT an aggregate. The case where it is an aggregate can be very useful though if the VRF contains a large number of host routes (e.g., as in dial-in), or if the VRF has an associated Local Area Network (LAN) interface (where there is a different outgoing layer 2 header for each system on the LAN, but a route is not distributed for each such system).

Whether or not each route has a distinct label is an implementation matter. There are a number of possible algorithms one could use to determine whether two routes get assigned the same label:

- One may choose to have a single label for an entire VRF, so that a single label is shared by all the routes from that VRF. Then when the egress PE receives a packet with that label, it must look up the packet's IP destination address in that VRF (the packet's "egress VRF"), in order to determine the packet's egress attachment circuit and the corresponding data link encapsulation.

- One may choose to have a single label for each attachment circuit, so that a single label is shared by all the routes with the same "outgoing attachment circuit". This enables one to avoid doing a lookup in the egress VRF, though some sort of lookup may need to be done in order to determine the data link encapsulation, e.g., an Address Resolution Protocol (ARP) lookup.

- One may choose to have a distinct label for each route. Then if a route is potentially reachable over more than one attachment circuit, the PE/CE routing can switch the preferred path for a route from one attachment circuit to another, without there being any need to distribute new a label for that route.

There may be other possible algorithms as well. The choice of algorithm is entirely at the discretion of the egress PE, and is otherwise transparent.

In using BGP-distributed MPLS labels in this manner, we presuppose that an MPLS packet carrying such a label can be tunneled from the router that installs the corresponding BGP-distributed route to the router that is the BGP next hop of that route. This requires either that a label switched path exist between those two routers or else that some other tunneling technology (e.g., [MPLS-in-IP-GRE]) can be used between them.

This tunnel may follow a "best effort" route, or it may follow a traffic-engineered route. Between a given pair of routers, there may be one such tunnel, or there may be several, perhaps with different Quality of Service (QoS) characteristics. All that matters for the VPN architecture is that some such tunnel exists. To ensure interoperability among systems that implement this VPN architecture using MPLS label switched paths as the tunneling technology, all such systems MUST support Label Distribution Protocol (LDP) [MPLS-LDP]. In particular, Downstream Unsolicited mode MUST be supported on interfaces that are neither Label Controlled ATM (LC-ATM) [MPLS-ATM] nor Label Controlled Frame Relay (LC-FR) [MPLS-FR] interfaces, and Downstream on Demand mode MUST be supported on LC-ATM interfaces and LC-FR interfaces.

If the tunnel follows a best-effort route, then the PE finds the route to the remote endpoint by looking up its IP address in the default forwarding table.

A PE router, UNLESS it is a route reflector (see Section 4.3.3) or an Autonomous System Border Router (ASBR) for an inter-provider VPN (see Section 10), should not install a VPN-IPv4 route unless it has at least one VRF with an Import Target identical to one of the route's Route Target attributes. Inbound filtering should be used to cause such routes to be discarded. If a new Import Target is later added to one of the PE's VRFs (a "VPN Join" operation), it must then acquire the routes it may previously have discarded. This can be done using the refresh mechanism described in [BGP-RFSH]. The outbound route filtering mechanism of [BGP-ORF] can also be used to advantage to make the filtering more dynamic.

Similarly, if a particular Import Target is no longer present in any of a PE's VRFs (as a result of one or more "VPN Prune" operations), the PE may discard all routes that, as a result, no longer have any of the PE's VRF's Import Targets as one of their Route Target attributes.

A router that is not attached to any VPN and that is not a Route Reflector (i.e., a P router) never installs any VPN-IPv4 routes at all.

Note that VPN Join and Prune operations are non-disruptive and do not require any BGP connections to be brought down, as long as the refresh mechanism of [BGP-RFSH] is used.

As a result of these distribution rules, no one PE ever needs to maintain all routes for all VPNs; this is an important scalability consideration.

4.3.3. Use of Route Reflectors

Rather than having a complete IBGP mesh among the PEs, it is advantageous to make use of BGP Route Reflectors [BGP-RR] to improve scalability. All the usual techniques for using route reflectors to improve scalability (e.g., route reflector hierarchies) are available.

Route reflectors are the only systems that need to have routing information for VPNs to which they are not directly attached. However, there is no need to have any one route reflector know all the VPN-IPv4 routes for all the VPNs supported by the backbone.

We outline below two different ways to partition the set of VPN-IPv4 routes among a set of route reflectors.

1. Each route reflector is preconfigured with a list of Route Targets. For redundancy, more than one route reflector may be preconfigured with the same list. A route reflector uses the preconfigured list of Route Targets to construct its inbound route filtering. The route reflector may use the techniques of [BGP-ORF] to install on each of its peers (regardless of whether the peer is another route reflector or a PE) the set of Outbound Route Filters (ORFs) that contains the list of its preconfigured Route Targets. Note that route reflectors should accept ORFs from other route reflectors, which means that route reflectors should advertise the ORF capability to other route reflectors.

   A service provider may modify the list of preconfigured Route Targets on a route reflector. When this is done, the route reflector modifies the ORFs it installs on all of its IBGP peers. To reduce the frequency of configuration changes on route reflectors, each route reflector may be preconfigured with a block of Route Targets. This way, when a new Route Target is needed for a new VPN, there is already one or more

route reflectors that are (pre)configured with this Route Target.

Unless a given PE is a client of all route reflectors, when a new VPN is added to the PE ("VPN Join"), it will need to become a client of the route reflector(s) that maintain routes for that VPN. Likewise, deleting an existing VPN from the PE ("VPN Prune") may result in a situation where the PE no longer needs to be a client of some route reflector(s). In either case, the Join or Prune operation is non-disruptive (as long as [BGP-RFSH] is used, and never requires a BGP connection to be brought down, only to be brought right back up.

(By "adding a new VPN to a PE", we really mean adding a new import Route Target to one of its VRFs, or adding a new VRF with an import Route Target not had by any of the PE's other VRFs.)

2. Another method is to have each PE be a client of some subset of the route reflectors. A route reflector is not preconfigured with the list of Route Targets, and does not perform inbound route filtering of routes received from its clients (PEs); rather, it accepts all the routes received from all of its clients (PEs). The route reflector keeps track of the set of the Route Targets carried by all the routes it receives. When the route reflector receives from its client a route with a Route Target that is not in this set, this Route Target is immediately added to the set. On the other hand, when the route reflector no longer has any routes with a particular Route Target that is in the set, the route reflector should delay (by a few hours) the deletion of this Route Target from the set.

   The route reflector uses this set to form the inbound route filters that it applies to routes received from other route reflectors. The route reflector may also use ORFs to install the appropriate outbound route filtering on other route reflectors. Just like with the first approach, a route reflector should accept ORFs from other route reflectors. To accomplish this, a route reflector advertises ORF capability to other route reflectors.

   When the route reflector changes the set, it should immediately change its inbound route filtering. In addition, if the route reflector uses ORFs, then the ORFs have to be immediately changed to reflect the changes in the set. If the route reflector doesn't use ORFs, and a new Route Target is added to

the set, the route reflector, after changing its inbound route filtering, must issue BGP Refresh to other route reflectors.

The delay of "a few hours" mentioned above allows a route reflector to hold onto routes with a given RT, even after it loses the last of its clients that are interested in such routes. This protects against the need to reacquire all such routes if the clients' "disappearance" is only temporary.

With this procedure, VPN Join and Prune operations are also non-disruptive.

Note that this technique will not work properly if some client PE has a VRF with an import Route Target that is not one of its export Route Targets.

In these procedures, a PE router which attaches to a particular VPN "auto-discovers" the other PEs that attach to the same VPN. When a new PE router is added, or when an existing PE router attaches to a new VPN, no reconfiguration of other PE routers is needed.

Just as there is no one PE router that needs to know all the VPN-IPv4 routes supported over the backbone, these distribution rules ensure that there is no one Route Reflector (RR) that needs to know all the VPN-IPv4 routes supported over the backbone. As a result, the total number of such routes that can be supported over the backbone is not bounded by the capacity of any single device, and therefore can increase virtually without bound.

4.3.4. How VPN-IPv4 NLRI Is Carried in BGP

The BGP Multiprotocol Extensions [BGP-MP] are used to encode the NLRI. If the Address Family Identifier (AFI) field is set to 1, and the Subsequent Address Family Identifier (SAFI) field is set to 128, the NLRI is an MPLS-labeled VPN-IPv4 address. AFI 1 is used since the network layer protocol associated with the NLRI is still IP. Note that this VPN architecture does not require the capability to distribute unlabeled VPN-IPv4 addresses.

In order for two BGP speakers to exchange labeled VPN-IPv4 NLRI, they must use BGP Capabilities Advertisement to ensure that they both are capable of properly processing such NLRI. This is done as specified in [BGP-MP], by using capability code 1 (multiprotocol BGP), with an AFI of 1 and an SAFI of 128.

The labeled VPN-IPv4 NLRI itself is encoded as specified in [MPLS-BGP], where the prefix consists of an 8-byte RD followed by an IPv4 prefix.

4.3.5. Building VPNs Using Route Targets

By setting up the Import Targets and Export Targets properly, one can construct different kinds of VPNs.

Suppose it is desired to create a fully meshed closed user group, i.e., a set of sites where each can send traffic directly to the other, but traffic cannot be sent to or received from other sites. Then each site is associated with a VRF, a single Route Target attribute is chosen, that Route Target is assigned to each VRF as both the Import Target and the Export Target, and that Route Target is not assigned to any other VRFs as either the Import Target or the Export Target.

Alternatively, suppose one desired, for whatever reason, to create a "hub and spoke" kind of VPN. This could be done by the use of two Route Target values, one meaning "Hub" and one meaning "Spoke". At the VRFs attached to the hub sites, "Hub" is the Export Target and

"Spoke" is the Import Target. At the VRFs attached to the spoke site, "Hub" is the Import Target and "Spoke" is the Export Target.

Thus, the methods for controlling the distribution of routing information among various sets of sites are very flexible, which in turn provides great flexibility in constructing VPNs.

4.3.6. Route Distribution Among VRFs in a Single PE

It is possible to distribute routes from one VRF to another, even if both VRFs are in the same PE, even though in this case one cannot say that the route has been distributed by BGP. Nevertheless, the decision to distribute a particular route from one VRF to another within a single PE is the same decision that would be made if the VRFs were on different PEs. That is, it depends on the Route Target attribute that is assigned to the route (or would be assigned if the route were distributed by BGP), and the import target of the second VRF.

5. Forwarding

If the intermediate routers in the backbone do not have any information about the routes to the VPNs, how are packets forwarded from one VPN site to another?

When a PE receives an IP packet from a CE device, it chooses a particular VRF in which to look up the packet’s destination address. This choice is based on the packet’s ingress attachment circuit.

Assume that a match is found. As a result we learn the packet's "next hop".

If the packet's next hop is reached directly over a VRF attachment circuit from this PE (i.e., the packet's egress attachment circuit is on the same PE as its ingress attachment circuit), then the packet is sent on the egress attachment circuit, and no MPLS labels are pushed onto the packet's label stack.

If the ingress and egress attachment circuits are on the same PE, but are associated with different VRFs, and if the route that best matches the destination address in the ingress attachment circuit's VRF is an aggregate of several routes in the egress attachment circuit's VRF, it may be necessary to look up the packet's destination address in the egress VRF as well.

If the packet's next hop is NOT reached through a VRF attachment circuit, then the packet must travel at least one hop through the backbone. The packet thus has a "BGP Next Hop", and the BGP Next Hop will have assigned an MPLS label for the route that best matches the packet's destination address. Call this label the "VPN route label". The IP packet is turned into an MPLS packet with the VPN route label as the sole label on the label stack.

The packet must then be tunneled to the BGP Next Hop.

If the backbone supports MPLS, this is done as follows:

- The PE routers (and any Autonomous System border routers) that redistribute VPN-IPv4 addresses need to insert /32 address prefixes for themselves into the IGP routing tables of the backbone. This enables MPLS, at each node in the backbone network, to assign a label corresponding to the route to each PE router. To ensure interoperability among different implementations, it is required to support LDP for setting up the label switched paths across the backbone. However, other methods of setting up these label switched paths are also possible. (Some of these other methods may not require the presence of the /32 address prefixes in the IGP.)

- If there are any traffic engineering tunnels to the BGP next hop, and if one or more of those is available for use by the packet in question, one of these tunnels is chosen. This tunnel will be associated with an MPLS label, the "tunnel label". The tunnel label gets pushed on the MPLS label stack, and the packet is forwarded to the tunnel's next hop.

- Otherwise,

  * The packet will have an "IGP Next Hop", which is the next hop along the IGP route to the BGP Next Hop.

  * If the BGP Next Hop and the IGP Next Hop are the same, and if penultimate hop popping is used, the packet is then sent to the IGP Next Hop, carrying only the VPN route label.

  * Otherwise, the IGP Next Hop will have assigned a label for the route that best matches the address of the BGP Next Hop. Call this the "tunnel label". The tunnel label gets pushed on as the packet's top label. The packet is then forwarded to the IGP Next Hop.

- MPLS will then carry the packet across the backbone to the BGP Next Hop, where the VPN label will be examined.

If the backbone does not support MPLS, the MPLS packet carrying only the VPN route label may be tunneled to the BGP Next Hop using the techniques of [MPLS-in-IP-GRE]. When the packet emerges from the tunnel, it will be at the BGP Next Hop, where the VPN route label will be examined.

At the BGP Next Hop, the treatment of the packet depends on the VPN route label (see Section 4.3.2). In many cases, the PE will be able to determine, from this label, the attachment circuit over which the packet should be transmitted (to a CE device), as well as the proper data link layer header for that interface. In other cases, the PE may only be able to determine that the packet's destination address needs to be looked up in a particular VRF before being forwarded to a CE device. There are also intermediate cases in which the VPN route label may determine the packet's egress attachment circuit, but a lookup (e.g., ARP) still needs to be done in order to determine the packet's data link header on that attachment circuit.

Information in the MPLS header itself, and/or information associated with the label, may also be used to provide QoS on the interface to the CE.

In any event, if the packet was an unlabeled IP packet when it arrived at its ingress PE, it will again be an unlabeled packet when it leaves its egress PE.

The fact that packets with VPN route labels are tunneled through the backbone is what makes it possible to keep all the VPN routes out of the P routers. This is crucial to ensuring the scalability of the

scheme. The backbone does not even need to have routes to the CEs, only to the PEs.

With respect to the tunnels, it is worth noting that this specification:

- DOES NOT require that the tunnels be point-to-point; multipoint-to-point can be used;

- DOES NOT require that there be any explicit setup of the tunnels, either via signaling or via manual configuration;

- DOES NOT require that there be any tunnel-specific signaling;

- DOES NOT require that there be any tunnel-specific state in the P or PE routers, beyond what is necessary to maintain the routing information and (if used) the MPLS label information.

Of course, this specification is compatible with the use of point-to-point tunnels that must be explicitly configured and/or signaled, and in some situations there may be reasons for using such tunnels.

The considerations that are relevant to choosing a particular tunneling technology are outside the scope of this specification.

6. Maintaining Proper Isolation of VPNs

To maintain proper isolation of one VPN from another, it is important that no router in the backbone accept a tunneled packet from outside the backbone, unless it is sure that both endpoints of that tunnel are outside the backbone.

If MPLS is being used as the tunneling technology, this means that a router in the backbone MUST NOT accept a labeled packet from any adjacent non-backbone device unless the following two conditions hold:

1. the label at the top of the label stack was actually distributed by that backbone router to that non-backbone device, and

2. the backbone router can determine that use of that label will cause the packet to leave the backbone before any labels lower in the stack will be inspected, and before the IP header will be inspected.

The first condition ensure that any labeled packets received from non-backbone routers have a legitimate and properly assigned label at

the top of the label stack. The second condition ensures that the backbone routers will never look below that top label. Of course, the simplest way to meet these two conditions is just to have the backbone devices refuse to accept labeled packets from non-backbone devices.

If MPLS is not being used as the tunneling technology, then filtering must be done to ensure that an MPLS-in-IP or MPLS-in-GRE packet can be accepted into the backbone only if the packet's IP destination address will cause it to be sent outside the backbone.

7. How PEs Learn Routes from CEs

The PE routers that attach to a particular VPN need to know, for each attachment circuit leading to that VPN, which of the VPN's addresses should be reached over that attachment circuit.

The PE translates these addresses into VPN-IPv4 addresses, using a configured RD. The PE then treats these VPN-IPv4 routes as input to BGP. Routes from a VPN site are NOT leaked into the backbone's IGP.

Exactly which PE/CE route distribution techniques are possible depends on whether or not a particular CE is in a "transit VPN". A "transit VPN" is one that contains a router that receives routes from a "third party" (i.e., from a router that is not in the VPN, but is not a PE router) and that redistributes those routes to a PE router. A VPN that is not a transit VPN is a "stub VPN". The vast majority of VPNs, including just about all corporate enterprise networks, would be expected to be "stubs" in this sense.

The possible PE/CE distribution techniques are:

1. Static routing (i.e., configuration) may be used. (This is likely to be useful only in stub VPNs.)

2. PE and CE routers may be Routing Information Protocol (RIP) [RIP] peers, and the CE may use RIP to tell the PE router the set of address prefixes that are reachable at the CE router's site. When RIP is configured in the CE, care must be taken to ensure that address prefixes from other sites (i.e., address prefixes learned by the CE router from the PE router) are never advertised to the PE. More precisely: if a PE router, say, PE1, receives a VPN-IPv4 route R1, and as a result distributes an IPv4 route R2 to a CE, then R2 must not be distributed back from that CE's site to a PE router, say, PE2, (where PE1 and PE2 may be the same router or different routers), unless PE2 maps R2 to a VPN-IPv4 route that is different than (i.e., contains a different RD than) R1.

3. The PE and CE routers may be OSPF peers. A PE router that is an OSPF peer of a CE router appears, to the CE router, to be an area 0 router. If a PE router is an OSPF peer of CE routers that are in distinct VPNs, the PE must of course be running multiple instances of OSPF.

   IPv4 routes that the PE learns from the CE via OSPF are redistributed into BGP as VPN-IPv4 routes. Extended Community attributes are used to carry, along with the route, all the information needed to enable the route to be distributed to other CE routers in the VPN in the proper type of OSPF Link State Advertisement (LSA). OSPF route tagging is used to ensure that routes received from the MPLS/BGP backbone are not sent back into the backbone.

   Specification of the complete set of procedures for the use of OSPF between PE and CE can be found in [VPN-OSPF] and [OSPF-2547-DNBIT].

4. The PE and CE routers may be BGP peers, and the CE router may use BGP (in particular, EBGP to tell the PE router the set of address prefixes that are at the CE router’s site. (This technique can be used in stub VPNs or transit VPNs.)

   This technique has a number of advantages over the others:

   a) Unlike the IGP alternatives, this does not require the PE to run multiple routing algorithm instances in order to talk to multiple CEs.

   b) BGP is explicitly designed for just this function: passing routing information between systems run by different administrations.

   c) If the site contains "BGP backdoors", i.e., routers with BGP connections to routers other than PE routers, this procedure will work correctly in all circumstances. The other procedures may or may not work, depending on the precise circumstances.

   d) Use of BGP makes it easy for the CE to pass attributes of the routes to the PE. A complete specification of the set of attributes and their use is outside the scope of this document. However, some examples of the way this may be used are the following:

- The CE may suggest a particular Route Target for each route, from among the Route Targets that the PE is authorized to attach to the route. The PE would then attach only the suggested Route Target, rather than the full set. This gives the CE administrator some dynamic control of the distribution of routes from the CE.

- Additional types of Extended Community attributes may be defined, where the intention is to have those attributes passed transparently (i.e., without being changed by the PE routers) from CE to CE. This would allow CE administrators to implement additional route filtering, beyond that which is done by the PEs. This additional filtering would not require coordination with the SP.

On the other hand, using BGP may be something new for the CE administrators.

If a site is not in a transit VPN, note that it need not have a unique Autonomous System Number (ASN). Every CE whose site is not in a transit VPN can use the same ASN. This can be chosen from the private ASN space, and it will be stripped out by the PE. Routing loops are prevented by use of the Site of Origin attribute (see below).

What if a set of sites constitutes a transit VPN? This will generally be the case only if the VPN is itself an Internet Service Provider's (ISP's) network, where the ISP is itself buying backbone services from another SP. The latter SP may be called a "carrier's carrier". In this case, the best way to provide the VPN is to have the CE routers support MPLS, and to use the technique described in Section 9.

When we do not need to distinguish among the different ways in which a PE can be informed of the address prefixes that exist at a given site, we will simply say that the PE has "learned" the routes from that site. This includes the case where the PE has been manually configured with the routes.

Before a PE can redistribute a VPN-IPv4 route learned from a site, it must assign a Route Target attribute (see Section 4.3.1) to the route, and it may assign a Site of Origin attribute to the route.

The Site of Origin attribute, if used, is encoded as a Route Origin Extended Community [BGP-EXTCOMM]. The purpose of this attribute is to uniquely identify the set of routes learned from a particular

site. This attribute is needed in some cases to ensure that a route learned from a particular site via a particular PE/CE connection is not distributed back to the site through a different PE/CE connection. It is particularly useful if BGP is being used as the PE/CE protocol, but different sites have not been assigned distinct ASNs.

8. How CEs Learn Routes from PEs

In this section, we assume that the CE device is a router.

If the PE places a particular route in the VRF it uses to route packets received from a particular CE, then in general, the PE may distribute that route to the CE. Of course, the PE may distribute that route to the CE only if this is permitted by the rules of the PE/CE protocol. (For example, if a particular PE/CE protocol has "split horizon", certain routes in the VRF cannot be redistributed back to the CE.) We add one more restriction on the distribution of routes from PE to CE: if a route's Site of Origin attribute identifies a particular site, that route must never be redistributed to any CE at that site.

In most cases, however, it will be sufficient for the PE to simply distribute the default route to the CE. (In some cases, it may even be sufficient for the CE to be configured with a default route pointing to the PE.) This will generally work at any site that does not itself need to distribute the default route to other sites. (E.g., if one site in a corporate VPN has the corporation's access to the Internet, that site might need to have default distributed to the other site, but one could not distribute default to that site itself.)

Whatever procedure is used to distribute routes from CE to PE will also be used to distribute routes from PE to CE.

9. Carriers' Carriers

Sometimes a VPN may actually be the network of an ISP, with its own peering and routing policies. Sometimes a VPN may be the network of an SP that is offering VPN services in turn to its own customers. VPNs like these can also obtain backbone service from another SP, the "carrier's carrier", using essentially the same methods described in this document. However, it is necessary in these cases that the CE routers support MPLS. In particular:

- The CE routers should distribute to the PE routers ONLY those routes that are internal to the VPN. This allows the VPN to be handled as a stub VPN.

- The CE routers should support MPLS, in that they should be able to receive labels from the PE routers, and send labeled packets to the PE routers. They do not need to distribute labels of their own, though.

- The PE routers should distribute, to the CE routers, labels for the routes they distribute to the CE routers.

  The PE must not distribute the same label to two different CEs unless one of the following conditions holds:

  * The two CEs are associated with exactly the same set of VRFs;

  * The PE maintains a different Incoming Label Map ([MPLS-ARCH]) for each CE.

  Further, when the PE receives a labeled packet from a CE, it must verify that the top label is one that was distributed to that CE.

- Routers at the different sites should establish BGP connections among themselves for the purpose of exchanging external routes (i.e., routes that lead outside of the VPN).

- All the external routes must be known to the CE routers.

Then when a CE router looks up a packet's destination address, the routing lookup will resolve to an internal address, usually the address of the packet's BGP next hop. The CE labels the packet appropriately and sends the packet to the PE. The PE, rather than looking up the packet's IP destination address in a VRF, uses the packet's top MPLS label to select the BGP next hop. As a result, if the BGP next hop is more than one hop away, the top label will be replaced by two labels, a tunnel label and a VPN route label. If the BGP next hop is one hop away, the top label may be replaced by just the VPN route label. If the ingress PE is also the egress PE, the top label will just be popped. When the packet is sent from its egress PE to a CE, the packet will have one fewer MPLS labels than it had when it was first received by its ingress PE.

In the above procedure, the CE routers are the only routers in the VPN that need to support MPLS. If, on the other hand, all the routers at a particular VPN site support MPLS, then it is no longer required that the CE routers know all the external routes. All that is required is that the external routes be known to whatever routers are responsible for putting the label stack on a hitherto unlabeled packet and that there be label switched path that leads from those routers to their BGP peers at other sites. In this case, for each

internal route that a CE router distributes to a PE router, it must also distribute a label.

10. Multi-AS Backbones

What if two sites of a VPN are connected to different Autonomous Systems (e.g., because the sites are connected to different SPs)? The PE routers attached to that VPN will then not be able to maintain IBGP connections with each other, or with a common route reflector. Rather, there needs to be some way to use EBGP to distribute VPN-IPv4 addresses.

There are a number of different ways of handling this case, which we present in order of increasing scalability.

a) VRF-to-VRF connections at the AS (Autonomous System) border routers.

   In this procedure, a PE router in one AS attaches directly to a PE router in another. The two PE routers will be attached by multiple sub-interfaces, at least one for each of the VPNs whose routes need to be passed from AS to AS. Each PE will treat the other as if it were a CE router. That is, the PEs associate each such sub-interface with a VRF, and use EBGP to distribute unlabeled IPv4 addresses to each other.

   This is a procedure that "just works", and that does not require MPLS at the border between ASes. However, it does not scale as well as the other procedures discussed below.

b) EBGP redistribution of labeled VPN-IPv4 routes from AS to neighboring AS.

   In this procedure, the PE routers use IBGP to redistribute labeled VPN-IPv4 routes either to an Autonomous System Border Router (ASBR), or to a route reflector of which an ASBR is a client. The ASBR then uses EBGP to redistribute those labeled VPN-IPv4 routes to an ASBR in another AS, which in turn distributes them to the PE routers in that AS, or perhaps to another ASBR which in turn distributes them, and so on.

   When using this procedure, VPN-IPv4 routes should only be accepted on EBGP connections at private peering points, as part of a trusted arrangement between SPs. VPN-IPv4 routes should neither be distributed to nor accepted from the public Internet, or from any BGP peers that are not trusted. An ASBR should never accept a labeled packet from an EBGP peer unless it has actually distributed the top label to that peer.

If there are many VPNs having sites attached to different Autonomous Systems, there does not need to be a single ASBR between those two ASes that holds all the routes for all the VPNs; there can be multiple ASBRs, each of which holds only the routes for a particular subset of the VPNs.

This procedure requires that there be a label switched path leading from a packet's ingress PE to its egress PE. Hence the appropriate trust relationships must exist between and among the set of ASes along the path. Also, there must be agreement among the set of SPs as to which border routers need to receive routes with which Route Targets.

c) Multi-hop EBGP redistribution of labeled VPN-IPv4 routes between source and destination ASes, with EBGP redistribution of labeled IPv4 routes from AS to neighboring AS.

In this procedure, VPN-IPv4 routes are neither maintained nor distributed by the ASBRs. An ASBR must maintain labeled IPv4 /32 routes to the PE routers within its AS. It uses EBGP to distribute these routes to other ASes. ASBRs in any transit ASes will also have to use EBGP to pass along the labeled /32 routes. This results in the creation of a label switched path from the ingress PE router to the egress PE router. Now PE routers in different ASes can establish multi-hop EBGP connections to each other, and can exchange VPN-IPv4 routes over those connections.

If the /32 routes for the PE routers are made known to the P routers of each AS, everything works normally. If the /32 routes for the PE routers are NOT made known to the P routers (other than the ASBRs), then this procedure requires a packet's ingress PE to put a three-label stack on it. The bottom label is assigned by the egress PE, corresponding to the packet's destination address in a particular VRF. The middle label is assigned by the ASBR, corresponding to the /32 route to the egress PE. The top label is assigned by the ingress PE's IGP Next Hop, corresponding to the /32 route to the ASBR.

To improve scalability, one can have the multi-hop EBGP connections exist only between a route reflector in one AS and a route reflector in another. (However, when the route reflectors distribute routes over this connection, they do not modify the BGP next hop attribute of the routes.) The actual PE routers would then only have IBGP connections to the route reflectors in their own AS.

This procedure is very similar to the "carrier's carrier" procedures described in Section 9. Like the previous procedure, it requires that there be a label switched path leading from a packet's ingress PE to its egress PE.

11. Accessing the Internet from a VPN

Many VPN sites will need to be able to access the public Internet, as well as to access other VPN sites. The following describes some of the alternative ways of doing this.

1. In some VPNs, one or more of the sites will obtain Internet access by means of an "Internet gateway" (perhaps a firewall) attached to a non-VRF interface to an ISP. The ISP may or may not be the same organization as the SP that is providing the VPN service. Traffic to/from the Internet gateway would then be routed according to the PE router's default forwarding table.

   In this case, the sites that have Internet access may be distributing a default route to their PEs, which in turn redistribute it to other PEs and hence into other sites of the VPN. This provides Internet access for all of the VPN's sites.

   In order to properly handle traffic from the Internet, the ISP must distribute, to the Internet, routes leading to addresses that are within the VPN. This is completely independent of any of the route distribution procedures described in this document. The internal structure of the VPN will in general not be visible from the Internet; such routes would simply lead to the non-VRF interface that attaches to the VPN's Internet gateway.

   In this model, there is no exchange of routes between a PE router's default forwarding table and any of its VRFs. VPN route distribution procedures and Internet route distribution procedures are completely independent.

   Note that although some sites of the VPN use a VRF interface to communicate with the Internet, ultimately all packets to/from the Internet traverse a non-VRF interface before leaving/entering the VPN, so we refer to this as "non-VRF Internet access".

   Note that the PE router to which the non-VRF interface attaches does not necessarily need to maintain all the Internet routes in its default forwarding table. The default forwarding table could have as few as one route, "default", which leads to

another router (probably an adjacent one) that has the Internet routes. A variation of this scheme is to tunnel packets received over the non-VRF interface from the PE router to another router, where this other router maintains the full set of Internet routes.

2. Some VPNs may obtain Internet access via a VRF interface ("VRF Internet access"). If a packet is received by a PE over a VRF interface, and if the packet's destination address does not match any route in the VRF, then it may be matched against the PE's default forwarding table. If a match is made there, the packet can be forwarded natively through the backbone to the Internet, instead of being forwarded by MPLS.

   In order for traffic to flow natively in the opposite direction (from Internet to VRF interface), some of the routes from the VRF must be exported to the Internet forwarding table. Needless to say, any such routes must correspond to globally unique addresses.

   In this scheme, the default forwarding table might have the full set of Internet routes, or it might have as little as a single default route leading to another router that does have the full set of Internet routes in its default forwarding table.

3. Suppose the PE has the capability to store "non-VPN routes" in a VRF. If a packet's destination address matches a "non-VPN route", then the packet is transmitted natively, rather than being transmitted via MPLS. If the VRF contains a non-VPN default route, all packets for the public Internet will match it, and be forwarded natively to the default route's next hop. At that next hop, the packets' destination addresses will be looked up in the default forwarding table, and may match more specific routes.

   This technique would only be available if none of the CE routers is distributing a default route.

4. It is also possible to obtain Internet access via a VRF interface by having the VRF contain the Internet routes. Compared with model 2, this eliminates the second lookup, but it has the disadvantage of requiring the Internet routes to be replicated in each such VRF.

   If this technique is used, the SP may want to make its interface to the Internet be a VRF interface, and to use the

techniques of Section 4 to distribute Internet routes, as VPN-IPv4 routes, to other VRFs.

It should be clearly understood that by default, there is no exchange of routes between a VRF and the default forwarding table. This is done ONLY upon agreement between a customer and an SP, and only if it suits the customer's policies.

12. Management VPNs

This specification does not require that the sub-interface connecting a PE router and a CE router be a "numbered" interface. If it is a numbered interface, this specification allows the addresses assigned to the interface to come from either the address space of the VPN or the address space of the SP.

If a CE router is being managed by the Service Provider, then the Service Provider will likely have a network management system that needs to be able to communicate with the CE router. In this case, the addresses assigned to the sub-interface connecting the CE and PE routers should come from the SP's address space, and should be unique within that space. The network management system should itself connect to a PE router (more precisely, be at a site that connects to a PE router) via a VRF interface. The address of the network management system will be exported to all VRFs that are associated with interfaces to CE routers that are managed by the SP. The addresses of the CE routers will be exported to the VRF associated with the network management system, but not to any other VRFs.

This allows communication between the CE and network management system, but does not allow any undesired communication to or among the CE routers.

One way to ensure that the proper route import/exports are done is to use two Route Targets; call them T1 and T2. If a particular VRF interface attaches to a CE router that is managed by the SP, then that VRF is configured to:

- import routes that have T1 attached to them, and
- attach T2 to addresses assigned to each end of its VRF interfaces.

If a particular VRF interface attaches to the SP's network management system, then that VRF is configured to attach T1 to the address of that system, and to import routes that have T2 attached to them.

13. Security Considerations

13.1. Data Plane

By security in the "data plane", we mean protection against the following possibilities:

- Packets from within a VPN travel to a site outside the VPN, other than in a manner consistent with the policies of the VPN.

- Packets from outside a VPN enter one of the VPN's sites, other than in a manner consistent with the policies of the VPN.

Under the following conditions:

1. a backbone router does not accept labeled packets over a particular data link, unless it is known that that data link attaches only to trusted systems, or unless it is known that such packets will leave the backbone before the IP header or any labels lower in the stack will be inspected, and

2. labeled VPN-IPv4 routes are not accepted from untrusted or unreliable routing peers,

3. no successful attacks have been mounted on the control plane,

the data plane security provided by this architecture is virtually identical to that provided to VPNs by Frame Relay or ATM backbones. If the devices under the control of the SP are properly configured, data will not enter or leave a VPN unless authorized to do so.

Condition 1 above can be stated more precisely. One should discard a labeled packet received from a particular neighbor unless one of the following two conditions holds:

- the packet's top label has a label value that the receiving system has distributed to that neighbor, or

- the packet's top label has a label value that the receiving system has distributed to a system beyond that neighbor (i.e., when it is known that the path from the system to which the label was distributed to the receiving system may be via that neighbor).

Condition 2 above is of most interest in the case of inter-provider VPNs (see Section 10). For inter-provider VPNs constructed according to scheme b) of Section 10, condition 2 is easily checked. (The issue of security when scheme (c) of Section 10 is used is for further study.)

It is worth noting that the use of MPLS makes it much simpler to provide data plane security than might be possible if one attempted to use some form of IP tunneling in place of the MPLS outer label. It is a simple matter to have one's border routers refuse to accept a labeled packet unless the first of the above conditions applies to it. It is rather more difficult to configure a router to refuse to accept an IP packet if that packet is an IP tunneled packet whose destination address is that of a PE router; certainly, this is not impossible to do, but it has both management and performance implications.

MPLS-in-IP and MPLS-in-GRE tunneling are specified in [MPLS-in-IP-GRE]. If it is desired to use such tunnels to carry VPN packets, then the security considerations described in Section 8 of that document must be fully understood. Any implementation of BGP/MPLS IP VPNs that allows VPN packets to be tunneled as described in that document MUST contain an implementation of IPsec that can be used as therein described. If the tunnel is not secured by IPsec, then the technique of IP address filtering at the border routers, described in Section 8.2 of that document, is the only means of ensuring that a packet that exits the tunnel at a particular egress PE was actually placed in the tunnel by the proper tunnel head node (i.e., that the packet does not have a spoofed source address). Since border routers frequently filter only source addresses, packet filtering may not be effective unless the egress PE can check the IP source address of any tunneled packet it receives, and compare it to a list of IP addresses that are valid tunnel head addresses. Any implementation that allows MPLS-in-IP and/or MPLS-in-GRE tunneling to be used without IPsec MUST allow the egress PE to validate in this manner the IP source address of any tunneled packet that it receives.

In the case where a number of CE routers attach to a PE router via a LAN interface, to ensure proper security, one of the following conditions must hold:

1. All the CE routers on the LAN belong to the same VPN, or

2. A trusted and secured LAN switch divides the LAN into multiple VLANs, with each VLAN containing only systems of a single VPN; in this case, the switch will attach the appropriate VLAN tag to any packet before forwarding it to the PE router.

Cryptographic privacy is not provided by this architecture, nor by Frame Relay or ATM VPNs. These architectures are all compatible with the use of cryptography on a CE-CE basis, if that is desired.

The use of cryptography on a PE-PE basis is for further study.

13.2. Control Plane

The data plane security of the previous section depends on the security of the control plane. To ensure security, neither BGP nor LDP connections should be made with untrusted peers. The TCP/IP MD5 authentication option [TCP-MD5] should be used with both these protocols. The routing protocol within the SP's network should also be secured in a similar manner.

13.3. Security of P and PE Devices

If the physical security of these devices is compromised, data plane security may also be compromised.

The usual steps should be taken to ensure that IP traffic from the public Internet cannot be used to modify the configuration of these devices, or to mount Denial of Service attacks on them.

14. Quality of Service

Although not the focus of this paper, Quality of Service is a key component of any VPN service. In MPLS/BGP VPNs, existing L3 QoS capabilities can be applied to labeled packets through the use of the "experimental" bits in the shim header [MPLS-ENCAPS], or, where ATM is used as the backbone, through the use of ATM QoS capabilities. The traffic engineering work discussed in [MPLS-RSVP] is also directly applicable to MPLS/BGP VPNs. Traffic engineering could even be used to establish label switched paths with particular QoS characteristics between particular pairs of sites, if that is desirable. Where an MPLS/BGP VPN spans multiple SPs, the architecture described in [PASTE] may be useful. An SP may apply either intserv (Integrated Services) or diffserv (Differentiated Services) capabilities to a particular VPN, as appropriate.

15. Scalability

We have discussed scalability issues throughout this paper. In this section, we briefly summarize the main characteristics of our model with respect to scalability.

The Service Provider backbone network consists of (a) PE routers, (b) BGP Route Reflectors, (c) P routers (that are neither PE routers nor Route Reflectors), and, in the case of multi-provider VPNs, (d) ASBRs.

P routers do not maintain any VPN routes. In order to properly forward VPN traffic, the P routers need only maintain routes to the PE routers and the ASBRs. The use of two levels of labeling is what makes it possible to keep the VPN routes out of the P routers.

A PE router maintains VPN routes, but only for those VPNs to which it is directly attached.

Route reflectors can be partitioned among VPNs so that each partition carries routes for only a subset of the VPNs supported by the Service Provider. Thus, no single route reflector is required to maintain routes for all VPNs.

For inter-provider VPNs, if the ASBRs maintain and distribute VPN-IPv4 routes, then the ASBRs can be partitioned among VPNs in a similar manner, with the result that no single ASBR is required to maintain routes for all the inter-provider VPNs. If multi-hop EBGP is used, then the ASBRs need not maintain and distribute VPN-IPv4 routes at all.

As a result, no single component within the Service Provider network has to maintain all the routes for all the VPNs. So the total capacity of the network to support increasing numbers of VPNs is not limited by the capacity of any individual component.

16. IANA Considerations

The Internet Assigned Numbers Authority (IANA) has created a new registry for the "Route Distinguisher Type Field" (see Section 4.2). This is a two-byte field. Types 0, 1, and 2 are defined by this document. Additional Route Distinguisher Type Field values with a high-order bit of 0 may be allocated by IANA on a "First Come, First Served" basis [IANA]. Values with a high-order bit of 1 may be allocated by IANA based on "IETF consensus" [IANA].

This document specifies (see Section 4.3.4) the use of the BGP Address Family Identifier (AFI) value 1, along with the BGP Subsequent Address Family Identifier (SAFI) value 128, to represent the address family "VPN-IPv4 Labeled Addresses", which is defined in this document.

The use of AFI value 1 for IP is as currently specified in the IANA registry "Address Family Identifier", so IANA need take no action with respect to it.

The SAFI value 128 was originally specified as "Private Use" in the IANA "Subsequent Address Family Identifier" registry. IANA has changed the SAFI value 128 from "private use" to "MPLS-labeled VPN address".

17. Acknowledgements

The full list of contributors can be found in Section 18.

Significant contributions to this work have also been made by Ravi Chandra, Dan Tappan, and Bob Thomas.

We also wish to thank Shantam Biswas for his review and contributions.

18. Contributors

Tony Bogovic
Telcordia Technologies
445 South Street, Room 1A264B
Morristown, NJ 07960

EMail: tjb@research.telcordia.com

Stephen John Brannon
Swisscom AG
Postfach 1570
CH-8301
Glattzentrum (Zuerich), Switzerland

EMail: stephen.brannon@swisscom.com

Marco Carugi
Nortel Networks S.A.
Parc d'activites de Magny-Les Jeunes Bois CHATEAUFORT
78928 YVELINES Cedex 9 - FRANCE

EMail: marco.carugi@nortelnetworks.com

Christopher J. Chase
AT&T
200 Laurel Ave
Middletown, NJ 07748
USA

EMail: chase@att.com

Ting Wo Chung
Bell Nexxia
181 Bay Street
Suite 350
Toronto, Ontario
M5J2T3

EMail: ting_wo.chung@bellnexxia.com

Eric Dean

Jeremy De Clercq
Alcatel Network Strategy Group
Francis Wellesplein 1
2018 Antwerp, Belgium

EMail: jeremy.de_clercq@alcatel.be

Luyuan Fang
AT&T
IP Backbone Architecture
200 Laurel Ave.
Middletown, NJ 07748

EMail: luyuanfang@att.com

Paul Hitchen
BT
BT Adastral Park
Martlesham Heath,
Ipswich IP5 3RE
UK

EMail: paul.hitchen@bt.com

Manoj Leelanivas
Juniper Networks, Inc.
385 Ravendale Drive
Mountain View, CA 94043 USA

EMail: manoj@juniper.net

Dave Marshall
Worldcom
901 International Parkway
Richardson, Texas 75081

EMail: dave.marshall@wcom.com

Luca Martini
Cisco Systems, Inc.
9155 East Nichols Avenue, Suite 400
Englewood, CO, 80112

EMail: lmartini@cisco.com

Monique Jeanne Morrow
Cisco Systems, Inc.
Glatt-com, 2nd floor
CH-8301
Glattzentrum, Switzerland

EMail: mmorrow@cisco.com

Ravichander Vaidyanathan
Telcordia Technologies
445 South Street, Room 1C258B
Morristown, NJ 07960

EMail: vravi@research.telcordia.com

Adrian Smith
BT
BT Adastral Park
Martlesham Heath,
Ipswich IP5 3RE
UK

EMail: adrian.ca.smith@bt.com

Vijay Srinivasan
1200 Bridge Parkway
Redwood City, CA 94065

EMail: vsriniva@cosinecom.com

Alain Vedrenne
Equant
Heraklion, 1041 route des Dolines, BP347
06906 Sophia Antipolis, Cedex, France

EMail: Alain.Vedrenne@equant.com

19. Normative References

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

[BGP-MP] Bates, T., Rekhter, Y., Chandra, R., and D. Katz, "Multiprotocol Extensions for BGP-4", RFC 2858, June 2000.

[BGP-EXTCOMM] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, February 2006.

[MPLS-ARCH] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.

[MPLS-BGP] Rekhter, Y. and E. Rosen, "Carrying Label Information in BGP-4", RFC 3107, May 2001.

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

20. Informative References

[BGP-AS4] Vohra, Q. and E. Chen, "BGP Support for Four-Octet AS Number Space", Work in Progress, March 2004.

[BGP-ORF] Chen, E. and Y. Rekhter, "Cooperative Route Filtering Capability for BGP-4", Work in Progress, March 2004.

[BGP-RFSH] Chen, E., "Route Refresh Capability for BGP-4", RFC 2918, September 2000.

[BGP-RR] Bates, T., Chandra, R., and E. Chen, "BGP Route Reflection - An Alternative to Full Mesh IBGP", RFC 2796, April 2000.

[IANA] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 2434, October 1998.

[MPLS-ATM] Davie, B., Lawrence, J., McCloghrie, K., Rosen, E., Swallow, G., Rekhter, Y., and P. Doolan, "MPLS using LDP and ATM VC Switching", RFC 3035, January 2001.

[MPLS/BGP-IPsec] Rosen, E., De Clercq, J., Paridaens, O., T'Joens, Y., and C. Sargor, "Architecture for the Use of PE-PE IPsec Tunnels in BGP/MPLS IP VPNs", Work in Progress, March 2004.

[MPLS-FR] Conta, A., Doolan, P., and A. Malis, "Use of Label Switching on Frame Relay Networks Specification", RFC 3034, January 2001.

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

[MPLS-LDP] Andersson, L., Doolan, P., Feldman, N., Fredette, A., and B. Thomas, "LDP Specification", RFC 3036, January 2001.

[MPLS-RSVP] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

[OSPFv2] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

[PASTE] Li, T. and Y. Rekhter, "A Provider Architecture for Differentiated Services and Traffic Engineering (PASTE)", RFC 2430, October 1998.

[RIP] Malkin, G., "RIP Version 2", STD 56, RFC 2453, November 1998.

[OSPF-2547-DNBIT] Rosen, E., Psenak, P., and P. Pillay-Esnault, "Using an LSA Options Bit to Prevent Looping in BGP/MPLS IP VPNs", Work in Progress, March 2004.

[TCP-MD5] Heffernan, A., "Protection of BGP Sessions via the TCP MD5 Signature Option", RFC 2385, August 1998.

[VPN-MCAST] Rosen, E., Cai, Y., and J. Wijsnands, "Multicast in MPLS/BGP VPNs", Work in Progress, May 2004.

[VPN-OSPF] Rosen, E., Psenak, P., and P. Pillay-Esnault, "OSPF as the PE/CE Protocol in BGP/MPLS VPNs", Work in Progress, February 2004.

Authors' Addresses

Eric C. Rosen
Cisco Systems, Inc.
1414 Massachusetts Avenue
Boxborough, MA 01719

EMail: erosen@cisco.com

Yakov Rekhter
Juniper Networks
1194 N. Mathilda Avenue
Sunnyvale, CA 94089

EMail: yakov@juniper.net

Full Copyright Statement

Intellectual Property

Acknowledgement

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).