Network Working Group
Request for Comments: 4110
Category: Informational

R. Callon
Juniper Networks
M. Suzuki
NTT Corporation

July 2005

A Framework for Layer 3 Provider-Provisioned Virtual Private Networks (PPVPNs)

Status of This Memo

Copyright Notice

Abstract

This document provides a framework for Layer 3 Provider-Provisioned Virtual Private Networks (PPVPNs). This framework is intended to aid in the standardization of protocols and mechanisms for support of layer 3 PPVPNs. It is the intent of this document to produce a coherent description of the significant technical issues that are important in the design of layer 3 PPVPN solutions. Selection of specific approaches, making choices regarding engineering tradeoffs, and detailed protocol specification, are outside of the scope of this framework document.

Table of Contents

1. Introduction

1.1. Objectives of the Document

This document provides a framework for Layer 3 Provider-Provisioned Virtual Private Networks (PPVPNs). This framework is intended to aid in standardizing protocols and mechanisms to support interoperable layer 3 PPVPNs.

The term "provider-provisioned VPNs" refers to Virtual Private Networks (VPNs) for which the Service Provider (SP) participates in management and provisioning of the VPN, as defined in section 1.3. There are multiple ways in which a provider can participate in managing and provisioning a VPN; therefore, there are multiple different types of PPVPNs. The framework document discusses layer 3 VPNs (as defined in section 1.3).

First, this document provides a reference model for layer 3 PPVPNs. Then technical aspects of layer 3 PPVPN operation are discussed, first from the customer's point of view, then from the providers point of view. Specifically, this includes discussion of the technical issues which are important in the design of standards and mechanisms for the operation and support of layer 3 PPVPNs. Furthermore, technical aspects of layer 3 PPVPN interworking are clarified. Finally, security issues as they apply to layer 3 PPVPNs are addressed.

This document takes a "horizontal description" approach. For each technical issue, it describes multiple approaches. To specify a particular PPVPN strategy, one must choose a particular way of solving each problem, but this document does not make choices, and does not select any particular approach to support VPNs.

The "vertical description" approach is taken in other documents, viz., in the documents that describe particular PPVPN solutions. Note that any specific solution will need to make choices based on SP requirements, customer needs, implementation cost, and engineering tradeoffs. Solutions will need to chose between flexibility

(supporting multiple options) and conciseness (selection of specific options in order to simplify implementation and deployment). While a framework document can discuss issues and criteria which are used as input to these choices, the specific selection of a solution is outside of the scope of a framework document.

1.2. Overview of Virtual Private Networks

The term "Virtual Private Network" (VPN) refers to a set of communicating sites, where (a) communication between sites outside the set and sites inside the set is restricted, but (b) communication between sites in the VPN takes place over a network infrastructure that is also used by sites which are not in the VPN. The fact that the network infrastructure is shared by multiple VPNs (and possibly also by non-VPN traffic) is what distinguishes a VPN from a private network. We will refer to this shared network infrastructure as the "VPN Backbone".

The logical structure of the VPN, such as addressing, topology, connectivity, reachability, and access control, is equivalent to part of or all of a conventional private network using private facilities [RFC2764] [VPN-2547BIS].

In this document, we are concerned only with the case where the shared network infrastructure (VPN backbone) is an IP and/or MPLS network. Further, we are concerned only with the case where the Service Provider's edge devices, whether at the provider edge (PE) or at the Customer Edge (CE), determine how to route VPN traffic by looking at the IP and/or MPLS headers of the packets they receive from the customer's edge devices; this is the distinguishing feature of Layer 3 VPNs.

In some cases, one SP may offer VPN services to another SP. The former SP is known as a carrier of carriers, and the service it offers is known as "carrier of carriers" service. In this document, in cases where the customer could be either an enterprise or SP network, we will make use of the term "customer" to refer to the user of the VPN services. Similarly we will use the term "customer network" to refer to the user's network.

VPNs may be intranets, in which the multiple sites are under the control of a single customer administration, such as multiple sites of a single company. Alternatively, VPNs may be extranets, in which the multiple sites are controlled by administrations of different customers, such as sites corresponding to a company, its suppliers, and its customers.

Figure 1.1. illustrates an example network, which will be used in the discussions below. PE1 and PE2 are Provider Edge devices within an SP network. CE1, CE2, and CE3 are Customer Edge devices within a customer network. Routers r3, r4, r5, and r6 are IP routers internal to the customer sites.

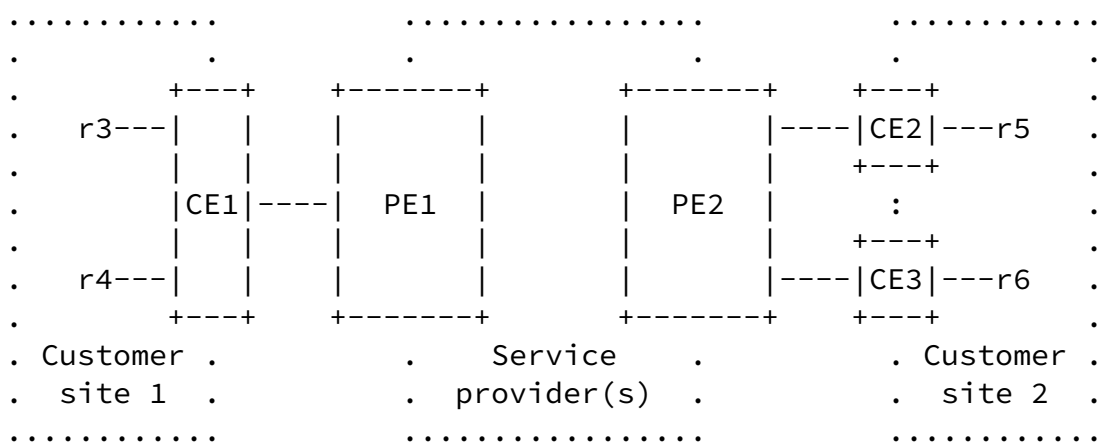


Figure 1.1.: VPN interconnecting two sites.

In many cases, Provider Edge (PE) and Customer Edge (CE) devices may be either routers or LSRs.

In this document, the Service Providers' network is an IP or MPLS network. It is desired to interconnect the customer network sites via the Service Providers' network. Some VPN solutions require that the VPN service be provided either over a single SP network, or over a small set of closely cooperating SP networks. Other VPN solutions are intended to allow VPN service to be provided over an arbitrary set of minimally cooperating SP networks (i.e., over the public Internet).

In many cases, customer networks will make use of private IP addresses [RFC1918] or other non-unique IP address (i.e., unregistered addresses); there is no guarantee that the IP addresses used in the customer network are globally unique. The addresses used in one customer's network may overlap the addresses used in others. However, a single PE device can be used to provide VPN service to multiple customer networks, even if those customer networks have overlapping addresses. In PE-based layer 3 VPNs, the PE devices may route the VPN traffic based on the customer addresses found in the IP headers; this implies that the PE devices need to maintain a level of isolation between the packets from different customer networks. In CE-based layer 3 VPNs, the PEs do not make routing decisions based on the customer's private addresses, so this issue does not arise. For either PE or CE-based VPNs, the fact that the VPNs do not necessarily use globally unique address spaces also implies that IP packets from a customer network cannot be transmitted over the SP network in their native form. Instead, some form of encapsulation/tunneling must be used.

Tunneling is also important for other reasons, such as providing isolation between different customer networks, allowing a wide range of protocols to be carried over an SP network, etc. Different QoS and security characteristics may be associated with different tunnels.

1.3. Types of VPNs

This section describes multiple types of VPNs, and some of the engineering tradeoffs between different types. It is not up to this document to decide between different types of VPNs. Different types of VPNs may be appropriate in different situations.

There is a wide spectrum of types of possible VPNs, and it is difficult to split the types of VPNs into clearly distinguished categories.

As an example, consider a company making use of a private network, with several sites interconnected via leased lines. All routing is done via routers which are internal to the private network.

At some point, the administrator of the private network might decide to replace the leased lines by ATM links (using an ATM service from an SP). Here again all IP-level routing is done between customer premises routers, and managed by the private network administrator.

In order to reduce the network management burden on the private network, the company may decide to make use of a provider-provisioned CE devices [VPN-CE]. Here the operation of the network might be unchanged, except that the CE devices would be provided by and managed by an SP.

The SP might decide that it is too difficult to manually configure each CE-CE link. This might lead the SP to replace the ATM links with a layer 2 VPN service between CE devices [VPN-L2]. Auto-discovery might be used to simplify configuration of links between CE devices, and an MPLS service might be used between CE devices instead of an ATM service (for example, to take advantage of the provider's high speed IP or MPLS backbone).

After a while the SP might decide that it is too much trouble to be managing a large number of devices at the customers' premises, and might instead physically move these routers to be on the provider premises. Each edge router at the provider premises might nonetheless be dedicated to a single VPN. The operation might remain unchanged (except that links from the edge routers to other routers in the private network become MAN links instead of LAN links, and the link from the edge routers to provider core routers become LAN links

instead of MAN links). The routers in question can now be considered

to be provider edge routers, and the service provided by the SP has now become essentially a layer 3 VPN service.

In order to minimize the cost of equipment, the provider might decide to replace several dedicated PE devices with a single physical router with the capability of running virtual routers (VR) [VPN-VR]. Protocol operation may remain unchanged. In this case the provider is offering a layer 3 VPN service making use of a VR capability. Note that autodiscovery might be used in a manner which is very similar to how it had been done in the layer 2 VPN case described above (for example, BGP might be used between VRs for discovery of other VRs supporting the same VPN).

Finally, in order to simplify operation of routing protocols for the private network over the SP network, the provider might decide to aggregate multiple instances of routing into a single instance of BGP [VPN-2547BIS].

In practice it is highly unlikely that any one network would actually evolve through all of these approaches at different points in time. However, this example illustrates that there is a continuum of possible approaches, and each approach is relatively similar to at least some of the other possible approaches for supporting VPN services. Some techniques (such as auto-discovery of VPN sites) may be common between multiple approaches.

1.3.1. CE- vs PE-based VPNs

The term "CE-based VPN" (or Customer Edge-based Virtual Private Network) refers to an approach in which the PE devices do not know anything about the routing or the addressing of the customer networks. The PE devices offer a simple IP service, and expect to receive IP packets whose headers contain only globally unique IP addresses. What makes a CE-based VPN into a Provider-Provisioned VPN is that the SP takes on the task of managing and provisioning the CE devices [VPN-CE].

In CE-based VPNs, the backbone of the customer network is a set of tunnels whose endpoints are the CE devices. Various kinds of tunnels may be used (e.g., GRE, IP-in-IP, IPsec, L2TP, MPLS), the only overall requirement being that sending a packet through the tunnel requires encapsulating it with a new IP header whose addresses are globally unique.

For customer provisioned CE-based VPNs, provisioning and management of the tunnels is the responsibility of the customer network administration. Typically, this makes use of manual configuration of

the tunnels. In this case the customer is also responsible for operation of the routing protocol between CE devices. (Note that discussion of customer provisioned CE-based VPNs is out of scope of the document).

For provider-provisioned CE-based VPNs, provisioning and management of the tunnels is the responsibility of the SP. In this case the provider may also configure routing protocols on the CE devices. This implies that routing in the private network is partially under the control of the customer, and partially under the control of the SP.

For CE-based VPNs (whether customer or provider-provisioned) routing in the customer network treats the tunnels as layer 2 links.

In a PE-based VPN (or Provider Edge-based Virtual Private Network), customer packets are carried through the SP networks in tunnels, just as they are in CE-based VPNs. However, in a PE-based VPN, the tunnel endpoints are the PE devices, and the PE devices must know how to route the customer packets, based on the IP addresses that they carry. In this case, the CE devices themselves do not have to have any special VPN capabilities, and do not even have to know that they are part of a VPN.

In this document we will use the generic term "VPN Edge Device" to refer to the device, attached to both the customer network and the VPN backbone, that performs the VPN-specific functions. In the case of CE-based VPNs, the VPN Edge Device is a CE device. In the case of PE-based VPNs, the VPN Edge Device is a PE device.

1.3.2. Types of PE-based VPNs

Different types of PE-based VPNs may be distinguished by the service offered.

- Layer 3 service

  When a PE receives a packet from a CE, it determines how to forward the packet by considering both the packet's incoming link, and the layer 3 information in the packet's header.

- Layer 2 service

  When a PE receives a frame from a CE, it determines how to forward the packet by considering both the packet's incoming link, and the layer 2 information in the frame header (such as FR, ATM, or MAC

header). (Note that discussion of layer 2 service is out of scope of the document).

1.3.3. Layer 3 PE-based VPNs

A layer 3 PE-based VPN is one in which the SP takes part in IP level forwarding based on the customer network's IP address space. In general, the customer network is likely to make use of private and/or non-unique IP addresses. This implies that at least some devices in the provider network needs to understand the IP address space as used in the customer network. Typically this knowledge is limited to the PE devices which are directly attached to the customer.

In a layer 3 PE-based VPN, the provider will need to participate in some aspects of management and provisioning of the VPNs, such as ensuring that the PE devices are configured to support the correct VPNs. This implies that layer 3 PE-based VPNs are by definition provider-provisioned VPNs.

Layer 3 PE-based VPNs have the advantage that they offload some aspects of VPN management from the customer network. From the perspective of the customer network, it looks as if there is just a normal network; specific VPN functionality is hidden from the customer network. Scaling of the customer network's routing might also be improved, since some layer 3 PE-based VPN approaches avoid the need for the customer's routing algorithm to see "N squared" (actually N*(N-1)/2) point to point duplex links between N customer sites.

However, these advantages come along with other consequences. Specifically, the PE devices must have some knowledge of the routing, addressing, and layer 3 protocols of the customer networks to which they attach. One consequence is that the set of layer 3 protocols which can be supported by the VPN is limited to those supported by the PE (which in practice means, limited to IP). Another consequence is that the PE devices have more to do, and the SP has more per-customer management to do.

An SP may offer a range of layer 3 PE-based VPN services. At one end of the range is a service limited to simply providing connectivity (optionally including QoS support) between specific customer network sites. This is referred to as "Network Connectivity Service". There

is a spectrum of other possible services, such as firewalls, user or site of origin authentication, and address assignment (e.g., using Radius or DHCP).

1.4. Scope of the Document

This framework document will discuss methods for providing layer 3 PE-based VPNs and layer 3 provider-provisioned CE-based VPNs. This may include mechanisms which will can be used to constrain

connectivity between sites, including the use and placement of firewalls, based on administrative requirements [PPVPN-REQ] [L3VPN-REQ]. Similarly the use and placement of NAT functionality is discussed. However, this framework document will not discuss methods for additional services such as firewall administration and address assignment. A discussion of specific firewall mechanisms and policies, and detailed discussion of NAT functionality, are outside of the scope of this document.

This document does not discuss those forms of VPNs that are outside of the scope of the IETF Provider-Provisioned VPN working group. Specifically, this document excludes discussion of PPVPNs using VPN native (non-IP, non-MPLS) protocols as the base technology used to provide the VPN service (e.g., native ATM service provided using ATM switches with ATM signaling). However, this does not mean to exclude multiprotocol access to the PPVPN by customers.

1.5. Terminology

Backdoor Links: Links between CE devices that are provided by the end customer rather than the SP; may be used to interconnect CE devices in multiple-homing arrangements.

CE-based VPN: An approach in which all the VPN-specific procedures are performed in the CE devices, and the PE devices are not aware in any way that some of the traffic they are processing is VPN traffic.

Customer: A single organization, corporation, or enterprise that administratively controls a set of sites belonging to a VPN.

Customer Edge (CE) Device: The equipment on the customer side of the SP-customer boundary (the customer interface).

IP Router: A device which forwards IP packets, and runs associated IP routing protocols (such as OSPF, IS-IS, RIP, BGP, or similar protocols). An IP router might optionally also be an LSR. The term "IP router" is often abbreviated as "router".

Label Switching Router: A device which forwards MPLS packets and runs associated IP routing and signaling protocols (such as LDP, RSVP-TE, CR-LDP, OSPF, IS-IS, or similar protocols). A label switching router is also an IP router.

PE-Based VPNs: The PE devices know that certain traffic is VPN traffic. They forward the traffic (through tunnels) based on the destination IP address of the packet, and optionally on based on other information in the IP header of the packet. The PE devices are

themselves the tunnel endpoints. The tunnels may make use of various encapsulations to send traffic over the SP network (such as, but not restricted to, GRE, IP-in-IP, IPsec, or MPLS tunnels).

Private Network: A network which allows communication between a restricted set of sites, over an IP backbone that is used only to carry traffic to and from those sites.

Provider Edge (PE) Device: The equipment on the SP side of the SP-customer boundary (the customer interface).

Provider-Provisioned VPNs (PPVPNs): VPNs, whether CE-based or PE-based, that are actively managed by the SP rather than by the end customer.

Route Reflectors: An SP-owned network element that is used to distribute BGP routes to the SP's BGP-enabled routers.

Virtual Private Network (VPN): Restricted communication between a set of sites, making use of an IP backbone which is shared by traffic that is not going to or coming from those sites.

Virtual Router (VR): An instance of one of a number of logical routers located within a single physical router. Each logical router emulates a physical router using existing mechanisms and tools for

configuration, operation, accounting, and maintenance.

VPN Forwarding Instance (VFI): A logical entity that resides in a PE that includes the router information base and forwarding information base for a VPN.

VPN Backbone: IP and/or MPLS network which is used to carry VPN traffic between the customer sites of a particular VPN.

VPN Edge Device: Device, attached to both the VPN backbone and the customer network, which performs VPN-specific functions. For PE-based VPNs, this is the PE device; for CE-based VPNs, this is the CE device.

VPN Routing: Routing that is specific to a particular VPN.

VPN Tunnel: A logical link between two PE or two CE entities, used to carry VPN traffic, and implemented by encapsulating packets that are transmitted between those two entities.

1.6. Acronyms

ATM	Asynchronous Transfer Mode
BGP	Border Gateway Protocol
CE	Customer Edge
CLI	Command Line Interface
CR-LDP	Constraint-based Routing Label Distribution Protocol
EBGP	External Border Gateway Protocol
FR	Frame Relay
GRE	Generic Routing Encapsulation
IBGP	Internal Border Gateway Protocol
IKE	Internet Key Exchange
IGP	Interior Gateway Protocol (e.g., RIP, IS-IS and OSPF are all IGPs)
IP	Internet Protocol (same as IPv4)
IPsec	Internet Protocol Security protocol
IPv4	Internet Protocol version 4 (same as IP)
IPv6	Internet Protocol version 6

IS-IS	Intermediate System to Intermediate System routing protocol
L2TP	Layer 2 Tunneling Protocol
LAN	Local Area Network
LDAP	Lightweight Directory Access Protocol
LDP	Label Distribution Protocol
LSP	Label Switched Path
LSR	Label Switching Router
MIB	Management Information Base
MPLS	Multi Protocol Label Switching
NBMA	Non-Broadcast Multi-Access
NMS	Network Management System
OSPF	Open Shortest Path First routing protocol
P	Provider equipment
PE	Provider Edge
PPVPN	Provider-Provisioned VPN
QoS	Quality of Service
RFC	Request For Comments
RIP	Routing Information Protocol
RSVP	Resource Reservation Protocol
RSVP-TE	Resource Reservation Protocol with Traffic Engineering Extensions
SNMP	Simple Network Management Protocol
SP	Service Provider
VFI	VPN Forwarding Instance
VPN	Virtual Private Network
VR	Virtual Router

2. Reference Models

This section describes PPVPN reference models. The purpose of discussing reference models is to clarify the common components and pieces that are needed to build and deploy a PPVPN. Two types of VPNs, layer 3 PE-based VPN and layer 3 provider-provisioned CE-based VPN are covered in separated sections below.

2.1. Reference Model for Layer 3 PE-based VPN

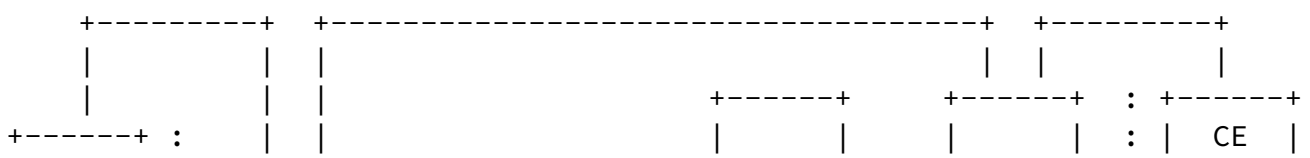
This subsection describes functional components and their

relationship for implementing layer 3 PE-based VPN.

Figure 2.1 shows the reference model for layer 3 PE-based VPNs and Figures 2.2 and 2.3 show relationship between entities in the reference model.

As shown in Figure 2.1, the customer interface is defined as the interface which exists between CE and PE devices, and the network interface is defined as the interface which exists between a pair of PE devices.

Figure 2.2 illustrates a single logical tunnel between each pair of VFIs supporting the same VPN. Other options are possible. For example, a single tunnel might occur between two PEs, with multiple per-VFI tunnels multiplexed over the PE to PE tunnel. Similarly, there may be multiple tunnels between two VFIs, for example to optimize forwarding within the VFI. Other possibilities will be discussed later in this framework document.



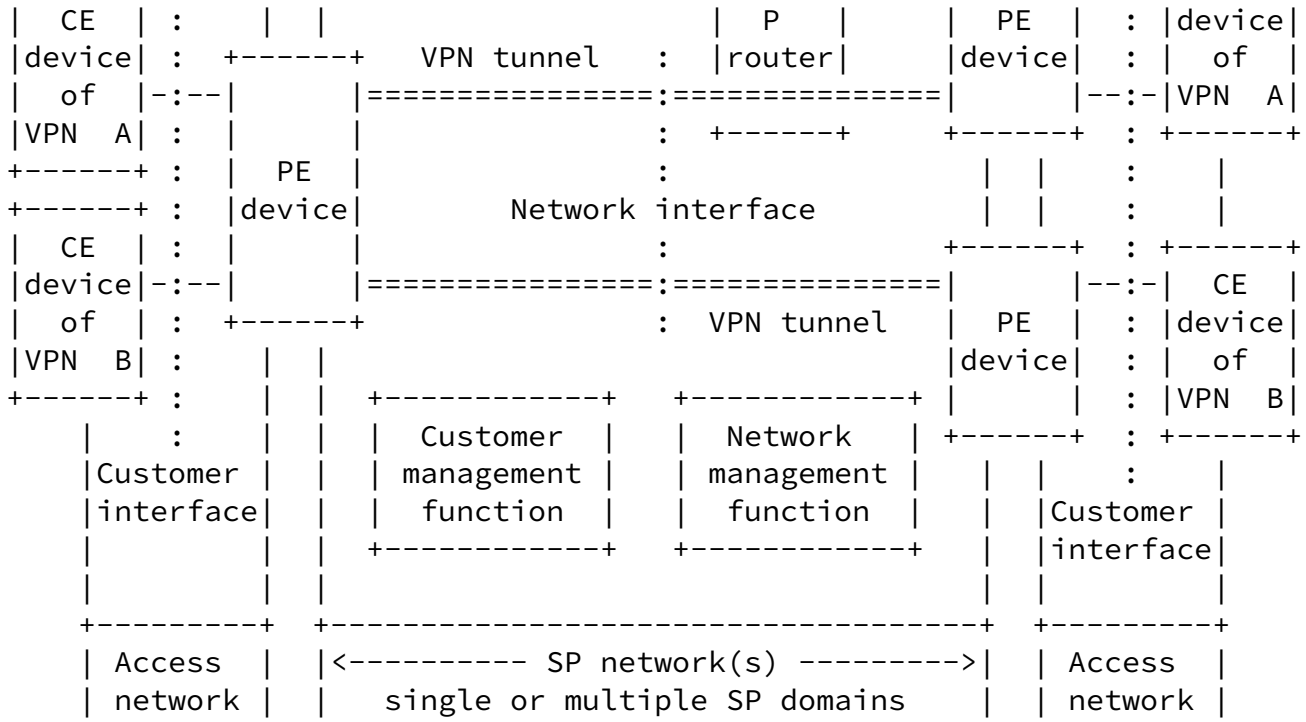


Figure 2.1: Reference model for layer 3 PE-based VPN.

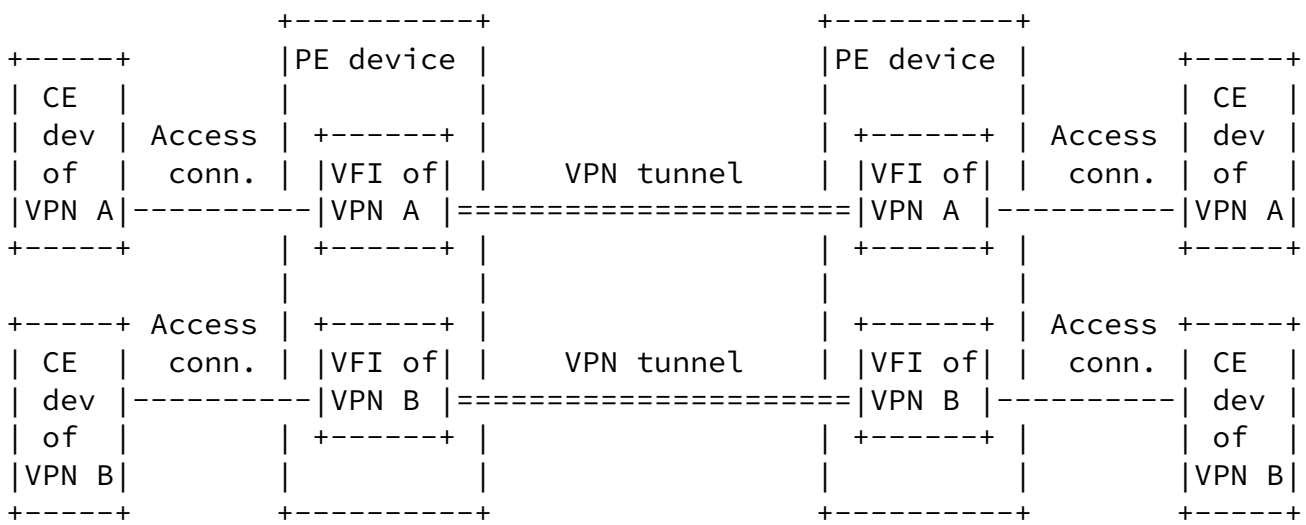


Figure 2.2: Relationship between entities in reference model (1).

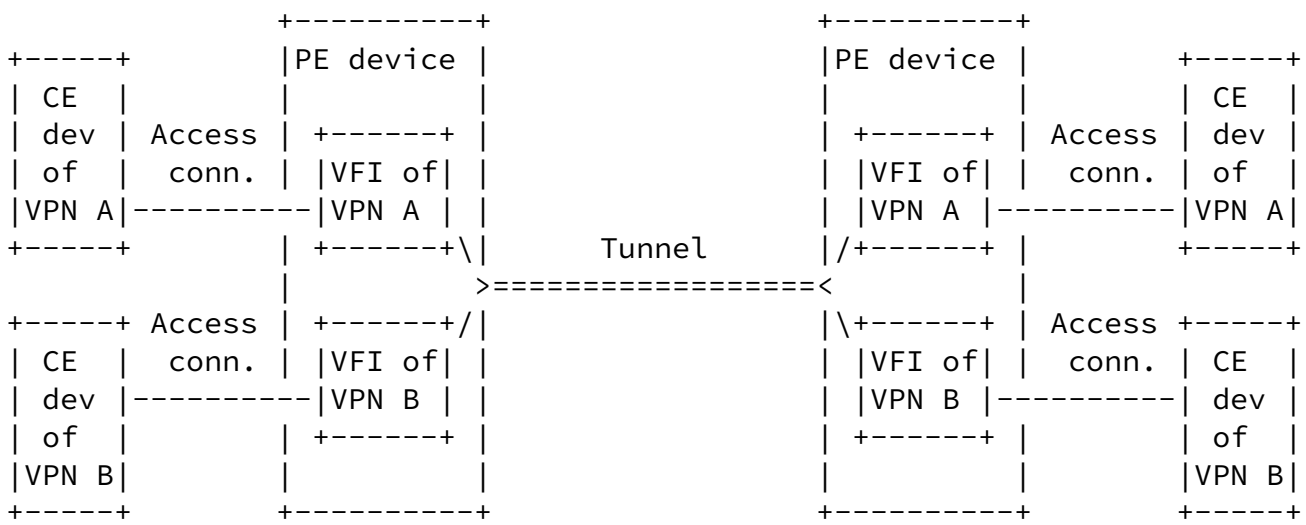


Figure 2.3: Relationship between entities in reference model (2).

2.1.1. Entities in the Reference Model

The entities in the reference model are described below.

- Customer edge (CE) device

  In the context of layer 3 provider-provisioned PE-based VPNs, a CE device may be a router, LSR, or host that has no VPN-specific functionality. It is attached via an access connection to a PE device.

- P router

  A router within a provider network which is used to interconnect PE devices, but which does not have any VPN state and does not have any direct attachment to CE devices.

- Provider edge (PE) device

  In the context of layer 3 provider-provisioned PE-based VPNs, a PE device implements one or more VFIs and maintains per-VPN state for the support of one or more VPNs. It may be a router, LSR, or other device that includes VFIs and provider edge VPN functionality such as provisioning, management, and traffic classification and separation. (Note that access connections are terminated by VFIs from the functional point of view). A PE device is attached via an access connection to one or more CE devices.

- Customer site

  A customer site is a set of users that have mutual IP reachability without use of a VPN backbone that goes beyond the site.

- SP networks

  An SP network is an IP or MPLS network administered by a single service provider.

- Access connection

  An access connection represents an isolated layer 2 connectivity between a CE device and a PE device. Access connections can be, e.g., dedicated physical circuits, logical circuits (such as FR, ATM, and MAC), or IP tunnels (e.g., using IPsec, L2TP, or MPLS).

- Access network

  An access network provides access connections between CE and PE devices. It may be a TDM network, layer 2 network (e.g., FR, ATM, and Ethernet), or IP network over which access is tunneled (e.g., using L2TP [RFC2661] or MPLS).

- VPN tunnel

  A VPN tunnel is a logical link between two VPN edge devices. A VPN packet is carried on a tunnel by encapsulating it before transmitting it over the VPN backbone.

  Multiple VPN tunnels at one level may be hierarchically multiplexed into a single tunnel at another level. For example, multiple per-VPN tunnels may be multiplexed into a single PE to PE tunnel (e.g., GRE, IP-in-IP, IPsec, or MPLS tunnel). This is illustrated in Figure 2.3. See section 4.3 for details.

- VPN forwarding instance (VFI)

  A single PE device is likely to be connected to a number of CE devices. The CE devices are unlikely to all be in the same VPN.

The PE device must therefore maintain a separate forwarding instances for each VPN to which it is connected. A VFI is a logical entity, residing in a PE, that contains the router information base and forwarding information base for a VPN. The interaction between routing and VFIs is discussed in section 4.4.2.

- Customer management function

  The customer management function supports the provisioning of customer specific attributes, such as customer ID, personal information (e.g., name, address, phone number, credit card number, and etc.), subscription services and parameters, access control policy information, billing and statistical information, and etc.

  The customer management function may use a combination of SNMP manager, directory service (e.g., LDAP [RFC3377]), or proprietary network management system.

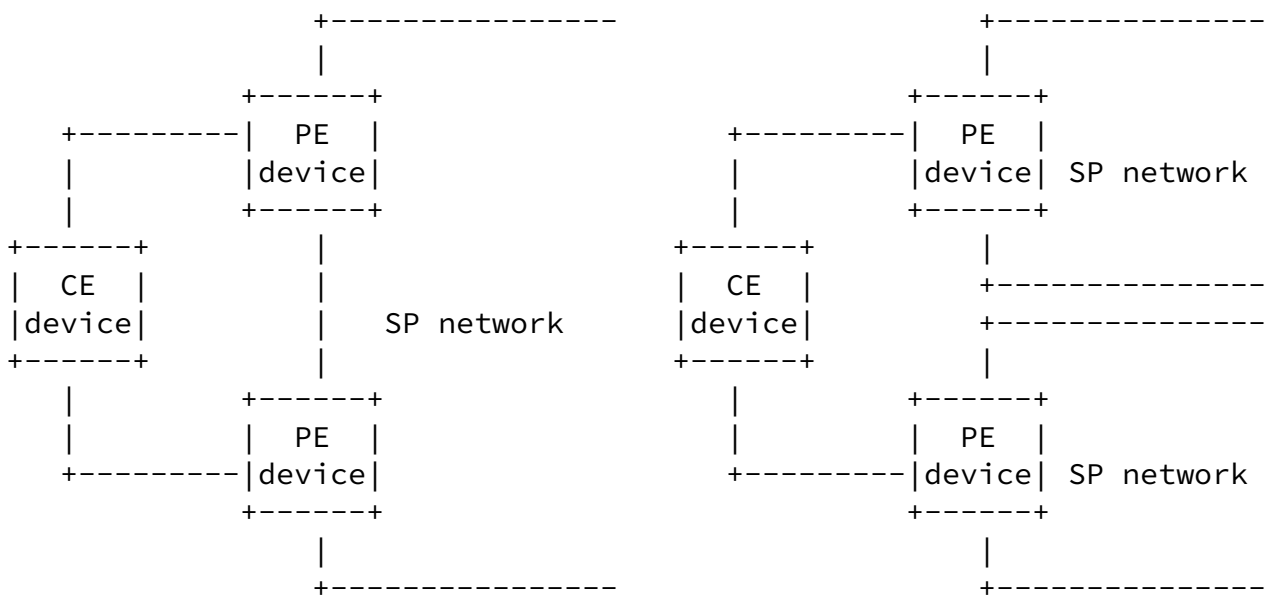
- Network management function

  The network management function supports the provisioning and monitoring of PE or CE device attributes and their relationships.

  The network management function may use a combination of SNMP manager, directory service (e.g., LDAP [RFC3377]), or proprietary network management system.

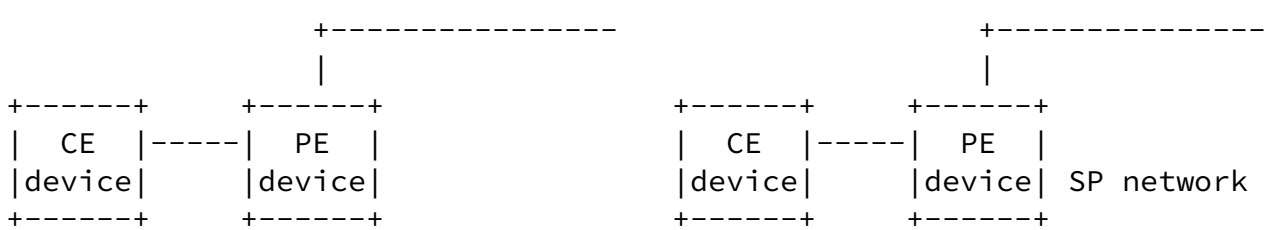
2.1.2. Relationship Between CE and PE

For robustness, a CE device may be connected to more than one PE device, resulting in a multi-homing arrangement. Four distinct types of multi-homing arrangements, shown in Figure 2.4, may be supported.



This type includes a CE device connected
to a PE device via two access connections.

(a) (b)



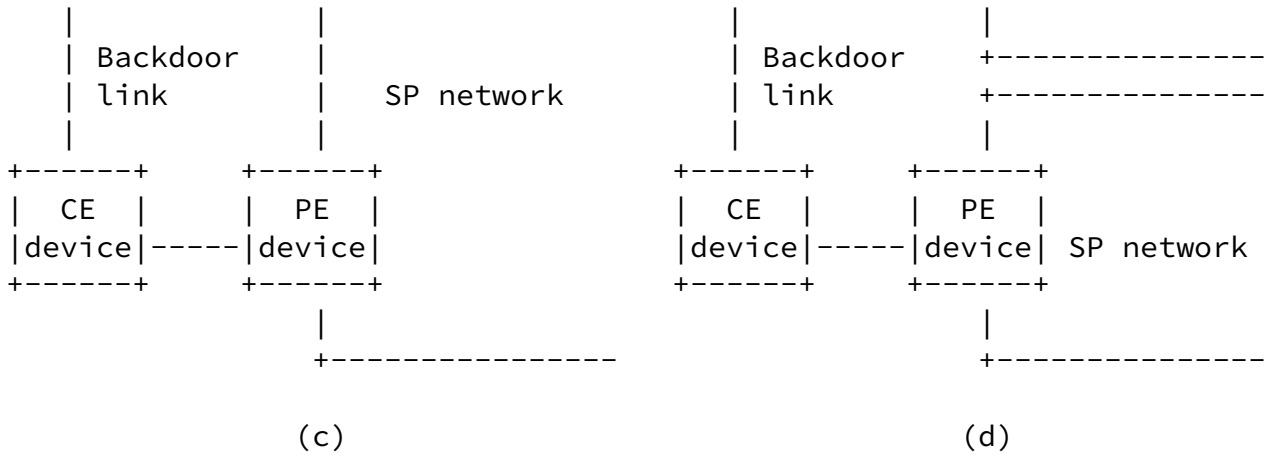


Figure 2.4: Four types of double-homing arrangements.

2.1.3. Interworking Model

It is quite natural to assume that multiple different layer 3 VPN approaches may be implemented, particularly if the VPN backbone includes more than one SP network. For example, (1) each SP chooses one or more layer 3 PE-based VPN approaches out of multiple vendor's implementations, implying that different SPs may choose different

approaches; and (2) an SP may deploy multiple networks of layer 3 PE-based VPNs (e.g., an old network and a new network). Thus it is important to allow interworking of layer 3 PE-based VPNs making use of multiple different layer 3 VPN approaches.

There are three scenarios that enable layer 3 PE-based VPN interworking among different approaches.

o Interworking function

This scenario enables interworking using a PE that is located at one or more points which are logically located between VPNs based on different layer 3 VPN approaches. For example, this PE may be located on the boundary between SP networks which make use of different layer 3 VPN approaches [VPN-DISC]. A PE at one of these points is called an interworking function (IWF), and an example configuration is shown in Figure 2.5.

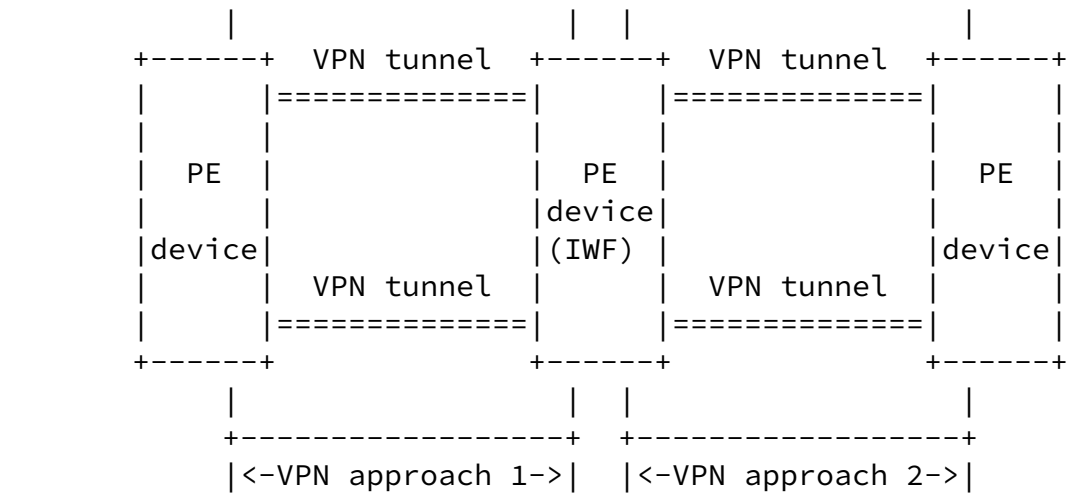
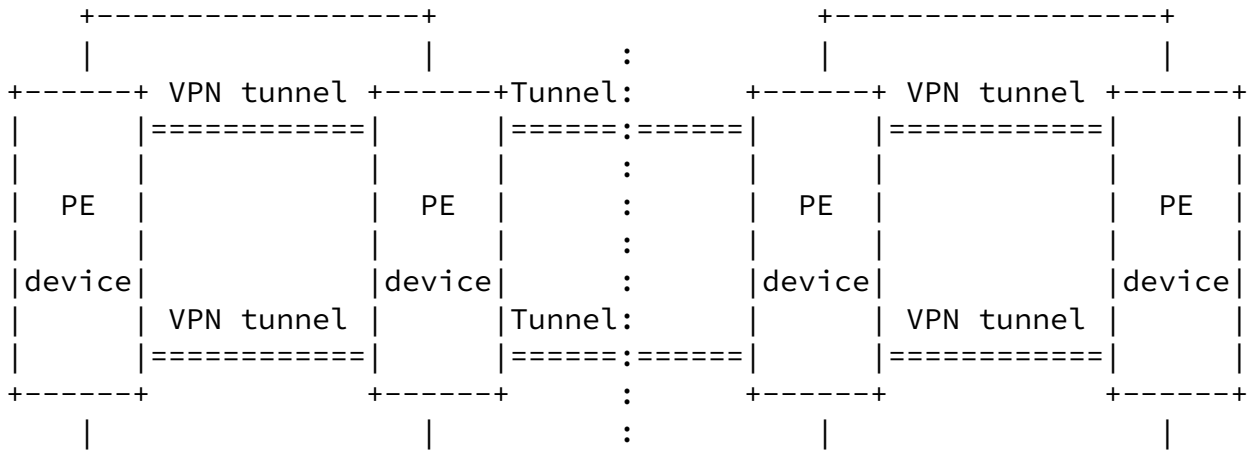


Figure 2.5: Interworking function.

o Interworking interface

This scenario enables interworking using tunnels between PEs supporting by different layer 3 VPN approaches. As shown in Figure 2.6, interworking interface is defined as the interface which exists between a pair of PEs and connects two SP networks implemented with different approaches. This interface is similar to the customer interface located between PE and CE, but the interface is supported by tunnels to identify VPNs, while the customer interface is supported by access connections.



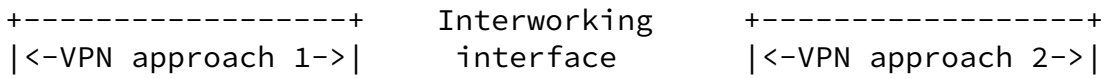


Figure 2.6: Interworking interface.

o Customer-based interworking

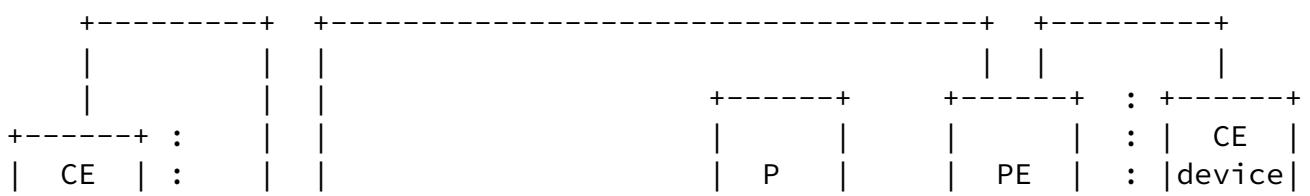
If some customer site has a CE attached to one kind of VPN, and a CE attached to another kind, communication between the two kinds of VPN occurs automatically.

2.2. Reference Model for Layer 3 Provider-Provisioned CE-based VPN

This subsection describes functional components and their relationship for implementing layer 3 provider-provisioned CE-based VPN.

Figure 2.7 shows the reference model for layer 3 provider-provisioned CE-based VPN. As shown in Figure 2.7, the customer interface is defined as the interface which exists between CE and PE devices.

In this model, a CE device maintains one or more VPN tunnel endpoints, and a PE device has no VPN-specific functionality. As a result, the interworking issues of section 2.1.3 do not arise.



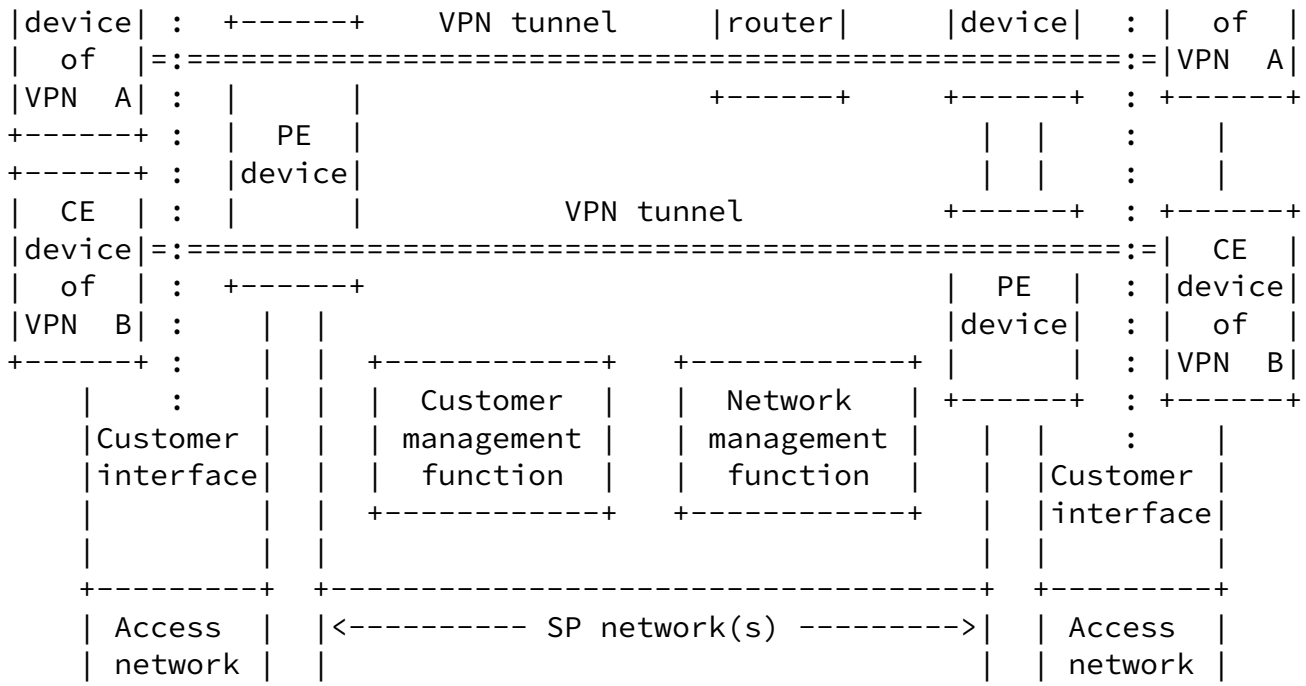


Figure 2.7: Reference model for layer 3
provider-provisioned CE-based VPN.

2.2.1. Entities in the Reference Model

The entities in the reference model are described below.

- Customer edge (CE) device

  In the context of layer 3 provider-provisioned CE-based VPNs, a CE device provides layer 3 connectivity to the customer site. It may be a router, LSR, or host that maintains one or more VPN tunnel endpoints. A CE device is attached via an access connection to a PE device and usually located at the edge of a customer site or co-located on an SP premises.

- P router (see section 2.1.1)

- Provider edge (PE) device

  In the context of layer 3 provider-provisioned CE-based VPNs, a PE device may be a router, LSR, or other device that has no VPN-specific functionality. It is attached via an access connection to one or more CE devices.

- Customer Site (see section 2.1.1)

- SP networks

  An SP network is a network administrated by a single service provider. It is an IP or MPLS network. In the context of layer 3 provider-provisioned CE-based VPNs, the SP network consists of the SP's network and the SP's management functions that manage both its own network and the customer's VPN functions on the CE device.

- Access connection (see section 2.1.1)

- Access network (see section 2.1.1)

- VPN tunnel

  A VPN tunnel is a logical link between two entities which is created by encapsulating packets within an encapsulating header for purpose of transmission between those two entities for support of VPNs. In the context of layer 3 provider-provisioned CE-based VPNs, a VPN tunnel is an IP tunnel (e.g., using GRE, IP-in-IP, IPsec, or L2TP) or an MPLS tunnel between two CE devices over the SP's network.

- Customer management function (see section 2.1.1)

- Network management function

  The network management function supports the provisioning and monitoring of PE or CE device attributes and their relationships, covering PE and CE devices that define the VPN connectivity of the customer VPNs.

  The network management function may use a combination of SNMP manager, directory service (e.g., LDAP [RFC3377]), or proprietary network management system.

3. Customer Interface

3.1. VPN Establishment at the Customer Interface

3.1.1. Layer 3 PE-based VPN

It is necessary for each PE device to know which CEs it is attached to, and what VPNs each CE is associated with.

VPN membership refers to the association of VPNs, CEs, and PEs. A given CE belongs to one or more VPNs. Each PE is therefore

associated with a set of VPNs, and a given VPN has a set of associated PEs which are supporting that VPN. If a PE has at least one attached CE belonging to a given VPN, then state information for that VPN (e.g., the VPN routes) must exist on that PE. The set of VPNs that exist on a PE may change over time as customer sites are added to or removed from the VPNs.

In some layer 3 PE-based PPVPN schemes, VPN membership information (i.e., information about which PEs are attached to which VPNs) is explicitly distributed. In others, the membership information is inferred from other information that is distributed. Different schemes use the membership information in different ways, e.g., some to determine what set of tunnels to set up, some to constrain the distribution of VPN routing information.

A VPN site may be added or deleted as a result of a provisioning operation carried out by the network administrator, or may be dynamically added or deleted as a result of a subscriber initiated operation; thus VPN membership information may be either static or dynamic, as discussed below.

3.1.1.1. Static Binding

Static binding occurs when a provisioning action binds a particular PE-CE access link to a particular VPN. For example, a network administrator may set up a dedicated link layer connection, such as an ATM VCC or a FR DLCI, between a PE device and a CE device. In this case the binding between a PE-CE access connection and a particular VPN to fixed at provisioning time, and remains the same until another provisioning action changes the binding.

3.1.1.2. Dynamic Binding

Dynamic binding occurs when some real-time protocol interaction causes a particular PE-CE access link to be temporarily bound to a particular VPN. For example, a mobile user may dial up the provider network and carry out user authentication and VPN selection procedures. Then the PE to which the user is attached is not one permanently associated with the user, but rather one that is typically geographically close to where the mobile user happens to be. Another example of dynamic binding is that of a permanent access connection between a PE and a CE at a public facility such as a hotel

or conference center, where the link may be accessed by multiple users in turn, each of which may wish to connect to a different VPN.

To support dynamically connected users, PPP and RADIUS are commonly used, as these protocols provide for user identification, authentication and VPN selection. Other mechanisms are also

possible. For example a user's HTTP traffic may be initially intercepted by a PE and diverted to a provider hosted web server. After a dialogue that includes user authentication and VPN selection, the user can then be connected to the required VPN. This is sometimes referred to as a "captive portal".

Independent of the particular mechanisms used for user authentication and VPN selection, an implication of dynamic binding is that a user for a given VPN may appear at any PE at any time. Thus VPN membership may change at any time as a result of user initiated actions, rather than as a result of network provisioning actions. This suggests that there needs to be a way to distribute membership information rapidly and reliably when these user-initiated actions take place.

3.1.2. Layer 3 Provider-Provisioned CE-based VPN

In layer 3 provider-provisioned CE-based VPNs, the PE devices have no knowledge of the VPNs. A PE device attached to a particular VPN has no knowledge of the addressing or routing information of that specific VPN.

CE devices have IP or MPLS connectivity via a connection to a PE device, which just provides ordinary connectivity to the global IP address space or to an address space which is unique in a particular SPs network. The IP connectivity may be via a static binding, or via some kind of dynamic binding.

The establishment of the VPNs is done at each CE device, making use of the IP or MPLS connectivity to the others. Therefore, it is necessary for a given CE device to know which other CE devices belong to the same VPN. In this context, VPN membership refers to the association of VPNs and CE devices.

3.2. Data Exchange at the Customer Interface

3.2.1. Layer 3 PE-based VPN

For layer 3 PE-based VPNs, the exchange is normal IP packets, transmitted in the same form which is available for interconnecting routers in general. For example, IP packets may be exchanged over Ethernet, SONET, T1, T3, dial-up lines, and any other link layer available to the router. It is important to note that those link layers are strictly local to the interface for the purpose of carrying IP packets, and are terminated at each end of the customer interface. The IP packets may contain addresses which, while unique within the VPN, are not unique on the VPN backbone. Optionally, the data exchange may use MPLS to carry the IP packets.

3.2.2. Layer 3 Provider-Provisioned CE-based VPN

The data exchanged at the customer interface are always normal IP packets that are routable on the VPN backbone, and whose addresses are unique on the VPN backbone. Optionally, MPLS frames can be used, if the appropriate label-switched paths exist across the VPN backbone. The PE device does not know whether these packets are VPN packets or not. At the current time, MPLS is not commonly offered as a customer-visible service, so that CE-based VPNs most commonly make use of IP services.

3.3. Customer Visible Routing

Once VPN tunnels are set up between pairs of VPN edge devices, it is necessary to set up mechanisms which ensure that packets from the customer network get sent through the proper tunnels. This routing function must be performed by the VPN edge device.

3.3.1. Customer View of Routing for Layer 3 PE-based VPNs

There is a PE-CE routing interaction which enables a PE to obtain those addresses, from the customer network, that are reachable via the CE. The PE-CE routing interaction also enables a CE device to obtain those addresses, from the customer network, which are reachable via the PE; these will generally be addresses that are at other sites in the customer network.

The PE-CE routing interaction can make use of static routing, an IGP

(such as RIP, OSPF, IS-IS, etc.), or BGP.

If the PE-CE interaction is done via an IGP, the PE will generally maintain at least several independent IGP instances; one for the backbone routing, and one for each VPN. Thus the PE participates in the IGP of the customer VPNs, but the CE does not participate in the backbone's IGP.

If the PE-CE interaction is done via BGP, the PE MAY support one instance of BGP for each VPN, as well as an additional instance of BGP for the public Internet routes. Alternatively, the PE might support a single instance of BGP, using, e.g., different BGP Address Families to distinguish the public Internet routes from the VPN routes.

Routing information which a PE learns from a CE in a particular VPN must be forwarded to the other PEs that are attached to the same VPN. Those other PEs must then forward the information in turn to the other CEs of that VPN.

The PE-PE routing distribution can be done as part of the same routing instance to which the PE-CE interface belongs. Alternatively, it can be done via a different routing instance, possibly using a different routing algorithm. In this case, the PE must redistribute VPN routes from one routing instance to another.

Note that VPN routing information is never distributed to the P routers. VPN routing information is known at the edge of the VPN backbone, but not in the core.

If the VPN's IGP is different than the routing algorithm running on the CE-PE link, then the CE must support two routing instances, and must redistribute the VPN's routes from one instance to the other (e.g., [VPN-BGP-OSPF]).

In the case of layer 3 PE-based VPNs a single PE device is likely to provide service for several different VPNs. Since different VPNs may have address spaces which are not mutually unique, a PE device must have several forwarding tables, in general one for each VPN to which it is attached. These will be referred to as VPN Forwarding Instances (VFIs). Each VFI is a logical entity internal to the PE

device. VFIs are defined in section 2.1.1, and discussed in more detail in section 4.4.2.

The scaling and management of the customer network (as well as the operation of the VPN) will depend upon the implementation approach and the manner in which routing is done.

3.3.1.1. Routing for Intranets

In the intranet case all of the sites to be interconnected belong to the same administration (for example, the same company). The options for routing within a single customer network include:

- A single IGP area (using OSPF, IS-IS, or RIP)
- Multiple areas within a single IGP
- A separate IGP within each site, with routes redistributed from each site to backbone routing (i.e., to a backbone as seen by the customer network).

Note that these options look at routing from the perspective of the overall routing in the customer network. This list does not specify whether PE device is considered to be in a site or not. This issue is discussed below.

A single IGP area (such as a single OSPF area, a single IS-IS area, or a single instance of RIP between routers) may be used. One could have, all routers within the customer network (including the PEs, or more precisely, including a VFI within each PE) appear within a single area. Tunnels between the PEs could also appear as normal links.

In some cases the multi-level hierarchy of OSPF or IS-IS may be used. One way to apply this to VPNs would be to have each site be a single OSPF or IS-IS area. The VFIs will participate in routing within each site as part of that area. The VFIs may then be interconnected as the backbone (OSPF area 0 or IS-IS level 2). If OSPF is used, the VFIs therefore appear to the customer network as area border routers. If IS-IS is used, the VFIs therefore participate in level 1 routing

within the local area, and appear to the customer network as if they are level 2 routers in the backbone.

Where an IGP is used across the entire network, it is straightforward for VPN tunnels, access connections, and backdoor links to be mixed in a network. Given that OSPF or IS-IS metrics will be assigned to all links, paths via alternate links can be compared and the shortest cost path will be used regardless of whether it is via VPN tunnels, access connections, or backdoor links. If multiple sites of a VPN do not use a common IGP, or if the backbone does not use the same common IGP as the sites, then special procedures may be needed to ensure that routes to/from other sites are treated as intra-area routes, rather than as external routes (depending upon the VPN approach taken).

Another option is to operate each site as a separate routing domain. For example each site could operate as a single OSPF area, a single IS-IS area, or a RIP domain. In this case the per-site routing domains will need to redistribute routes into a backbone routing domain (Note: in this context the "backbone routing domain" refers to a backbone as viewed by the customer network). In this case it is optional whether or not the VFIs participate in the routing within each site.

3.3.1.2. Routing for Extranets

In the extranet case the sites to be interconnected belong to multiple different administrations. In this case IGPs (such as OSPF, IS-IS, or RIP) are normally not used across the interface between organizations. Either static routes or BGP may be used between sites. If the customer network administration wishes to maintain control of routing between its site and other networks, then either

static routing or BGP may be used across the customer interface. If the customer wants to outsource all such control to the provider, then an IGP or static routes may be used at this interface.

The use of BGP between sites allows for policy based routing between sites. This is particularly useful in the extranet case. Note that private IP addresses or non-unique IP address (e.g., unregistered

addresses) should not be used for extranet communication.

3.3.1.3. CE and PE Devices for Layer 3 PE-based VPNs

When using a single IGP area across an intranet, the entire customer network participates in a single area of an IGP. In this case, for layer 3 PE-based VPNs both CE and PE devices participate as normal routers within the area.

The other options make a distinction between routing within a site, and routing between sites. In this case, a CE device would normally be considered as part of the site where it is located. However, there is an option regarding how the PE devices should be considered.

In some cases, from the perspective of routing within the customer network, a PE device (or more precisely a VFI within a PE device) may be considered to be internal to the same area or routing domain as the site to which it is attached. This simplifies the management responsibilities of the customer network administration, since inter-area routing would be handled by the provider.

For example, from the perspective of routing within the customer network, the CE devices may be the area border or AS boundary routers of the IGP area. In this case, static routing, BGP, or whatever routing is used in the backbone, may be used across the customer interface.

3.3.2. Customer View of Routing for Layer 3 Provider-Provisioned CE-based VPNs

For layer 3 provider-provisioned CE-based VPNs, the PE devices are not aware of the set of addresses which are reachable at particular customer sites. The CE and PE devices do not exchange the customer's routing information.

Customer sites that belong to the same VPN may exchange routing information through the CE-CE VPN tunnels that appear, to the customers IGP, as router adjacencies. Alternatively, instead of

exchanging routing information through the VPN tunnels, the SP's management system may take care of the configuration of the static route information of one site towards the other sites in the VPN.

Routing within the customer site may be done in any possible way, using any kind of routing protocols (see section 3.3.3).

As the CE device receives an IP or MPLS service from the SP, the CE and PE devices may exchange routing information that is meaningful within the SP routing realm.

Moreover, as the forwarding of tunneled customer packets in the SP network will be based on global IP forwarding, the routes to the various CE devices must be known in the entire SP's network.

This means that a CE device may need to participate in two different routing processes:

- routing in its own private network (VPN routing), within its own site and with the other VPN sites through the VPN tunnels, possibly using private addresses.

- routing in the SP network (global routing), as such peering with its PE.

However, in many scenarios, the use of static/default routes at the CE-PE interface might be all the global routing that is required.

3.3.3. Options for Customer Visible Routing

The following technologies are available for the exchange of routing information.

- Static routing

  Routing tables may be configured through a management system.

- RIP (Routing Information Protocol) [RFC2453]

  RIP is an interior gateway protocol and is used within an autonomous system. It sends out routing updates at regular intervals and whenever the network topology changes. Routing information is then propagated by the adjacent routers to their neighbors and thus to the entire network. A route from a source to a destination is the path with the least number of routers. This number is called the "hop count" and its maximum value is 15. This implies that RIP is suitable for a small- or medium-sized networks.

o OSPF (Open Shortest Path First) [RFC2328]

OSPF is an interior gateway protocol and is applied to a single autonomous system. Each router distributes the state of its interfaces and neighboring routers as a link state advertisement, and maintains a database describing the autonomous system's topology. A link state is advertised every 30 minutes or when the topology is reconfigured.

Each router maintains an identical topological database, from which it constructs a tree of shortest paths with itself as the root. The algorithm is known as the Shortest Path First or SPF. The router generates a routing table from the tree of shortest paths. OSPF supports a variable length subnet mask, which enables effective use of the IP address space.

OSPF allows sets of networks to be grouped together into an area. Each area has its own topological database. The topology of the area is invisible from outside its area. The areas are interconnected via a "backbone" network. The backbone network distributes routing information between the areas. The area routing scheme can reduce the routing traffic and compute the shortest path trees and is indispensable for larger scale networks.

Each multi-access network with multiple routers attached has a designated router. The designated router generates a link state advertisement for the multi-access network and synchronizes the topological database with other adjacent routers in the area. The concept of designated router can thus reduce the routing traffic and compute shortest path trees. To achieve high availability, a backup designated router is used.

o IS-IS (intermediate system to intermediate system) [RFC1195]

IS-IS is a routing protocol designed for the OSI (Open Systems Interconnection) protocol suites. Integrated IS-IS is derived from IS-IS in order to support the IP protocol. In the Internet community, IS-IS means integrated IS-IS. In this, a link state is advertised over a connectionless network service. IS-IS has the same basic features as OSPF. They include: link state advertisement and maintenance of a topological database within an area, calculation of a tree of shortest paths, generation of a routing table from a tree of shortest paths, the area routing

scheme, a designated router, and a variable length subnet mask.

o BGP-4 (Border Gateway Protocol version 4) [RFC1771]

BGP-4 is an exterior gateway protocol and is applied to the routing of inter-autonomous systems. A BGP speaker establishes a session with other BGP speakers and advertises routing information to them. A session may be an External BGP (EBGP) that connects two BGP speakers within different autonomous systems, or an internal BGP (IBGP) that connects two BGP speakers within a single autonomous system. Routing information is qualified with path attributes, which differentiate routes for the purpose of selecting an appropriate one from possible routes. Also, routes are grouped by the community attribute [RFC1997] [BGP-COM].

The IBGP mesh size tends to increase dramatically with the number of BGP speakers in an autonomous system. BGP can reduce the number of IBGP sessions by dividing the autonomous system into smaller autonomous systems and grouping them into a single confederation [RFC3065]. Route reflection is another way to reduce the number of IBGP sessions [RFC1966]. BGP divides the autonomous system into clusters. Each cluster establishes the IBGP full mesh within itself, and designates one or more BGP speakers as "route reflectors," which communicate with other clusters via their route reflectors. Route reflectors in each cluster maintain path and attribute information across the autonomous system. The autonomous system still functions like a fully meshed autonomous system. On the other hand, confederations provide finer control of routing within the autonomous system by allowing for policy changes across confederation boundaries, while route reflection requires the use of identical policies.

BGP-4 has been extended to support IPv6, IPX, and others as well as IPv4 [RFC2858]. Multiprotocol BGP-4 carries routes from multiple "address families".

4. Network Interface and SP Support of VPNs

4.1. Functional Components of a VPN

The basic functional components of an implementation of a VPN are:

- A mechanism to acquire VPN membership/capability information
- A mechanism to tunnel traffic between VPN sites
- For layer 3 PE-based VPNs, a means to learn customer routes, distribute them between the PEs, and to advertise reachable destinations to customer sites.

Based on the actual implementation, these functions could be implemented on a per-VPN basis or could be accomplished via a common mechanism shared by all VPNs. For instance, a single process could handle the routing information for all the VPNs or a separate process may be created for each VPN.

Logically, the establishment of a VPN can be thought of as composed of the following three stages. In the first stage, the VPN edge devices learn of each other. In the second stage, they establish tunnels to each other. In the third stage, they exchange routing information with each other. However, not all VPN solutions need be decomposed into these three stages. For example, in some VPN solutions, tunnels are not established after learning membership information; rather, pre-existing tunnels are selected and used. Also, in some VPN solutions, the membership information and the routing information are combined.

In the membership/capability discovery stage, membership and capability information needs to be acquired to determine whether two particular VPN edge devices support any VPNs in common. This can be accomplished, for instance, by exchanging VPN identifiers of the configured VPNs at each VPN edge device. The capabilities of the VPN edge devices need to be determined, in order to be able to agree on a common mechanism for tunneling and/or routing. For instance, if site A supports both IPsec and MPLS as tunneling mechanisms and site B supports only MPLS, they can both agree to use MPLS for tunneling. In some cases the capability information may be determined implicitly, for example some SPs may implement a single VPN solution. Likewise, the routing information for VPNs can be distributed using

the methods discussed in section 4.4.

In the tunnel establishment stage, mechanisms may need to be invoked to actually set up the tunnels. With IPsec, for instance, this could involve the use of IKE to exchange keys and policies for securing the data traffic. However, if IP tunneling, e.g., is used, there may not be any need to explicitly set up tunnels; if MPLS tunnels are used, they may be pre-established as part of normal MPLS functioning.

In the VPN routing stage, routing information for the VPN sites must be exchanged before data transfer between the sites can take place. Based on the VPN model, this could involve the use of static routes, IGPs such as OSPF/ISIS/RIP, or an EGP such as BGP.

VPN membership and capability information can be distributed from a central management system, using protocols such as, e.g., LDAP. Alternatively, it can be distributed manually. However, as manual configuration does not scale and is error prone, its use is discouraged. As a third alternative, VPN information can be

distributed via protocols that ensure automatic and consistent distribution of information in a timely manner, much as routing protocols do for routing information. This may suggest that the information be carried in routing protocols themselves, though only if this can be done without negatively impacting the essential routing functions.

It can be seen that quite a lot of information needs to be exchanged in order to establish and maintain a VPN. The scaling and stability consequences need to be analyzed for any VPN approach.

While every VPN solution must address the functionality of all three components, the combinations of mechanisms used to provide the needed functionality, and the order in which different pieces of functionality are carried out, may differ.

For layer 3 provider-provisioned CE-based VPNs, the VPN service is offering tunnels between CE devices. IP routing for the VPN is done by the customer network. With these solutions, the SP is involved in the operation of the membership/capability discovery stage and the tunnel establishment stage. The IP routing functional component may be entirely up to the customer network, or alternatively, the SP's

management system may be responsible for the distribution of the reachability information of the VPN sites to the other sites of the same VPN.

4.2. VPN Establishment and Maintenance

For a layer 3 provider-provisioned VPN the SP is responsible for the establishment and maintenance of the VPNs. Many different approaches and schemes are possible in order to provide layer 3 PPVPNs, however there are some generic problems that any VPN solution must address, including:

- For PE-based VPNs, when a new site is added to a PE, how do the other PEs find out about it? When a PE first gets attached to a given VPN, how does it determine which other PEs are attached to the same VPN. For CE-based VPNs, when a new site is added, how does its CE find out about all the other CEs at other sites of the same VPN?

- In order for layer 3 PE-based VPNs to scale, all routes for all VPNs cannot reside on all PEs. How is the distribution of VPN routing information constrained so that it is distributed to only those devices that need it?

- An administrator may wish to provision different topologies for different VPNs (e.g., a full mesh or a hub & spoke topology). How is this achieved?

  This section looks at some of these generic problems and at some of the mechanisms that can be used to solve them.

4.2.1. VPN Discovery

Mechanisms are needed to acquire information that allows the establishment and maintenance of VPNs. This may include, for example, information on VPN membership, topology, and VPN device capabilities. This information may be statically configured, or distributed by an automated protocol. As a result of the operation of these mechanisms and protocols, a device is able to determine

where to set up tunnels, and where to advertise the VPN routes for each VPN.

With a physical network, the equivalent problem can by solved by the control of the physical interconnection of links, and by having a router run a discovery/hello protocol over its locally connected links. With VPNs both the routers and the links (tunnels) may be logical entities, and thus some other mechanisms are needed.

A number of different approaches are possible for VPN discovery. One scheme uses the network management system to configure and provision the VPN edge devices. This approach can also be used to distribute VPN discovery information, either using proprietary protocols or using standard management protocols and MIBs. Another approach is where the VPN edge devices act as clients of a centralized directory or database server that contains VPN discovery information. Another possibility is where VPN discovery information is piggybacked onto a routing protocol running between the VPN edge devices [VPN-DISC].

4.2.1.1. Network Management for Membership Information

SPs use network management extensively to configure and monitor the various devices that are spread throughout their networks. This approach could be also used for distributing VPN related information. A network management system (either centralized or distributed) could be used by the SP to configure and provision VPNs on the VPN edge devices at various locations. VPN configuration information could be entered into a network management application and distributed to the remote sites via the same means used to distribute other network management information. This approach is most natural when all the devices that must be provisioned are within a single SP's network,

since the SP has access to all VPN edge devices in its domain. Security and access control are important, and could be achieved for example using SNMPv3, SSH, or IPsec tunnels.

4.2.1.2. Directory Servers

An SP typically needs to maintain a database of VPN configuration/membership information, regardless of the mechanisms

used to distribute it. LDAPv3 [RFC3377] is a standard directory protocol which makes it possible to use a common mechanism for both storing such information and distributing it.

To facilitate interoperability between different implementations, as well as between the management systems of different SPs, a standard schema for representing VPN membership and configuration information would have to be developed.

LDAPv3 supports authentication of messages and associated access control, which can be used to limit access to VPN information to authorized entities.

4.2.1.3. Augmented Routing for Membership Information

Extensions to the use of existing BGP mechanisms, for distribution of VPN membership information, are proposed in [VPN-2547BIS]. In that scheme, BGP is used to distribute VPN routes, and each route carries a set of attributes which indicate the VPN (or VPNs) to which the route belongs. This allows the VPN discovery information and routing information to be combined in a single protocol. Information needed to establish per-VPN tunnels can also be carried as attributes of the routes. This makes use of the BGP protocol's ability to effectively carry large amounts of routing information.

It is also possible to use BGP to distribute just the membership/capability information, while using a different technique to distribute the routing. BGP's update message would be used to indicate that a PE is attached to a particular VPN; BGP's withdraw message would be used to indicate that a PE has ceased to be attached to a particular VPN. This makes use of the BGP protocol's ability to dynamically distribute real-time changes in a reliable and fairly rapid manner. In addition, if a BGP route reflector is used, PEs never have to be provisioned with each other's IP addresses at all. Both cases make use of BGP's mechanisms, such as route filters, for constraining the distribution of information.

Augmented routing may be done in combination with aggregated routing, as discussed in section 4.4.4. Of course, when using BGP for distributing any kind of VPN-specific information, one must ensure

that one is not disrupting the classical use of BGP for distributing

public Internet routing information. For further discussion of this, see the discussion of aggregated routing, section 4.4.4.

4.2.1.4. VPN Discovery for Inter-SP VPNs

When two sites of a VPN are connected to different SP networks, the SPs must support a common mechanism for exchanging membership/capability information. This might make use of manual configuration or automated exchange of information between the SPs. Automated exchange may be facilitated if one or more mechanisms for VPN discovery are standardized and supported across the multiple SPs. Inter-SP trust relationships will need to be established, for example to determine which information and how much information about the VPNs may be exchanged between SPs.

In some cases different service providers may deploy different approaches for VPN discovery. Where this occurs, this implies that for multi-SP VPNs, some manual coordination and configuration may be necessary.

The amount of information which needs to be shared between SPs may vary greatly depending upon the number of size of the multi-SP VPNs. The SPs will therefore need to determine and agree upon the expected amount of membership information to be exchanged, and the dynamic nature of this information. Mechanisms may also be needed to authenticate the VPN membership information.

VPN information should be distributed only to places where it needs to go, whether that is intra-provider or inter-provider. In this way, the distribution of VPN information is unlike the distribution of inter-provider routing information, as the latter needs to be distributed throughout the Internet. In addition, the joint support of a VPN by two SPs should not require any third SP to maintain state for that VPN. Again, notice the difference with respect to inter-provider routing; in inter-provider routing: sending traffic from one SP to another may indeed require routing state in a third SP.

As one possible example: Suppose that there are two SPs A and C, which want to support a common VPN. Suppose that A and C are interconnected via SP B. In this case B will need to know how to route traffic between A and C, and therefore will need to know something about A and C (such as enough routing information to forward IP traffic and/or connect MPLS LSPs between PEs or route reflectors in A and C). However, for scaling purposes it is desirable that B not need to know VPN-specific information about the VPNs which are supported by A and C.

4.2.2. Constraining Distribution of VPN Routing Information

In layer 3 provider-provisioned CE-based VPNs, the VPN tunnels connect CE devices. In this case, distribution of IP routing information occurs between CE devices on the customer sites. No additional constraints on the distribution of VPN routing information are necessary.

In layer 3 PE-based VPNs, however, the PE devices must be aware of VPN routing information (for the VPNs to which they are attached). For scalability reasons, one does not want a scheme in which all PEs contain all routes for all VPNs. Rather, only the PEs that are attached to sites in a given VPN should contain the routing information for that VPN. This means that the distribution of VPN routing information between PE devices must be constrained.

As VPN membership may change dynamically, it is necessary to have a mechanism that allows VPN route information to be distributed to any PE where there is an attached user for that VPN, and allows for the removal of this information when it is no longer needed.

In the Virtual Router scheme, per-VPN tunnels must be established before any routes for a VPN are distributed, and the routes are then distributed through those tunnels. Thus by establishing the proper set of tunnels, one implicitly constrains and controls the distribution of per-VPN routing information. In this scheme, the distribution of membership information consists of the set of VPNs that exists on each PE, as well as information about the desired topology. This enables a PE to determine the set of remote PEs to which it must establish tunnels for a particular VPN.

In the aggregated routing scheme (see section 4.4.4), the distribution of VPN routing information is constrained by means of route filtering. As VPN membership changes on a PE, the route filters in use between the PE and its peers can be adjusted. Each peer may then adjust the filters in use with each of its peers in turn, and thus the changes propagate across the network. When BGP is used, this filtering may take place at route reflectors as discussed in section 4.4.4.

4.2.3. Controlling VPN Topology

The topology for a VPN consists of a set of nodes interconnected via tunnels. The topology may be a full mesh, a hub and spoke topology, or an arbitrary topology. For a VPN the set of nodes will include all VPN edge devices that have attached sites for that VPN.

Naturally, whatever the topology, all VPN sites are reachable from each other; the topology simply constrains the way traffic is routed

among the sites. For example, in one topology traffic between site A and site B goes from one to the other directly over the VPN backbone; in another topology, traffic from site A to site B must traverse site C before reaching site B.

The simplest topology is a full mesh, where a tunnel exists between every pair of VPN edge devices. If we assume the use of point-to-point tunnels (rather than multipoint-to-point), then with a full mesh topology there are N*(N-1)/2 duplex tunnels or N*(N-1) simplex tunnels for N VPN edge devices. Each tunnel consumes some resources at a VPN edge device, and depending on the type of tunnel, may or may not consume resources in intermediate routers or LSRs. One reason for using a partial mesh topology is to reduce the number of tunnels a VPN edge device, and/or the network, needs to support. Another reason is to support the scenario where an administrator requires all traffic from certain sites to traverse some particular site for policy or control reasons, such as to force traffic through a firewall, or for monitoring or accounting purposes. Note that the topologies used for each VPN are separate, and thus the same VPN edge device may be part of a full mesh topology for one VPN, and of a partial mesh topology for another VPN.

An example of where a partial mesh topology could be suitable is for a VPN that supports a large number of telecommuters and a small number of corporate sites. Most traffic will be between telecommuters and the corporate sites, not between pairs of telecommuters. A hub and spoke topology for the VPN would thus map onto the underlying traffic flow, with the telecommuters attached to spoke VPN edge devices and the corporate sites attached to hub VPN edge devices. Traffic between telecommuters is still supported, but this traffic traverses a hub VPN edge device.

The selection of a topology for a VPN is an administrative choice, but it is useful to examine protocol mechanisms that can be used to automate the construction of the desired topology, and thus reduce the amount of configuration needed. To this end it is useful for a VPN edge device to be able to advertise per-VPN topology information to other VPN edge devices. It may be simplest to advertise this at the same time as the membership information is advertised, using the

same mechanisms.

A simple scheme is where a VPN edge device advertises itself either as a hub or as a spoke, for each VPN that it has. When received by other VPN edge devices this information can be used when determining whether to establish a tunnel. A more comprehensive scheme allows a VPN edge device to advertise a set of topology groups, with tunnels established between a pair of VPN edge devices if they have a group in common.

4.3. VPN Tunneling

VPN solutions use tunneling in order to transport VPN packets across the VPN backbone, from one VPN edge device to another. There are different types of tunneling protocols, different ways of establishing and maintaining tunnels, and different ways to associate tunnels with VPNs (e.g., shared versus dedicated per-VPN tunnels). Sections 4.3.1 through 4.3.5 discusses some common characteristics shared by all forms of tunneling, and some common problems to which tunnels provide a solution. Section 4.3.6 provides a survey of available tunneling techniques. Note that tunneling protocol issues are generally independent of the mechanisms used for VPN membership and VPN routing.

One motivation for the use of tunneling is that the packet addressing used in a VPN may have no relation to the packet addressing used between the VPN edge devices. For example the customer VPN traffic could use non-unique or private IP addressing [RFC1918]. Also an IPv6 VPN could be implemented across an IPv4 provider backbone. As such the packet forwarding between the VPN edge devices must use information other than that contained in the VPN packets themselves. A tunneling protocol adds additional information, such an extra header or label, to a VPN packet, and this additional information is then used for forwarding the packet between the VPN edge devices.

Another capability optionally provided by tunneling is that of isolation between different VPN traffic flows. The QoS and security requirements for these traffic flows may differ, and can be met by using different tunnels with the appropriate characteristics. This allows a provider to offer different service characteristics for traffic in different VPNs, or to subsets of traffic flows within a single VPN.

The specific tunneling protocols considered in this section are GRE, IP-in-IP, IPsec, and MPLS, as these are the most suitable for carrying VPN traffic across the VPN backbone. Other tunneling protocols, such as L2TP [RFC2661], may be used as access tunnels, carrying traffic between a PE and a CE. As backbone tunneling is independent of and orthogonal to access tunneling, protocols for the latter are not discussed here.

4.3.1. Tunnel Encapsulations

All tunneling protocols use an encapsulation that adds additional information to the encapsulated packet; this information is used for forwarding across the VPN backbone. Examples are provided in section 4.3.6.

One characteristic of a tunneling protocol is whether per-tunnel state is needed in the SP network in order to forward the encapsulated packets. For IP tunneling schemes (GRE, IP-in-IP, and IPsec) per-tunnel state is completely confined to the VPN edge devices. Other routers are unaware of the tunnels, and forward according to the IP header. For MPLS, per-tunnel state is needed, since the top label in the label stack must be examined and swapped by intermediate LSRs. The amount of state required can be minimized by hierarchical multiplexing, and by use of multi-point to point tunnels, as discussed below.

Another characteristic is the tunneling overhead introduced. With IPsec the overhead may be considerable as it may include, for example, an ESP header, ESP trailer and an additional IP header. The other mechanisms listed use less overhead, with MPLS being the most lightweight. The overhead inherent in any tunneling mechanism may result in additional IP packet fragmentation, if the resulting packet is too large to be carried by the underlying link layer. As such it is important to report any reduced MTU sizes via mechanisms such as path MTU discovery in order to avoid fragmentation wherever possible.

Yet another characteristic is something we might call "transparency to the Internet". IP-based encapsulation can carry be used to carry a packet anywhere in the Internet. MPLS encapsulation can only be used to carry a packet on IP networks that support MPLS. If an

MPLS-encapsulated packet must cross the networks of multiple SPs, the adjacent SPs must bilateral agreements to accept MPLS packets from each other. If only a portion of the path across the backbone lacks MPLS support, then an MPLS-in-IP encapsulation can be used to move the MPLS packets across that part of the backbone. However, this does add complexity. On the other hand, MPLS has efficiency advantages, particularly in environments where encapsulations may need to be nested.

Transparency to the Internet is sometimes a requirement, but sometimes not. This depends on the sort of service which a SP is offering to its customer.

4.3.2. Tunnel Multiplexing

When a tunneled packet arrives at the tunnel egress, it must be possible to infer the packet's VPN from its encapsulation header. In MPLS encapsulations, this must be inferred from the packet's label stack. In IP-based encapsulations, this can be inferred from some combination of the IP source address, the IP destination address, and a "multiplexing field" in the encapsulation header. The multiplexing

field might be one which was explicitly designed for multiplexing, or one that wasn't originally designed for this but can be pushed into service as a multiplexing field. For example:

- GRE: Packets associated to VPN by source IP address, destination IP address, and Key field, although the key field was originally intended for authentication.
- IP-in-IP: Packets associated to VPN by IP destination address in outer header.
- IPsec: Packets associated to VPN by IP source address, IP destination address, and SPI field.
- MPLS: Packets associated to VPN by label stack.

Note that IP-in-IP tunneling does not have a real multiplexing field, so a different IP destination address must be used for every VPN

supported by a given PE. In the other IP-based encapsulations, a given PE need have only a single IP address, and the multiplexing field is used to distinguish the different VPNs supported by a PE. Thus the IP-in-IP solution has the significant disadvantage that it requires the allocation and assignment of a potentially large number of IP addresses, all of which have to be reachable via backbone routing.

In the following, we will use the term "multiplexing field" to refer to whichever field in the encapsulation header must is used to distinguish different VPNs at a given PE. In the IP-in-IP encapsulation, this is the destination IP address field, in the other encapsulations it is a true multiplexing field.

4.3.3. Tunnel Establishment

When tunnels are established, the tunnel endpoints must agree on the multiplexing field values which are to be used to indicate that particular packets are in particular VPNs. The use of "well known" or explicitly provisioned values would not scale well as the number of VPNs increases. So it is necessary to have some sort of protocol interaction in which the tunnel endpoints agree on the multiplexing field values.

For some tunneling protocols, setting up a tunnel requires an explicit exchange of signaling messages. Generally the multiplexing field values would be agreed upon as part of this exchange. For example, if an IPsec encapsulation is used, the SPI field plays the role of the multiplexing field, and IKE signaling is used to distribute the SPI values; if an MPLS encapsulation is used, LDP,

CR-LDP or RSVP-TE can be used to distribute the MPLS label value used as the multiplexing field. Information about the identity of the VPN with which the tunnel is to be associated needs to be exchanged as part of the signaling protocol (e.g., a VPN-ID can be carried in the signaling protocol). An advantage of this approach is that per-tunnel security, QoS and other characteristics may also be negotiable via the signaling protocol. A disadvantage is that the signaling imposes overhead, which may then lead to scalability considerations, discussed further below.

For some tunneling protocols, there is no explicit protocol

interaction that sets up the tunnel, and the multiplexing field values must be exchanged in some other way. For example, for MPLS tunnels, MPLS labels can be piggybacked on the protocols used to distribute VPN routes or VPN membership information. GRE and IP-in-IP have no associated signaling protocol, and thus by necessity the multiplexing values are distributed via some other mechanism, such as via configuration, control protocol, or piggybacked in some manner on a VPN membership protocol.

The resources used by the different tunneling establishment mechanisms may vary. With a full mesh VPN topology, and explicit signaling, each VPN edge device has to establish a tunnel to all the other VPN edge devices for in each VPN. The resources needed for this on a VPN edge device may be significant, and issues such as the time needed to recover following a device failure may need to be taken into account, as the time to recovery includes the time needed to reestablish a large number of tunnels.

4.3.4. Scaling and Hierarchical Tunnels

If tunnels require state to be maintained in the core of the network, it may not be feasible to set up per-VPN tunnels between all adjacent devices that are adjacent in some VPN topology. This would violate the principle that there is no per-VPN state in the core of the network, and would make the core scale poorly as the number of VPNs increases. For example, MPLS tunnels require that core network devices maintain state for the topmost label in the label stack. If every core router had to maintain one or more labels for every VPN, scaling would be very poor.

There are also scaling considerations related to the use of explicit signaling for tunnel establishment. Even if the tunneling protocol does not maintain per tunnel state in the core, the number of tunnels that a single VPN edge device needs to handle may be large, as this grows according to the number of VPNs and the number of neighbors per VPN. One way to reduce the number of tunnels in a network is to use

a VPN topology other than a full mesh. However this may not always be desirable, and even with hub and spoke topologies the hubs VPN edge devices may still need to handle large numbers of tunnels.

If the core routers need to maintain any per-tunnel state at all, scaling can be greatly improved by using hierarchical tunnels. One tunnel can be established between each pair of VPN edge devices, and multiple VPN-specific tunnels can then be carried through the single "outer" tunnel. Now the amount of state is dependent only on the number of VPN edge devices, not on the number of VPNs. Scaling can be further improved by having the outer tunnels be multipoint-to-point "merging" tunnels. Now the amount of state to be maintained in the core is on the order of the number of VPN edge devices, not on the order of the square of that number. That is, the amount of tunnel state is roughly equivalent to the amount of state needed to maintain IP routes to the VPN edge devices. This is almost (if not quite) as good as using tunnels which do not require any state to be maintained in the core.

Using hierarchical tunnels may also reduce the amount of state to be maintained in the VPN edge devices, particularly if maintaining the outer tunnels requires more state than maintaining the per-VPN tunnels that run inside the outer tunnels.

There are other factors relevant to determining the number of VPN edge to VPN edge "outer" tunnels to use. While using a single such tunnel has the best scaling properties, using more than one may allow different QoS capabilities or different security characteristics to be used for different traffic flows (from the same or from different VPNs).

When tunnels are used hierarchically, the tunnels in the hierarchy may all be of the same type (e.g., an MPLS label stack) or they may be of different types (e.g., a GRE tunnel carried inside an IPsec tunnel).

One example using hierarchical tunnels is the establishment of a number of different IPsec security associations, providing different levels of security between a given pair of VPN edge devices. Per-VPN GRE tunnels can then be grouped together and then carried over the appropriate IPsec tunnel, rather than having a separate IPsec tunnel per-VPN. Another example is the use of an MPLS label stack. A single PE-PE LSP is used to carry all the per-VPN LSPs. The mechanisms used for label establishment are typically different. The PE-PE LSP could be established using LDP, as part or normal backbone operation, with the per-VPN LSP labels established by piggybacking on VPN routing (e.g., using BGP) discussed in sections 3.3.1.3 and 4.1.

4.3.5. Tunnel Maintenance

Once a tunnel is established it is necessary to know that the tunnel is operational. Mechanisms are needed to detect tunnel failures, and to respond appropriately to restore service.

There is a potential issue regarding propagation of failures when multiple tunnels are multiplexed hierarchically. Suppose that multiple VPN-specific tunnels are multiplexed inside a single PE to PE tunnel. In this case, suppose that routing for the VPN is done over the VPN-specific tunnels (as may be the case for CE-based and VR approaches). Suppose that the PE to PE tunnel fails. In this case multiple VPN-specific tunnels may fail, and layer 3 routing may simultaneously respond for each VPN using the failed tunnel. If the PE to PE tunnel is subsequently restored, there may then be multiple VPN-specific tunnels and multiple routing protocol instances which also need to recover. Each of these could potentially require some exchange of control traffic.

When a tunnel fails, if the tunnel can be restored quickly, it might therefore be preferable to restore the tunnel without any response by high levels (such as other tunnels which were multiplexed inside the failed tunnels). By having high levels delay response to a lower level failed tunnel, this may limit the amount of control traffic needed to completely restore correct service. However, if the failed tunnel cannot be quickly restored, then it is necessary for the tunnels or routing instances multiplexed over the failed tunnel to respond, and preferable for them to respond quickly and without explicit action by network operators.

With most layer 3 provider-provisioned CE-based VPNs and the VR scheme, a per-VPN instance of routing is running over the tunnel, thus any loss of connectivity between the tunnel endpoints will be detected by the VPN routing instance. This allows rapid detection of tunnel failure. Careful adjustment of timers might be needed to avoid failure propagation as discussed the above. With the aggregated routing scheme, there isn't a per-VPN instance of routing running over the tunnel, and therefore some other scheme to detect loss of connectivity is needed in the event that the tunnel cannot be rapidly restored.

Failure of connectivity in a tunnel can be very difficult to detect reliably. Among the mechanisms that can be used to detect failure are loss of the underlying connectivity to the remote endpoint (as indicated, e.g., by "no IP route to host" or no MPLS label), timeout of higher layer "hello" mechanisms (e.g., IGP hellos, when the tunnel is an adjacency in some IGP), and timeout of keep alive mechanisms in

the tunnel establishment protocols (if any). However, none of these techniques provides completely reliable detection of all failure modes. Additional monitoring techniques may also be necessary.

With hierarchical tunnels it may suffice to only monitor the outermost tunnel for loss of connectivity. However there may be failure modes in a device where the outermost tunnel is up but one of the inner tunnels is down.

4.3.6. Survey of Tunneling Techniques

Tunneling mechanisms provide isolated communication between two CE-PE devices. Available tunneling mechanisms include (but are not limited to): GRE [RFC2784] [RFC2890], IP-in-IP encapsulation [RFC2003] [RFC2473], IPsec [RFC2401] [RFC2402], and MPLS [RFC3031] [RFC3035].

Note that the following subsections address tunnel overhead to clarify the risk of fragmentation. Some SP networks contain layer 2 switches that enforce the standard/default MTU of 1500 bytes. In this case, any encapsulation whatsoever creates a significant risk of fragmentation. However, layer 2 switch vendors are in general aware of IP tunneling as well as stacked VLAN overhead, thus many switches practically allow an MTU of approximately 1512 bytes now. In this case, up to 12 bytes of encapsulation can be used before there is any risk of fragmentation. Furthermore, to improve TCP and NFS performance, switches that support 9K bytes "jumbo frames" are also on the market. In this case, there is no risk of fragmentation.

4.3.6.1. GRE [RFC2784] [RFC2890]

Generic Routing Encapsulation (GRE) specifies a protocol for encapsulating an arbitrary payload protocol over an arbitrary delivery protocol [RFC2784]. In particular, it can be used where both the payload and the delivery protocol are IP as is the case in layer 3 VPNs. A GRE tunnel is a tunnel whose packets are encapsulated by GRE.

o Multiplexing

The GRE specification [RFC2784] does not explicitly support

multiplexing. But the key field extension to GRE is specified in [RFC2890] and it may be used as a multiplexing field.

- o QoS/SLA

  GRE itself does not have intrinsic QoS/SLA capabilities, but it inherits whatever capabilities exist in the delivery protocol (IP). Additional mechanisms, such as Diffserv or RSVP extensions [RFC2746], can be applied.

- o Tunnel setup and maintenance

  There is no standard signaling protocol for setting up and maintaining GRE tunnels.

- o Large MTUs and minimization of tunnel overhead

  When GRE encapsulation is used, the resulting packet consists of a delivery protocol header, followed by a GRE header, followed by the payload packet. When the delivery protocol is IPv4, and if the key field is not present, GRE encapsulation adds at least 28 bytes of overhead (36 bytes if key field extension is used.)

- o Security

  GRE encapsulation does not provide any significant security. The optional key field can be used as a clear text password to aid in the detection of misconfigurations, but it does not provide integrity or authentication. An SP network which supports VPNs must do extensive IP address filtering at its borders to prevent spoofed packets from penetrating the VPNs. If multi-provider VPNs are being supported, it may be difficult to set up these filters.

4.3.6.2. IP-in-IP Encapsulation [RFC2003] [RFC2473]

IP-in-IP specifies the format and procedures for IP-in-IP

encapsulation. This allows an IP datagram to be encapsulated within another IP datagram. That is, the resulting packet consists of an outer IP header, followed immediately by the payload packet. There is no intermediate header as in GRE. [RFC2003] and [RFC2473] specify IPv4 and IPv6 encapsulations respectively. Once the encapsulated datagram arrives at the intermediate destination (as specified in the outer IP header), it is decapsulated, yielding the original IP datagram, which is then delivered to the destination indicated by the original destination address field.

- Multiplexing

  The IP-in-IP specifications don't explicitly support multiplexing. But if a different IP address is used for every VPN then the IP address field can be used for this purpose. (See section 4.3.2 for detail).

- QoS/SLA

  IP-in-IP itself does not have intrinsic QoS/SLA capabilities, but of course it inherits whatever capabilities exist for IP. Additional mechanisms, such as RSVP extensions [RFC2764] or DiffServ extensions [RFC2983], may be used with it.

- Tunnel setup and maintenance

  There is no standard setup and maintenance protocol for IP-in-IP.

- Large MTUs and minimization of tunnel overhead

  When the delivery protocol is IPv4, IP-in-IP adds at least 20 bytes of overhead.

- Security

  IP-in-IP encapsulation does not provide any significant security.

An SP network which supports VPNs must do extensive IP address filtering at its borders to prevent spoofed packets from penetrating the VPNs. If multi-provider VPNs are being supported, it may be difficult to set up these filters.

4.3.6.3. IPsec [RFC2401] [RFC2402] [RFC2406] [RFC2409]

IP Security (IPsec) provides security services at the IP layer [RFC2401]. It comprises authentication header (AH) protocol [RFC2402], encapsulating security payload (ESP) protocol [RFC2406], and Internet key exchange (IKE) protocol [RFC2409]. AH protocol provides data integrity, data origin authentication, and an anti-replay service. ESP protocol provides data confidentiality and limited traffic flow confidentiality. It may also provide data integrity, data origin authentication, and an anti-replay service. AH and ESP may be used in combination.

IPsec may be employed in either transport or tunnel mode. In transport mode, either an AH or ESP header is inserted immediately after the payload packet's IP header. In tunnel mode, an IP packet is encapsulated with an outer IP packet header. Either an AH or ESP header is inserted between them. AH and ESP establish a

unidirectional secure communication path between two endpoints, which is called a security association. In tunnel mode, PE-PE tunnel (or a CE-CE tunnel) consists of a pair of unidirectional security associations. The IPsec and IKE protocols are used for setting up IPsec tunnels.

- Multiplexing

  The SPI field of AH and ESP is used to multiplex security associations (or tunnels) between two peer devices.

- QoS/SLA

  IPsec itself does not have intrinsic QoS/SLA capabilities, but it inherits whatever mechanisms exist for IP. Other mechanisms such as "RSVP Extensions for IPsec Data Flows" [RFC2207] or DiffServ extensions [RFC2983] may be used with it.

- Tunnel setup and maintenance

The IPsec and IKE protocols are used for the setup and maintenance of tunnels.

- Large MTUs and minimization of tunnel overhead

  IPsec transport mode adds at least 8 bytes of overhead. IPsec tunnel mode adds at least 28 bytes of overhead. IPsec transport mode adds minimal overhead. In PE-based PPVPNs, the processing overhead of IPsec (due to its cryptography) may limit the PE's performance, especially if privacy is being provided; this is not generally an issue in CE-based PPVPNs.

- Security

  When IPsec tunneling is used in conjunction with IPsec's cryptographic capabilities, excellent authentication and integrity functions can be provided. Privacy can also be optionally provided.

4.3.6.4. MPLS [RFC3031] [RFC3032] [RFC3035]

Multiprotocol Label Switching (MPLS) is a method for forwarding packets through a network. Routers at the edge of a network apply simple labels to packets. A label may be inserted between the data link and network headers, or may be carried in the data link header (e.g., the VPI/VCI field in an ATM header). Routers in the network

switch packets according to the labels, with minimal lookup overhead. A path, or a tunnel in the PPVPN, is called a "label switched path (LSP)".

- Multiplexing

  LSPs may be multiplexed within other LSPs.

- QoS/SLA

  MPLS does not have intrinsic QoS or SLA management mechanisms, but bandwidth may be allocated to LSPs, and their routing may be

explicitly controlled. Additional techniques such as DiffServ and DiffServ aware traffic engineering may be used with it [RFC3270] [MPLS-DIFF-TE]. QoS capabilities from IP may be inherited.

- Tunnel setup and maintenance

  LSPs are set up and maintained by LDP (Label Distribution Protocol), RSVP (Resource Reservation Protocol) [RFC3209], or BGP.

- Large MTUs and minimization of tunnel overhead.

  MPLS encapsulation adds four bytes per label. VPN-2547BIS's [VPN-2547BIS] approach uses at least two labels for encapsulation and adds minimal overhead.

- Encapsulation

  MPLS packets may optionally be encapsulated in IP or GRE, for cases where it is desirable to carry MPLS packets over an IP-only infrastructure.

- Security

  MPLS encapsulation does not provide any significant security. An SP which is providing VPN service can refuse to accept MPLS packets from outside its borders. This provides the same level of assurance as would be obtained via IP address filtering when IP-based encapsulations are used. If a VPN is jointly provided by multiple SPs, care should be taken to ensure that a labeled packet is accepted from a neighboring router in another SP only if its top label is one which was actually distributed to that router.

- Applicability

  MPLS is the only one of the encapsulation techniques that cannot be guaranteed to run over any IP network. Hence it would not be applicable when transparency to the Internet is a requirement.

If the VPN backbone consists of several cooperating SP networks which support MPLS, then the adjacent networks may support MPLS at their interconnects. If two cooperating SP networks which support MPLS are separated by a third which does not support MPLS, then MPLS-in-IP or MPLS-in-IPsec tunneling may be done between them.

4.4. PE-PE Distribution of VPN Routing Information

In layer 3 PE-based VPNs, PE devices examine the IP headers of packets they receive from the customer networks. Forwarding is based on routing information received from the customer network. This implies that the PE devices need to participate in some manner in routing for the customer network. Section 3.3 discussed how routing would be done in the customer network, including the customer interface. In this section, we discuss ways in which the routing information from a particular VPN may be passed, over the shared VPN backbone, among the set of PEs attaching to that VPN.

The PEs needs to distribute two types of routing information to each other: (i) Public Routing: routing information which specifies how to reach addresses on the VPN backbone (i.e., "public addresses"); call this "public routing information" (ii) VPN Routing: routing information obtained from the CEs, which specifies how to reach addresses ("private addresses") that are in the VPNs.

The way in which routing information in the first category is distributed is outside the scope of this document; we discuss only the distribution of routing information in the second category. Of course, one of the requirements for distributing VPN routing information is that it be kept separate and distinct from the public information. Another requirement is that the distribution of VPN routing information not destabilize or otherwise interfere with the distribution of public routing information.

Similarly, distribution of VPN routing information associated with one VPN should not destabilize or otherwise interfere with the operation of other VPNs. These requirements are, for example, relevant in the case that a private network might be suffering from instability or other problems with its internal routing, which might be propagated to the VPN used to support that private network.

Note that this issue does not arise in CE-based VPNs, as in CE-based VPNs, the PE devices do not see packets from the VPN until after the packets haven been encapsulated in an outer header that has only public addresses.

4.4.1. Options for VPN Routing in the SP

The following technologies can be used for exchanging VPN routing information discussed in sections 3.3.1.3 and 4.1.

- Static routing
- RIP [RFC2453]
- OSPF [RFC2328]
- BGP-4 [RFC1771]

4.4.2. VPN Forwarding Instances (VFIs)

In layer 3 PE-based VPNs, the PE devices receive unencapsulated IP packets from the CE devices, and the PE devices use the IP destination addresses in these packets to help make their forwarding decisions. In order to do this properly, the PE devices must obtain routing information from the customer networks. This implies that the PE device participates in some manner in the customer network's routing.

In layer 3 PE-based VPNs, a single PE device connected to several CE devices that are in the same VPN, and it may also be connected to CE devices of different VPNs. The route which the PE chooses for a given IP destination address in a given packet will depend on the VPN from which the packet was received. A PE device must therefore have a separate forwarding table for each VPN to which it is attached. We refer to these forwarding tables as "VPN Forwarding Instances" (VFIs), as defined in section 2.1.

A VFI contains routes to locally attached VPN sites, as well as routes to remote VPN sites. Section 4.4 discusses the way in which routes to remote sites are obtained.

Routes to local sites may be obtained in several ways. One way is to explicitly configure static routes into the VFI. This can be useful in simple deployments, but it requires that one or more devices in the customer's network be configured with static routes (perhaps just a default route), so that traffic will be directed from the site to the PE device.

Another way is to have the PE device be a routing peer of the CE device, in a routing algorithm such as RIP, OSPF, or BGP. Depending on the deployment scenario, the PE might need to advertise a large number of routes to each CE (e.g., all the routes which the PE obtained from remote sites in the CE's VPN), or it might just need to advertise a single default route to the CE.

A PE device uses some resources in proportion to the number of VFIs that it has, particularly if a distinct dynamic routing protocol instance is associated with each VFI. A PE device also uses some resources in proportion to the total number of routes it supports, where the total number of routes includes all the routes in all its VFIs, and all the public routes. These scaling factors will limit the number of VPNs which a single PE device can support.

When dynamic routing is used between a PE and a CE, it is not necessarily the case that each VFI is associated with a single routing protocol instance. A single routing protocol instance may provide routing information for multiple VFIs, and/or multiple routing protocol instances might provide information for a single VFI. See sections 4.4.3, 4.4.4, 3.3.1, and 3.3.1.3 for details.

There are several options for how VPN routes are carried between the PEs, as discussed below.

4.4.3. Per-VPN Routing

One option is to operate separate instances of routing protocols between the PEs, one instance for each VPN. When this is done, routing protocol packets for each customer network need to be tunneled between PEs. This uses the same tunneling method, and optionally the same tunnels, as is used for transporting VPN user data traffic between PEs.

With per-VPN routing, a distinct routing instance corresponding to each VPN exists within the corresponding PE device. VPN-specific tunnels are set up between PE devices (using the control mechanisms that were discussed in sections 3 and 4). Logically these tunnels are between the VFIs which are within the PE devices. The tunnels then used as if they were normal links between normal routers. Routing protocols for each VPN operate between VFIs and the routers within the customer network.

This approach establishes, for each VPN, a distinct "control plane" operating across the VPN backbone. There is no sharing of control plane by any two VPNs, nor is there any sharing of control plane by

the VPN routing and the public routing. With this approach each PE device can logically be thought of as consisting of multiple independent routers.

The multiple routing instances within the PE device may be separate processes, or may be in the same process with different data structures. Similarly, there may be mechanisms internal to the PE devices to partition memory and other resources between routing instances. The mechanisms for implementing multiple routing instances within a single physical PE are outside of the scope of this framework document, and are also outside of the scope of other standards documents.

This approach tends to minimize the explicit interactions between different VPNs, as well as between VPN routing and public routing. However, as long as the independent logical routers share the same hardware, there is some sharing of resources, and interactions are still possible. Also, each independent control plane has its associated overheads, and this can raise issues of scale. For example, the PE device must run a potentially large number of independent routing "decision processes," and must also maintain a potentially very large number of routing adjacencies.

4.4.4. Aggregated Routing Model

Another option is to use one single instance of a routing protocol for carrying VPN routing information between the PEs. In this method, the routing information for multiple different VPNs is aggregated into a single routing protocol.

This approach greatly reduces the number of routing adjacencies which the PEs must maintain, since there is no longer any need to maintain more than one such adjacency between a given pair of PEs. If the single routing protocol supports a hierarchical route distribution mechanism (such as BGP's "route reflectors"), the PE-PE adjacencies

can be completely eliminated, and the number of backbone adjacencies can be made into a small constant which is independent of the number of PE devices. This improves the scaling properties.

Additional routing instances may still be needed to support the exchange of routing information between the PE and its locally attached CEs. These can be eliminated, with a consequent further improvement in scalability, by using static routing on the PE-CE interfaces, or possibly by having the PE-CE routing interaction use the same protocol instance that is used to distribute VPN routes across the VPN backbone (see section 4.4.4.2 for a way to do this).

With this approach, the number of routing protocol instances in a PE device does not depend on the number of CEs supported by the PE device, if the routing between PE and CE devices is static or BGP-4. However, CE and PE devices in a VPN exchange route information inside a VPN using a routing protocol except for BGP-4, the number of routing protocol entities in a PE device depends on the number of CEs supported by the PE device.

In principle it is possible for routing to be aggregated using either BGP or on an IGP.

4.4.4.1. Aggregated Routing with OSPF or IS-IS

When supporting VPNs, it is likely that there can be a large number of VPNs supported within any given SP network. In general only a small number of PE devices will be interested in the operation of any one VPN. Thus while the total amount of routing information related to the various customer networks will be very large, any one PE needs to know about only a small number of such networks.

Generally SP networks use OSPF or IS-IS for interior routing within the SP network. There are very good reasons for this choice, which are outside of the scope of this document.

Both OSPF and IS-IS are link state routing protocols. In link state routing, routing information is distributed via a flooding protocol. The set of routing peers is in general not fully meshed, but there is a path from any router in the set to any other. Flooding ensures

that routing information from any one router reaches all the others. This requires all routers in the set to maintain the same routing information. One couldn't withhold any routing information from a particular peer unless it is known that none of the peers further downstream will need that information, and in general this cannot be known.

As a result, if one tried to do aggregated routing by using OSPF, with all the PEs in the set of routing peers, all the PEs would end up with the exact same routing information; there is no way to constrain the distribution of routing information to a subset of the PEs. Given the potential magnitude of the total routing information required for supporting a large number of VPNs, this would have unfortunate scaling implications.

In some cases VPNs may span multiple areas within a provider, or span multiple providers. If VPN routing information were aggregated into the IGP used within the provider, then some method would need to be used to extend the reach of IGP routing information between areas and between SPs.

4.4.4.2. Aggregated Routing with BGP

In order to use BGP for aggregated routing, the VPN routing information must be clearly distinguished from the public Internet routing information. This is typically done by making use of BGP's capability of handling multiple address families, and treating the VPN routes as being in a different address family than the public Internet routes. Typically a VPN route also carries attributes which depend on the particular VPN or VPNs to which that route belongs.

When BGP is used for carrying VPN information, the total amount of information carried in BGP (including the Internet routes and VPN routes) may be quite large. As noted above, there may be a large number of VPNs which are supported by any particular provider, and the total amount of routing information associated with all VPNs may be quite large. However, any one PE will in general only need to be aware of a small number of VPNs. This implies that where VPN routing information is aggregated into BGP, it is desirable to be able to limit which VPN information is distributed to which PEs.

In "Interior BGP" (IBGP), routing information is not flooded; it is

sent directly, over a TCP connection, to the peer routers (or to a route reflector). These peer routers (unless they are route reflectors) are then not even allowed to redistribute the information to each other. BGP also has a comprehensive set of mechanisms for constraining the routing information that any one peer sends to another, based on policies established by the network administration. Thus IBGP satisfies one of the requirements for aggregated routing within a single SP network - it makes it possible to ensure that routing information relevant to a particular VPN is processed only by the PE devices that attach to that VPN. All that is necessary is that each VPN route be distributed with one or more attributes which identify the distribution policies. Then distribution can be constrained by filtering against these attributes.

In "Exterior BGP" (EBGP), routing peers do redistribute routing information to each other. However, it is very common to constrain the distribution of particular items of routing information so that they only go to those exterior peers who have a "need to know," although this does require a priori knowledge of which paths may validly lead to which addresses. In the case of VPN routing, if a VPN is provided by a small set of cooperating SPs, such constraints can be applied to ensure that the routing information relevant to that VPN does not get distributed anywhere it doesn't need to be. To the extent that a particular VPN is supported by a small number of cooperating SPs with private peering arrangements, this is

particularly straightforward, as the set of EBGP neighbors which need to know the routing information from a particular VPN is easier to determine.

BGP also has mechanisms (such as "Outbound Route Filtering," ORF) which enable the proper set of VPN routing distribution constraints to be dynamically distributed. This reduces the management burden of setting up the constraints, and hence improves scalability.

Within a single routing domain (in the layer 3 VPN context, this typically means within a single SP's network), it is common to have the IBGP routers peer directly with one or two route reflectors, rather than having them peer directly with each other. This greatly reduces the number of IBGP adjacencies which any one router must

support. Further, a route reflector does not merely redistribute routing information, it "digests" the information first, by running its own decision processes. Only routes which survive the decision process are redistributed.

As a result, when route reflectors are used, the amount of routing information carried around the network, and in particular, the amount of routing information which any given router must receive and process, is greatly reduced. This greatly increases the scalability of the routing distribution system.

It has already been stated that a given PE has VPN routing information only for those PEs to which it is directly attached. It is similarly important, for scalability, to ensure that no single route reflector should have to have all the routing information for all VPNs. It is after all possible for the total number of VPN routes (across all VPNs supported by an SP) to exceed the number which can be supported by a single route reflector. Therefore, the VPN routes may themselves be partitioned, with some route reflectors carrying one subset of the VPN routes and other route reflectors carrying a different subset. The route reflectors which carry the public Internet routes can also be completely separate from the route reflectors that carry the VPN routes.

The use of outbound route filters allows any one PE and any one route reflector to exchange information about only those VPNs which the PE and route reflector are both interested in. This in turn ensures that each PE and each route reflector receives routing information only about the VPNs which it is directly supporting. Large SPs which support a large number of VPNs therefore can partition the information which is required for support of those VPNs.

Generally a PE device will be restricted in the total number of routes it can support, whether those are public Internet routes or VPN routes. As a result, a PE device may be able to be attached to a larger number of VPNs if it does not also need to support Internet routes.

The way in which VPN routes are partitioned among PEs and/or route

reflectors is a deployment issue. With suitable deployment procedures, the limited capacity of these devices will not limit the number of VPNs that can be supported.

Similarly, whether a given PE and/or route reflector contains Internet routes as well as VPN routes is a deployment issue. If the customer networks served by a particular PE do not need the Internet access, then that PE does not need to be aware of the Internet routes. If some or all of the VPNs served by a particular PE do need the Internet access, but the PE does not contain Internet routes, then the PE can maintain a default route that routes all the Internet traffic from that PE to a different router within the SP network, where that other router holds the full the Internet routing table. With this approach the PE device needs only a single default route for all the Internet routes.

For the reasons given above, the BGP protocol seems to be a reasonable protocol to use for distributing VPN routing information. Additional reasons for the use of BGP are:

- BGP has been proven to be useful for distributing very large amounts of routing information; there isn't any routing distribution protocol which is known to scale any better.

- The same BGP instance that is used for PE-PE distribution of VPN routes can be used for PE-CE route distribution, if CE-PE routing is static or BGP. PEs and CEs are really parts of distinct Autonomous Systems, and BGP is particularly well-suited for carrying routing information between Autonomous Systems.

On the other hand, BGP is also used for distributing public Internet routes, and it is crucially important that VPN route distributing not compromise the distribution of public Internet routes in any way. This issue is discussed in the following section.

4.4.5. Scalability and Stability of Routing with Layer 3 PE-based VPNs

For layer 3 PE-based VPNs, there are likely to be cases where a service provider supports Internet access over the same link that is used for VPN service. Thus, a particular CE to PE link may carry both private network IP packets (for transmission between sites of the private network using VPN services) as well as public Internet traffic (for transmission from the private site to the Internet, and for transmission to the private site from the Internet). This section looks at the scalability and stability of routing in this case. It is worth noting that this sort of issue may be applicable where per-VPN routing is used, as well as where aggregated routing is used.

For layer 3 PE-based VPNs, it is necessary for the PE devices to be able to forward IP packets using the addresses spaces of the supported private networks, as well as using the full Internet address space. This implies that PE devices might in some cases participate in routing for the private networks, as well as for the public Internet.

In some cases the routing demand on the PE might be low enough, and the capabilities of the PE, might be great enough, that it is reasonable for the PE to participate fully in routing for both private networks and the public Internet. For example, the PE device might participate in normal operation of BGP as part of the global Internet. The PE device might also operate routing protocols (or in some cases use static routing) to exchange routes with CE devices.

For large installations, or where PE capabilities are more limited, it may be undesirable for the PE to fully participate in routing for both VPNs as well as the public Internet. For example, suppose that the total volume of routes and routing instances supported by one PE across multiple VPNs is very large. Suppose furthermore that one or more of the private networks suffers from routing instabilities, for example resulting in a large number of routing updates being transmitted to the PE device. In this case it is important to prevent such routing from causing any instability in the routing used in the global Internet.

In these cases it may be necessary to partition routing, so that the PE does not need to maintain as large a collection of routes, and so that the PE is not able to adversely effect Internet routing. Also, given that the total number of route prefixes and the total number of routing instances which the PE needs to maintain might be very large, it may be desirable to limit the participation in Internet routing for those PEs which are supporting a large number of VPNs or which are supporting large VPNs.

Consider a case where a PE is supporting a very large number of VPNs, some of which have a large number of sites. To pick a VERY large example, let's suppose 1000 VPNs, with an average of 100 sites each, plus 10 prefixes per site on average. Consider that the PE also needs to be able to route traffic to the Internet in general. In this example the PE might need to support approximately 1,000,000 prefixes for the VPNs, plus more than 100,000 prefixes for the Internet. If augmented and aggregated routing is used, then this implies a large number of routes which may be advertised in a single routing protocol (most likely BGP). If the VR approach is used, then there are also 100,000 neighbor adjacencies in the various per-VPN routing protocol instances. In some cases this number of routing prefixes and/or this number of adjacencies might be difficult to support in one device.

In this case, an alternate approach is to limit the PE's participation in Internet routing to the absolute minimum required: Specifically the PE will need to know which Internet address prefixes are reachable via directly attached CE devices. All other Internet routes may be summarized into a single default route pointing to one or more P routers. In many cases the P routers to which the default routes are directed may be the P routers to which the PE device is directly attached (which are the ones which it needs to use for forwarding most Internet traffic). Thus if there are M CE devices directly connected to the PE, and if these M CE devices are the next hop for a total of N globally addressable Internet address prefixes, then the PE device would maintain N+1 routes corresponding to globally routable Internet addresses.

In this example, those PE devices which provide VPN service run routing to compute routes for the VPNs, but don't operate Internet routing, and instead use only a default route to route traffic to all Internet destinations (not counting the addresses which are reachable via directly attached CE devices). The P routers need to maintain Internet routes, and therefore take part in Internet routing protocols. However, the P routers don't know anything about the VPN routes.

In some cases the maximum number of routes and/or routing instances supportable via a single PE device may limit the number of VPNs which can be supported by that PE. For example, in some cases this might require that two different PE devices be used to support VPN services for a set of multiple CEs, even if one PE might have had sufficient

throughput to handle the data traffic from the full set of CEs. Similarly, the amount of resources which any one VPN is permitted to use in a single PE might be restricted.

There will be cases where it is not necessary to partition the routing, since the PEs will be able to maintain all VPN routes and all Internet routes without a problem. However, it is important that VPN approaches allow partitioning to be used where needed in order to prevent future scaling problems. Again, making the system scalable is a matter of proper deployment.

It may be wondered whether it is ever desirable to have both Internet routing and VPN routing running in a single PE device or route reflector. In fact, if there is even a single system running both Internet routing and VPN routing, doesn't that raise the possibility that a disruption within the VPN routing system will cause a disruption within the Internet routing system?

Certainly this possibility exists in theory. To minimize that possibility, BGP implementations which support multiple address families should be organized so as to minimize the degree to which the processing and distribution of one address family affects the processing and distribution of another. This could be done, for example, by suitable partitioning of resources. This partitioning may be helpful both to protect Internet routing from VPN routing, and to protect well behaved VPN customers from "mis-behaving" VPNs. Or one could try to protect the Internet routing system from the VPN routing system by giving preference to the Internet routing. Such implementation issues are outside the scope of this document. If one has inadequate confidence in an implementation, deployment procedures can be used, as explained above, to separate the Internet routing from the VPN routing.

4.5. Quality of Service, SLAs, and IP Differentiated Services

The following technologies for QoS/SLA may be applicable to PPVPNs.

4.5.1. IntServ/RSVP [RFC2205] [RFC2208] [RFC2210] [RFC2211] [RFC2212]

Integrated services, or IntServ for short, is a mechanism for

providing QoS/SLA by admission control. RSVP is used to reserve network resources. The network needs to maintain a state for each reservation. The number of states in the network increases in proportion to the number of concurrent reservations.

In some cases, IntServ on the edge of a network (e.g., over the customer interface) may be mapped to DiffServ in the SP network.

4.5.2. DiffServ [RFC2474] [RFC2475]

IP differentiated service, or DiffServ for short, is a mechanism for providing QoS/SLA by differentiating traffic. Traffic entering a network is classified into several behavior aggregates at the network edge and each is assigned a corresponding DiffServ codepoint. Within the network, traffic is treated according to its DiffServ codepoint. Some behavior aggregates have already been defined. Expedited forwarding behavior [RFC3246] guarantees the QoS, whereas assured forwarding behavior [RFC2597] differentiates traffic packet precedence values.

When DiffServ is used, network provisioning is done on a per-traffic-class basis. This ensures a specific class of service can be achieved for a class (assuming that the traffic load is controlled). All packets within a class are then treated equally within an SP network. Policing is done at input to prevent any one user from exceeding their allocation and therefore defeating the provisioning for the class as a whole. If a user exceeds their traffic contract, then the excess packets may optionally be discarded, or may be marked as "over contract". Routers throughout the network can then preferentially discard over contract packets in response to congestion, in order to ensure that such packets do not defeat the service guarantees intended for in contract traffic.

4.6. Concurrent Access to VPNs and the Internet

In some scenarios, customers will need to concurrently have access to their VPN network and to the public Internet.

Two potential problems are identified in this scenario: the use of private addresses and the potential security threads.

o The use of private addresses

  The IP addresses used in the customer's sites will possibly belong to a private routing realm, and as such be unusable in the public Internet. This means that a network address translation function (e.g., NAT) will need to be implemented to allow VPN customers to access the Public Internet.

  In the case of layer 3 PE-based VPNs, this translation function will be implemented in the PE to which the CE device is connected. In the case of layer 3 provider-provisioned CE-based VPNs, this translation function will be implemented on the CE device itself.

o Potential security threat

  As portions of the traffic that flow to and from the public Internet are not necessarily under the SP's nor the customer's control, some traffic analyzing function (e.g., a firewall function) will be implemented to control the traffic entering and leaving the VPN.

  In the case of layer 3 PE-based VPNs, this traffic analyzing function will be implemented in the PE device (or in the VFI supporting a specific VPN), while in the case of layer 3 provider provisioned CE-based VPNs, this function will be implemented in the CE device.

o Handling of a customer IP packet destined for the Internet

  In the case of layer 3 PE-based VPNs, an IP packet coming from a customer site will be handled in the corresponding VFI. If the IP destination address in the packet's IP header belongs to the Internet, multiple scenarios are possible, based on the adapted policy. As a first possibility, when Internet access is not allowed, the packet will be dropped. As a second possibility, when

(controlled) Internet access is allowed, the IP packet will go through the translation function and eventually through the traffic analyzing function before further processing in the PE's global Internet forwarding table.

Note that different implementation choices are possible. One can choose to implement the translation and/or the traffic analyzing function in every VFI (or CE device in the context of layer 3 provider-provisioned CE-based VPNs), or alternatively in a subset or even in only one VPN network element. This would mean that the traffic to/from the Internet from/to any VPN site needs to be routed through that single network element (this is what happens in a hub and spoke topology for example).

4.7. Network and Customer Management of VPNs

4.7.1. Network and Customer Management

Network and customer management systems responsible for managing VPN networks have several challenges depending on the type of VPN network or networks they are required to manage.

For any type of provider-provisioned VPN it is useful to have one place where the VPN can be viewed and optionally managed as a whole. The NMS may therefore be a place where the collective instances of a VPN are brought together into a cohesive picture to form a VPN. To

be more precise, the instances of a VPN on their own do not form the VPN; rather, the collection of disparate VPN sites together forms the VPN. This is important because VPNs are typically configured at the edges of the network (i.e., PEs) either through manual configuration or auto-configuration. This results in no state information being kept in within the "core" of the network. Sometimes little or no information about other PEs is configured at any particular PE.

Support of any one VPN may span a wide range of network equipment, potentially including equipment from multiple implementors. Allowing a unified network management view of the VPN therefore is simplified through use of standard management interfaces and models. This will also facilitate customer self-managed (monitored) network devices or systems.

In cases where significant configuration is required whenever a new service is provisioned, it is important for scalability reasons that the NMS provide a largely automated mechanism for this operation. Manual configuration of VPN services (i.e., new sites, or re-provisioning existing ones), could lead to scalability issues, and should be avoided. It is thus important for network operators to maintain visibility of the complete picture of the VPN through the NMS system. This must be achieved using standard protocols such as SNMP, XML, or LDAP. Use of proprietary command-line interfaces has the disadvantage that proprietary interfaces do not lend themselves to standard representations of managed objects.

To achieve the goals outlined above for network and customer management, device implementors should employ standard management interfaces to expose the information required to manage VPNs. To this end, devices should utilize standards-based mechanisms such as SNMP, XML, or LDAP to achieve this goal.

4.7.2. Segregated Access of VPN Information

Segregated access of VPNs information is important in that customers sometimes require access to information in several ways. First, it is important for some customers (or operators) to access PEs, CEs or P devices within the context of a particular VPN on a per-VPN-basis in order to access statistics, configuration or status information. This can either be under the guise of general management, operator-initiated provisioning, or SLA verification (SP, customer or operator).

Where users outside of the SP have access to information from PE or P devices, managed objects within the managed devices must be accessible on a per-VPN basis in order to provide the customer, the SP or the third party SLA verification agent with a high degree of security and convenience.

Security may require authentication or encryption of network management commands and information. Information hiding may use

encryption or may isolate information through a mechanism that provides per-VPN access. Authentication or encryption of both requests and responses for managed objects within a device may be employed. Examples of how this can be achieved include IPsec tunnels, SNMPv3 encryption for SNMP-based management, or encrypted telnet sessions for CLI-based management.

In the case of information isolation, any one customer should only be able to view information pertaining to its own VPN or VPNs. Information isolation can also be used to partition the space of managed objects on a device in such a way as to make it more convenient for the SP to manage the device. In certain deployments, it is also important for the SP to have access to information pertaining to all VPNs, thus it may be important for the SP to create virtual VPNs within the management domain which overlap across existing VPNs.

If the user is allowed to change the configuration of their VPN, then in some cases customers may make unanticipated changes or even mistakes, thereby causing their VPN to mis-behave. This in turn may require an audit trail to allow determination of what went wrong and some way to inform the carrier of the cause.

The segregation and security access of information on a per-VPN basis is also important when the carrier of carrier's paradigm is employed. In this case it may be desirable for customers (i.e., sub-carriers or VPN wholesalers) to manage and provision services within their VPNs on their respective devices in order to reduce the management overhead cost to the carrier of carrier's SP. In this case, it is important to observe the guidelines detailed above with regard to information hiding, isolation and encryption. It should be noted that there may be many flavors of information hiding and isolation employed by the carrier of carrier's SP. If the carrier of carriers SP does not want to grant the sub-carrier open access to all of the managed objects within their PEs or P routers, it is necessary for devices to provide network operators with secure and scalable per-VPN network management access to their devices. For the reasons outlined above, it therefore is desirable to provide standard mechanisms for achieving these goals.

5. Interworking Interface

This section describes interworking between different layer 3 VPN approaches. This may occur either within a single SP network, or at an interface between SP networks.

5.1. Interworking Function

Figure 2.5 (see section 2.1.3) illustrates a case where one or more PE devices sits at the logical interface between two different layer 3 VPN approaches. With this approach the interworking function occurs at a PE device which participates in two or more layer 3 VPN approaches. This might be physically located at the boundary between service providers, or might occur at the logical interface between different approaches within a service provider.

With layer 3 VPNs, the PE devices are in general layer 3 routers, and are able to forward layer 3 packets on behalf of one or more private networks. For example, it may be common for a PE device supporting layer 3 VPNs to contain multiple logical VFIs (sections 1, 2, 3.3.1, 4.4.2) each of which supports forwarding and routing for a private network.

The PE which implements an interworking function needs to participate in the normal manner in the operation of multiple approaches for supporting layer 3 VPNs. This involves the functions discussed elsewhere in this document, such as VPN establishment and maintenance, VPN tunneling, routing for the VPNs, and QoS maintenance.

VPN establishment and maintenance information, as well as VPN routing information will need to be passed between VPN approaches. This might involve passing of information between approaches as part of the interworking function. Optionally this might involve manual configuration so that, for example, all of the participants in the VPN on one side of the interworking function considers the PE performing the interworking function to be the point to use to contact a large number of systems (comprising all systems supported by the VPN located on the other side of the interworking function).

5.2. Interworking Interface

Figure 2.6 (see section 2.1.3) illustrates a case where interworking is performed by use of tunnels between PE devices. In this case each PE device participates in the operation of one layer 3 VPN approach. Interworking between approaches makes use of per-VPN tunnels set up between PE. Each PEs operates as if it is a normal PEs, and considers each tunnel to be associated with a particular VPN.

Information can then be transmitted over the interworking interface in the same manner that it is transmitted over a CE to PE interface.

In some cases establishment of the interworking interfaces may require manual configuration, for example to allow each PE to determine which tunnels should be set up, and which private network is associated with each tunnel.

5.2.1. Tunnels at the Interworking Interface

In order to implement an interworking interface between two SP networks for supporting one or more PPVPN spanning both SP networks, a mechanism for exchanging customer data as well as associated control data (e.g., routing data) should be provided.

Since PEs of SP networks to be interworked may only communicate over a network cloud, an appropriate tunnel established through the network cloud will be used for exchanging data associated with the PPVPN realized by interworked SP networks.

In this way, each interworking tunnel is assigned to an associated layer 3 PE-based VPN; in other words, a tunnel is terminated by a VFI (associated with the PPVPN) in a PE device. This scenario results in implementation of traffic isolation for PPVPNs supported by an Interworking Interface and spanning multiple SP networks (in each SP network, there is no restriction in applied technology for providing PPVPN so that both sides may adopt different technologies). The way of the assignment of each tunnel for a PE-based VPN is specific to implementation technology used by the SP network that is inter-connected to the tunnel at the PE device.

The identifier of layer 3 PE-based VPN at each end is meaningful only in the context of the specific technology of an SP network and need not be understood by another SP network interworking through the tunnel.

The following tunneling mechanisms may be used at the interworking interface. Available tunneling mechanisms include (but are not limited to): GRE, IP-in-IP, IP over ATM, IP over FR, IPsec, and MPLS.

o GRE

  The tunnels at interworking interface may be provided by GRE [RFC2784] with key and sequence number extensions [RFC2890].

- IP-in-IP

  The tunnels at interworking interface may be provided by IP-in-IP [RFC2003] [RFC2473].

- IP over ATM AAL5

  The tunnels at interworking interface may be provided by IP over ATM AAL5 [RFC2684] [RFC2685].

- IP over FR

  The tunnels at interworking interface may be provided by IP over FR.

- IPsec

  The tunnels at interworking interface may be provided by IPsec [RFC2401] [RFC2402].

- MPLS

  The tunnels at interworking interface may be provided by MPLS [RFC3031] [RFC3035].

5.3. Support of Additional Services

This subsection describes additional usages for supporting QoS/SLA, customer visible routing, and customer visible multicast routing, as services of layer 3 PE-based VPNs spanning multiple SP networks.

- QoS/SLA

  QoS/SLA management mechanisms for GRE, IP-in-IP, IPsec, and MPLS tunnels were discussed in sections 4.3.6 and 4.5. See these sections for details. FR and ATM are capable of QoS guarantee. Thus, QoS/SLA may also be supported at the interworking interface.

- Customer visible routing

As described in section 3.3, customer visible routing enables the exchange of unicast routing information between customer sites using a routing protocol such as OSPF, IS-IS, RIP, and BGP-4. On the interworking interface, routing packets, such as OSPF packets, are transmitted through a tunnel associated with a layer 3 PE-based VPN in the same manner as that for user data packets within the VPN.

- Customer visible multicast routing

  Customer visible multicast routing enables the exchange of multicast routing information between customer sites using a routing protocol such as DVMRP and PIM. On the interworking interface, multicast routing packets are transmitted through a tunnel associated with a layer 3 PE-based VPN in the same manner as that for user data packets within the VPN. This enables a multicast tree construction within the layer 3 PE-based VPN.

5.4. Scalability Discussion

This subsection discusses scalability aspect of the interworking scenario.

- Number of routing protocol instances

  In the interworking scenario discussed in this section, the number of routing protocol instances and that of layer 3 PE-based VPNs are the same. However, the number of layer 3 PE-based VPNs in a PE device is limited due to resource amount and performance of the PE device. Furthermore, each tunnel is expected to require some bandwidth, but total of the bandwidth is limited by the capacity of a PE device; thus, the number of the tunnels is limited by the capabilities of the PE. This limit is not a critical drawback.

- Performance of packet transmission

  The interworking scenario discussed in this section does not place any additional burden on tunneling technologies used at interworking interface. Since performance of packet transmission depends on a tunneling technology applied, it should be carefully

selected when provisioning interworking. For example, IPsec places computational requirements for encryption/decryption.

6. Security Considerations

Security is one of the key requirements concerning VPNs. In network environments, the term security currently covers many different aspects of which the most important from a networking perspective are shortly discussed hereafter.

Note that the Provider-Provisioned VPN requirements document explains the different security requirements for Provider-Provisioned VPNs in more detail.

6.1. System Security

Like in every network environment, system security is the most important security aspect that must be enforced. Care must be taken that no unauthorized party can gain access to the network elements that control the VPN functionality (e.g., PE and CE devices).

As the VPN customers are making use of the shared SP's backbone, the SP must ensure the system security of its network elements and management systems.

6.2. Access Control

When a network or parts of a network are private, one of the requirements is that access to that network (part) must be restricted to a limited number of well-defined customers. To accomplish this requirement, the responsible authority must control every possible access to the network.

In the context of PE-based VPNs, the access points to a VPN must be limited to the interfaces that are known by the SP.

6.3. Endpoint Authentication

When one receives data from a certain entity, one would like to be

sure of the identity of the sending party. One would like to be sure that the sending entity is indeed whom he or she claims to be, and that the sending entity is authorized to reach a particular destination.

In the context of layer 3 PE-based VPNs, both the data received by the PEs from the customer sites via the SP network and destined for a customer site should be authenticated.

Note that different methods for authentication exist. In certain circumstances, identifying incoming packets with specific customer interfaces might be sufficient. In other circumstances, (e.g., in temporary access (dial-in) scenarios), a preliminary authentication phase might be requested. For example, when PPP is used. Or alternatively, an authentication process might need to be present in every data packet transmitted (e.g., in remote access via IPsec).

For layer 3 PE-based VPNs, VPN traffic is tunneled from PE to PE and the VPN tunnel endpoint will check the origin of the transmitted packet. When MPLS is used for VPN tunneling, the tunnel endpoint

checks whether the correct labels are used. When IPsec is used for VPN tunneling, the tunnel endpoint can make use of the IPsec authentication mechanisms.

In the context of layer 3 provider-provisioned CE-based VPNs, the endpoint authentication is enforced by the CE devices.

6.4. Data Integrity

When information is exchanged over a certain part of a network, one would like to be sure that the information that is received by the receiving party of the exchange is identical to the information that was sent by the sending party of the exchange.

In the context of layer 3 PE-based VPNs, the SP assures the data integrity by ensuring the system security of every network element. Alternatively, explicit mechanisms may be implemented in the used tunneling technique (e.g., IPsec).

In the context of layer 3 provider-provisioned CE-based VPNs, the underlying network that will tunnel the encapsulated packets will not always be of a trusted nature, and the CE devices that are responsible for the tunneling will also ensure the data integrity, e.g., by making use of the IPsec architecture.

6.5. Confidentiality

One would like that the information that is being sent from one party to another is not received and not readable by other parties. With traffic flow confidentiality one would like that even the characteristics of the information sent is hidden from third parties. Data privacy is the confidentiality of the user data.

In the context of PPVPNs, confidentiality is often seen as the basic service offered, as the functionalities of a private network are offered over a shared infrastructure.

In the context of layer 3 PE-based VPNs, as the SP network (and more precisely the PE devices) participates in the routing and forwarding of the customer VPN data, it is the SP's responsibility to ensure confidentiality. The technique used in PE-based VPN solutions is the ensuring of PE to PE data separation. By implementing VFI's in the PE devices and by tunneling VPN packets through the shared network infrastructure between PE devices, the VPN data is always kept in a separate context and thus separated from the other data.

In some situations, this data separation might not be sufficient. Circumstances where the VPN tunnel traverses other than only trusted and SP controlled network parts require stronger confidentiality measures such as cryptographic data encryption. This is the case in certain inter-SP VPN scenarios or when the considered SP is on itself a client of a third party network provider.

For layer 3 provider-provisioned CE-based VPNs, the SP network does not bare responsibility for confidentiality assurance, as the SP just offers IP connectivity. The confidentiality will then be enforced at the CE and will lie in the tunneling (data separation) or in the

cryptographic encryption (e.g., using IPsec) by the CE device.

Note that for very sensitive user data (e.g., used in banking operations) the VPN customer may not outsource his data privacy enforcement to a trusted SP. In those situations, PE-to-PE confidentiality will not be sufficient and end-to-end cryptographic encryption will be implemented by the VPN customer on its own private equipment (e.g., using CE-based VPN technologies or cryptographic encryption over the provided VPN connectivity).

6.6. User Data and Control Data

An important remark is the fact that both the user data and the VPN control data must be protected.

Previous subsections were focused on the protection of the user data, but all the control data (e.g., used to set up the VPN tunnels, used to configure the VFI's or the CE devices (in the context of layer 3 provider-provisioned CE-based VPNs)) will also be secured by the SP to prevent deliberate misconfiguration of provider-provisioned VPNs.

6.7. Security Considerations for Inter-SP VPNs

In certain scenarios, a single VPN will need to cross multiple SPs.

The fact that the edge-to-edge part of the data path does not fall under the control of the same entity can have security implications, for example with regards to endpoint authentication.

Another point is that the SPs involved must closely interact to avoid conflicting configuration information on VPN network elements (such as VFIs, PEs, CE devices) connected to the different SPs.

Appendix A: Optimizations for Tunnel Forwarding

A.1. Header Lookups in the VFIs

If layer 3 PE-based VPNs are implemented in the most straightforward manner, then it may be necessary for PE devices to perform multiple header lookups in order to forward a single data packet. This section discusses an example of how multiple lookups might be needed with the most straightforward implementation. Optimizations which might optionally be used to reduce the number of lookups are discussed in the following sections.

As an example, in many cases a tunnel may be set up between VFIs within PEs for support of a given VPN. When a packet arrives at the egress PE, the PE may need to do a lookup on the outer header to determine which VFI the packet belongs to. The PE may then need to do a second lookup on the packet that was encapsulated across the VPN tunnel, using the forwarding table specific to that VPN, before forwarding the packet.

For scaling reasons it may be desired in some cases to set up VPN tunnels, and then multiplex multiple VPN-specific tunnels within the VPN tunnels.

This implies that in the most straightforward implementation three header lookups might be necessary in a single PE device: One lookup may identify that this is the end of the VPN tunnel (implying the need to strip off the associated header). A second lookup may identify that this is the end of the VPN-specific tunnel. This lookup will result in stripping off the second encapsulating header, and will identify the VFI context for the final lookup. The last lookup will make use of the IP address space associated with the VPN, and will result in the packet being forwarded to the correct CE within the correct VPN.

A.2. Penultimate Hop Popping for MPLS

Penultimate hop popping is an optimization which is described in the MPLS architecture document [RFC3031].

Consider the egress node of any MPLS LSP. The node looks at the label, and discovers that it is the last node. It then strips off the label header, and looks at the next header in the packet (which may be an IP header, or which may have another MPLS header in the case that hierarchical nesting of LSPs is used). For the last node on the LSP, the outer MPLS header doesn't actually convey any useful information (except for one situation discussed below).

For this reason, the MPLS standards allow the egress node to request that the penultimate node strip the MPLS header. If requested, this implies that the penultimate node does not have a valid label for the LSP, and must strip the MPLS header. In this case, the egress node receives the packet with the corresponding MPLS header already stripped, and can forward the packet properly without needing to strip the header for the LSP which ends at that egress node.

There is one case in which the MPLS header conveys useful information: This is in the case of a VPN-specific LSP terminating at a PE device. In this case, the value of the label tells the PE which LSP the packet is arriving on, which in turn is used to determine which VFI is used for the packet (i.e., which VPN-specific forwarding table needs to be used to forward the packet).

However, consider the case where multiple VPN-specific LSPs are multiplexed inside one PE-to-PE LSP. Also, let's suppose that in this case the egress PE has chosen all incoming labels (for all LSPs) to be unique in the context of that PE. This implies that the label associated with the PE-to-PE LSP is not needed by the egress node. Rather, it can determine which VFI to use based on the VPN-specific LSP. In this case, the egress PE can request that the penultimate LSR performs penultimate label popping for the PE-to-PE LSP. This eliminates one header lookup in the egress LSR.

Note that penultimate node label popping is only applicable for VPN standards which use multiple levels of LSPs. Even in this case penultimate node label popping is only done when the egress node specifically requests it from the penultimate node.

A.3. Demultiplexing to Eliminate the Tunnel Egress VFI Lookup

Consider a VPN standard which makes use of MPLS as the tunneling mechanism. Any standard for encapsulating VPN traffic inside LSPs needs to specify what degree of granularity is available in terms of the manner in which user data traffic is assigned to LSPs. In other words, for any given LSP, the ingress or egress PE device needs to know which LSPs need to be set up, and the ingress PE needs to know which set of VPN packets are allowed to be mapped to any particular LSP.

Suppose that a VPN standard allows some flexibility in terms of the mapping of packets to LSPs, and suppose that the standard allows the egress node to determine the granularity. In this case the egress node would need to have some way to indicate the granularity to the ingress node, so that the ingress node will know which packets can be mapped to each LSP.

In this case, the egress node might decide to have packets mapped to LSPs in a manner which simplifies the header lookup function at the egress node. For example, the egress node could determine which set of packets it will forward to a particular neighbor CE device. The egress node can then specify that the set of IP packets which are to use a particular LSP correspond to that specific set of packets. For packets which arrive on the specified LSP, the egress node does not need to do a header lookup on the VPN's customer address space: It can just pop the MPLS header and forward the packet to the appropriate CE device. If all LSPs are set up accordingly, then the egress node does not need to do any lookup for VPN traffic which arrives on LSPs from other PEs (in other words, the PE device will not need to do a second lookup in its role as an egress node).

Note that PE devices will most likely also be an ingress routers for traffic going in the other direction. The PE device will need to do an address lookup in the customer network's address space in its role as an ingress node. However, in this direction the PE still needs to do only a single header lookup.

When used with MPLS tunnels, this optional optimization reduces the need for header lookups, at the cost of possibly increasing the number of label values which need to be assigned (since one label would need to be assigned for each next-hop CE device, rather than just one label for every VFI).

The same approach is also possible when other encapsulations are used, such as GRE [RFC2784] [RFC2890], IP-in-IP [RFC2003] [RFC2473], or IPsec [RFC2401] [RFC2402]. This requires that distinct values are used for the multiplexing field in the tunneling protocol. See section 4.3.2 for detail.

Acknowledgments

This document is output of the framework document design team of the PPVPN WG. The members of the design team are listed in the "contributors" and "author's addresses" sections below.

However, sources of this document are based on various inputs from colleagues of authors and contributors. We would like to thank

Junichi Sumimoto, Kosei Suzuki, Hiroshi Kurakami, Takafumi Hamano, Naoto Makinae, Kenichi Kitami, Rajesh Balay, Anoop Ghanwani, Harpreet Chadha, Samir Jain, Lianghwa Jou, Vijay Srinivasan, and Abbie Matthews.

We would also like to thank Yakov Rekhter, Scott Bradner, Dave McDysan, Marco Carugi, Pascal Menezes, Thomas Nadeau, and Alex Zinin for their valuable comments and suggestions.

Normative References

[PPVPN-REQ] Nagarajan, A., Ed., "Generic Requirements for Provider Provisioned Virtual Private Networks (PPVPN)", RFC 3809, June 2004.

[L3VPN-REQ] Carugi, M., Ed. and D. McDysan, Ed., "Service Requirements for Layer 3 Provider Provisioned Virtual Private Networks (PPVPNs)", RFC 4031, April 2005.

Informative References

[BGP-COM] Sangli, S., et al., "BGP Extended Communities Attribute", Work In Progress, February 2005.

[MPLS-DIFF-TE] Le Faucheur, F., Ed., "Protocol extensions for support of Differentiated-Service-aware MPLS Traffic Engineering", Work In Progress, December 2004.

[VPN-2547BIS] Rosen, E., et al., "BGP/MPLS VPNs", Work In Progress.

[VPN-BGP-OSPF] Rosen, E., et al., "OSPF as the Provider/Customer Edge Protocol for BGP/MPLS IP VPNs", Work In Progress, May 2005.

[VPN-CE] De Clercq, J., et al., "An Architecture for Provider Provisioned CE-based Virtual Private Networks using IPsec", Work In Progress.

[VPN-DISC] Ould-Brahim, H., et al., "Using BGP as an Auto-Discovery Mechanism for Layer-3 and Layer-2 VPNs," Work In Progress.

[VPN-L2] Andersson, L. and E. Rosen, Eds., "Framework for Layer 2 Virtual Private Networks (L2VPNs)", Work In Progress.

[VPN-VR] Knight, P., et al., "Network based IP VPN Architecture Using Virtual Routers", Work In Progress, July 2002.

[RFC1195] Callon, R., "Use of OSI IS-IS for Routing in TCP/IP and Dual Environments", RFC 1195, December 1990.

[RFC1771] Rekhter, Y. and T. Li, "A Border Gateway Protocol 4 (BGP-4)", RFC 1771, March 1995.

[RFC1918] Rekhter, Y., Moskowitz, B., Karrenberg, D., de Groot, G., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, February 1996.

[RFC1966] Bates, T., "BGP Route Reflection: An alternative to full mesh IBGP", RFC 1966, June 1996.

[RFC1997] Chandra, R., Traina, P., and T. Li, "BGP Communities Attribute", RFC 1997, February 2001.

[RFC2003] Perkins, C., "IP Encapsulation within IP", RFC 2003, October 1996.

[RFC2205] Braden, R., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.

[RFC2208] Mankin, A., Ed., Baker, F., Braden, B., Bradner, S., O'Dell, M., Romanow, A., Weinrib, A., and L. Zhang, "Resource ReSerVation Protocol (RSVP) Version 1 Applicability Statement Some Guidelines on Deployment", RFC 2208, September 1997.

[RFC2210] Wroclawski, J., "The Use of RSVP with IETF Integrated Services", RFC 2210, September 1997.

[RFC2211] Wroclawski, J., "Specification of the Controlled-Load Network Element Service", RFC 2211, September 1997.

[RFC2212] Shenker, S., Partridge, C., and R. Guerin, "Specification of Guaranteed Quality of Service", RFC 2212, September 1997.

[RFC2207] Berger, L. and T. O'Malley, "RSVP Extensions for IPSEC Data Flows", RFC 2207, September 1997.

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

[RFC2401] Kent, S. and R. Atkinson, "Security Architecture for the Internet Protocol", RFC 2401, November 1998.

[RFC2402] Kent, S. and R. Atkinson, "IP Authentication Header", RFC 2402, November 1998.

[RFC2406] Kent, S. and R. Atkinson, "IP Encapsulating Security Payload (ESP)", RFC 2406, November 1998.

[RFC2409] Harkins, D. and D. Carrel, "The Internet Key Exchange (IKE)", RFC 2409, November 1998.

[RFC2453] Malkin, G., "RIP Version 2", STD 56, RFC 2453, November 1994.

[RFC2473] Conta, A. and S. Deering, "Generic Packet Tunneling in IPv6 Specification", RFC 2473, December 1998.

[RFC2474] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, December 1998.

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An architecture for Differentiated

Services", RFC 2475, December 1998.

[RFC2597] Heinanen, J., Baker, F., Weiss, W., and J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, June 1999.

[RFC2661] Townsley, W., Valencia, A., Rubens, A., Pall, G., Zorn, G., and B. Palter, "Layer Two Tunneling Protocol 'L2TP'", RFC 2661, August 1999.

[RFC2684] Grossman, D. and J. Heinanen, "Multiprotocol Encapsulation Over ATM Adaptation Layer 5", RFC 2684, September 1999.

[RFC2685] Fox B. and B. Gleeson, "Virtual Private Networks Identifier," RFC 2685, September 1999.

[RFC2746] Terzis, A., Krawczyk, J., Wroclawski, J., and L. Zhang, "RSVP Operation Over IP Tunnels", RFC 2746, January 2000.

[RFC2764] Gleeson, B., Lin, A., Heinanen, J., Armitage, G., and A. Malis, "A Framework for IP Based Virtual Private Networks", RFC 2764, February 2000.

[RFC2784] Farinacci, D., Li, T., Hanks, S., Meyer, D., and P. Traina, "Generic Routing Encapsulation (GRE)", RFC 2784, March 2000.

[RFC2890] Dommety, G., "Key and Sequence Number Extensions to GRE", RFC 2890, September 2000.

[RFC2858] Bates, T., Rekhter, Y., Chandra, R., and D. Katz, "Multiprotocol Extensions for BGP-4", RFC 2858, June 2000.

[RFC2983] Black, D., "Differentiated Services and Tunnels", RFC 2983, October 2000.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC

3031, January 2001.

[RFC3032] Rosen E., Tappan, D., Fedorkow, G., Rekhter, Y., Farinacci, D., Li, T., and A. Conta, "MPLS Label Stack Encoding", RFC 3032, January 2001.

[RFC3035] 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.

[RFC3065] Traina, P., McPherson, D., and J. Scudder, "Autonomous System Confederations for BGP", RFC 3065, June 1996.

[RFC3209] 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.

[RFC3246] Davie, B., Charny, A., Bennet, J.C.R., Benson, K., Le Boudec, J.Y., Courtney, W., Davari, S., Firoiu, V., and D. Stiliadis, "An Expedited Forwarding PHB (Per-Hop Behavior)", RFC 3246, March 2002.

[RFC3270] Le Faucheur, F., Wu, L., Davie, B., Davari, S., Vaananen, P., Krishnan, R., Cheval, P., and J. Heinanen, "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services", RFC 3270, May 2002.

[RFC3377] Hodges, J. and R. Morgan, "Lightweight Directory Access Protocol (v3): Technical Specification", RFC 3377, September 2002.

Contributors' Addresses

Jeremy De Clercq
Alcatel
Fr. Wellesplein 1,

2018 Antwerpen, Belgium

EMail: jeremy.de_clercq@alcatel.be

Bryan Gleeson
Nokia
313 Fairchild Drive,
Mountain View, CA 94043 USA.

EMail: bryan.gleeson@nokia.com

Andrew G. Malis
Tellabs
90 Rio Robles Drive
San Jose, CA 95134 USA

EMail: andy.malis@tellabs.com

Karthik Muthukrishnan
Lucent Technologies
1 Robbins Road
Westford, MA 01886, USA

EMail: mkarthik@lucent.com

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

EMail: erosen@cisco.com

Chandru Sargor
Redback Networks
300 Holger Way
San Jose, CA 95134, USA

EMail: apricot+l3vpn@redback.com

Jieyun Jessica Yu
University of California, Irvine
5201 California Ave., Suite 150,
Irvine, CA, 92697 USA

EMail: jyy@uci.edu

Authors' Addresses

Ross Callon
Juniper Networks
10 Technology Park Drive
Westford, MA 01886-3146, USA

EMail: rcallon@juniper.net

Muneyoshi Suzuki
NTT Information Sharing Platform Labs.
3-9-11, Midori-cho,
Musashino-shi, Tokyo 180-8585, Japan

EMail: suzuki.muneyoshi@lab.ntt.co.jp

Full Copyright Statement

Intellectual Property

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet
gement