Network Working Group
Internet Draft
Expiration Date: November 2001

Eric C. Rosen
Clarence Filsfils
Cisco Systems, Inc.

May 2001

# An Architecture for L2VPNs

draft-rosen-ppvpn-l2vpn-00.txt

## Status of this Memo



## Abstract

Service Providers may offer a "Layer 2 VPN Service" over an IP backbone by provisioning point-to-point "virtual circuits" that run through IP tunnels. This document discusses the signaling, encapsulation, and configuration issues that arise. Its purpose is to provide an architecture which allows different kinds of point-to-point virtual circuits to be provided through different kinds of IP tunnels.

Internet Draft draft-rosen-ppvpn-l2vpn-00.txt May 2001

Table of Contents



## 1. Boilerplate for Sub-IP Area Drafts

This draft is targeted at the PPVPN WG, as it addresses the following work item from the PPVPN WG charter:

> "The working group is expected to consider at least three specific approaches including ... port-based VPNs (i.e., where the SP provides a Layer 2 interface, such as Frame Relay or ATM, to the VPN customer, while using IP-based mechanisms in the provider infrastructure"

The set of related documents may be found in the "References" section.

## 2. Introduction

Enterprises have long built their own wide-area networks by purchasing wide-area point-to-point data link layer connectivity from service providers, and then building their own layer 3 infrastructure on top of that. Originally the data links from the service provider were leased lines, and the layer 3 overlays were termed "private networks". Later, virtual circuits of various sorts (X.25, Frame Relay, ATM) began to replace leased lines, and the layer 3 overlays were termed "virtual private networks". (Though what makes a leased line less virtual than an ATM VC is difficult to understand.)

We will refer to these VPNs as "Layer 2 VPNs" because the service provider providers only a layer 2 interface to its customer, and the customer is responsible for creating and managing the layer 3 overlay.

Today, layer 2 VPNs are usually offered over a Frame Relay or ATM infrastructure.

There are still many enterprises that wish to manage their own layer 3 overlays, and many service providers who wish to provide layer 2 interfaces to such customers. However, many of these service providers would like to replace their Frame Relay or ATM infrastructures with an IP infrastructure. So it is desirable to have standard ways of using an IP infrastructure to provide a layer 2 interface to customers. In particular, it is desirable to have standard ways of using an IP infrastructure to provide virtual circuits between pairs of customer sites.

The term "Layer 2 VPN" may be somewhat misleading, in that the SP does not actually provide a VPN to the customer. The SP provides layer 2 connectivity, and the customer builds his own VPN, using the provided layer 2 connectivity as one of the building blocks. The problem is really how to provide the layer 2 connectivity over an IP backbone, rather than how to provide a network service over an IP

backbone.

Typically a customer will expect a certain amount of bandwidth from each data link. The customer may build his network using data links obtained from a variety of providers.

In an L2VPN service, the SP need not know about the customer's topology, about the customer's policies, or about the customer's routing. The SP need not even know whether all the point-to-point links he is providing are used by the customer as part of the same network, or whether a customer network has point-to-point links from other providers as well. In essence the customer builds his own





network, using data link resources obtained from the SP. Nevertheless, we will continue to use the term "Layer 2 VPN" (L2VPN), as it is apparently well entrenched in the vernacular, despite its technical inaccuracy.

We adopt the "Provider Edge" (PE) and "Customer Edge" (CE) terminology from RFC2547 [RFC2547bis].

Not all L2VPNs are built on top of point-to-point data link connections. It is possible for an SP to provide an "emulated LAN" service instead. In this case, the PE device is a LAN switch which can serve multiple customers, and which can do SA learning and Spanning Tree on a per-customer basis. Some sort of multicast technique must be used for transmitting customer LAN multicasts and unknown DA frames among the PEs which attach to a common customer. The primary focus of this draft will however be on the provision of point-to-point data link services. If the CE devices attaching to an L2VPN's PE devices are LAN switches, the L2VPN may be thought of as a "Transparent LAN Service", even if the SP provides only point-to-point data link connections.

The provision of Emulated LAN services over IP backbone networks can be added at a later date if there is sufficient interest.

## 3. Outline of Architecture

The general architecture for providing L2VPN services is the

following. A CE devices attaches to a PE device via some sort of virtual circuit. We will call this the "Attachment VC". To provide a layer 2 connection between CE1 and CE2, where CE1 attaches to PE1 and CE2 attaches to PE2, an "Emulated VC" must be carried across the IP backbone from PE1 to PE2. At each of PE1 and PE2, the Emulated VC is associated with an Attachment VC. In effect, the ordered triple <CE1-PE1 Attachment VC, PE1-PE2 Emulated VC, PE2-CE2 Attachment VC> functions as a VC between CE1 and CE2.

The Emulated VC is carried in a "Tunnel" from PE1 to PE2. When there are multiple Emulated VCs running from PE1 to PE2, a single Tunnel should be able to carry a large number of Emulated VCs. There should be no requirement that two Emulated VCs in a common tunnel have the same CE endpoints. When PE2 removes a packet from a Tunnel, it associates the packet with an Emulated VC. The association of the Emulated VC with an Attachment VC determines the CE to which the packet is sent.

There should be no requirement that all VCs going from PE1 to PE2





travel in the same tunnel.

If the two Attachment VCs associated with an Emulated VC are of the same type (e.g., both Frame Relay, both ATM), the Emulated VC may need to carry some type-specific information with each packet. If the two Attachment VCs are not of the same type, one or the other end of the Emulated VC must perform some sort of inter-working function.

It is desired to allow a variety of different tunneling technologies to be used for the PE-PE Tunnels.

In order to provide this sort of L2VPN architecture in a standard way, the following need to be standardized:

- Tunneling Protocols. The architecture allows any number of different tunneling protocols to be used, but each should be standardized.

    * Signaling. The standard for a tunneling protocol will generally include a signaling protocol, so that tunnels can be set up dynamically, and tunnel control information can be passed from one tunnel endpoint to another.

    * Encapsulation The standard for a tunneling protocol always includes a way to encapsulate data packets in the tunnel.

    * Multiplexing. Most tunneling protocols allow for multiple streams to be encapsulated inside a single tunnel. They do this by supporting a multiplexing field. To support L2VPNs, each Emulated VC in the tunnel should correspond to a distinct value of the tunnel multiplexing field. It is important to note that this multiplexing field belongs to the tunneling protocol, not to the data packet that is encapsulated in the tunnel.

- Signaling Protocols for the Emulated VCs. A protocol is needed to setup and maintain the Emulated VCs that are carried within the tunnels. This protocol does not set up the tunnels, but only the Emulated VCs within the tunnels.

  Note though that to set up an Emulated VC, one must set up the multiplexing field value which the tunnel protocol will use when carrying packets of that Emulated VC. This strongly suggests that the signaling protocol for setting up an Emulated VC be specific to the particular tunneling technology that is being used.





- Encapsulations for the user data frames. For each kind of layer 2 frame that may be received on an Attachment VC, an encapsulation must be specified that allows the frame and its related per-frame control information to be carried within the Emulated VC.

While not strictly required for standardization, it is also important to discuss the configuration model for setting up the L2VPN service. If it turns out that the configuration model can be considerably simplified through the use of an auto-discovery mechanism, the auto-discovery mechanism will need to be standardized.

## 4. Encapsulating

An encapsulation for User Data Frames is proposed in [L2ENCAPS]. Actually, that draft proposes both a tunnel-independent encapsulation for user data frames, as well as the method for encapsulating the frames within an MPLS tunnel. We propose to adopt the tunnel-independent encapsulation specified therein.

## 5. Signaling

As we stated in the introduction, the signaling used to setup and maintain the Emulated VCs should depend on the particular tunneling technology. If MPLS is to be used as the tunneling technology, the procedures specified in [L2SIG]. That draft proposes to use extensions to the standard MPLS signaling (LDP) to set up the multiplexing field (itself an MPLS label) for the Emulated VC, as well as to setup and pass other control information needed to maintain the Emulated VC.

If the tunneling technology is L2TP or IPsec, then the signaling protocols which are native to those tunnel technologies should be similarly extended.

## 6. Tunneling

In the case where MPLS is the tunneling technology, [L2SIG] specifies the way in which frames on one or more Emulated VCs are to be carried in an MPLS LSP.

Similar drafts are needed for L2TP and IPsec tunnels.





## 7. Configuration Models

It can be somewhat unwieldy to configure a large number of point-to-point VCs. Therefore a number of L2VPN proposals focus on methods of simplifying the configuration, either by adding additional signaling mechanisms, or by adding auto-configuration and auto-discovery mechanisms, or both. The current proposal is to separate the signaling from any auto-configuration or auto-discovery mechanisms.

Then we can discuss separately the procedures for signaling, given a particular configuration, and the procedures for creating the that configuration in the first place.

This approach differs from the approach of RFC2547 for layer 3 VPNs; there the auto-discovery of VPN sites is combined with the signaling of MPLS labels for VPN routes. What we propose here is more like the way auto-discovery is separated from virtual circuit setup in the Virtual Router (VR) model of layer 3 VPNs. (See, e.g., [AUTO].) In both the L2VPN case and the VR case, it is necessary to set up data link connections which go from one PE to another. In the RFC2547 case, there are no cross-network data link connections set up. Setting up a point-to-point data link connection requires signaling of the sort specified in [L2SIG], and cannot be properly automated as a side effect of the auto-discovery procedures.

## 7.1. Brute Force

In order to provide L2PVN service connecting CE1 and CE2, one configuration model is the following:

- CE1 must be configured with an Attachment VC to PE1. Call this A1.
- PE1 must be configured with that same Attachment VC (to CE1).
- CE2 must be configured with an Attachment VC to PE2. Call this A2.
- PE2 must be configured with that same attachment VC (to CE2).
- PE1 must be configured with an identified Emulated VC to PE2.
- PE2 must be configured with an identified Emulated VC to PE1; the identifier should be the same at both PEs, and the triple <PE1, PE2, identifier> must be unique. Call this E.
- PE1 must be configured to associate A1 with E.
- PE2 must be configured to associate A2 with E.
- PE1 must be configured with a tunnel, T1, to PE2, and a tunnel, T2, from PE2. PE2 must be configured with a tunnel, T1, from PE1, and a tunnel, T2, to PE1.





- PE1 must be configured to associate E with T1 and T2.

- PE2 must be configured to associate E with T1 and T2.

Note also that A1, E, and A2 may have specific properties that need to be configured, e.g., QoS, bandwidth, etc.

## 7.2. Eliminating Tunnel Configuration

If MPLS is the tunneling technology, and LDP downstream unsolicited label distribution is used, it is NOT necessary to configure T1 and T2, or to explicitly associate E with them. The necessary tunnel exists automatically as long as there is a route from one PE to the other.

## 7.3. Single Endpoint Configuration of Emulated VCs

We assumed above that the Emulated VC signaling required that a particular Emulated VC be configured at both PE endpoints. This isn't necessarily the case. If the Emulated VC signaling protocol allows an Emulated VC to be set up based on the configuration of just a single endpoint, there is no need to configure the other endpoint, and those steps can be removed.

This enables the configuration model to be simplified to:

- CE1 must be configured with an Attachment VC to PE1. Call this A1.
- PE1 must be configured with that same Attachment VC (to CE1).
- CE2 must be configured with an Attachment VC to PE2. Call this A2.
- PE2 must be configured with that same attachment VC (to CE2).
- PE1 must be configured with an identified Emulated VC to PE2.
- PE1 must be configured to associate A1 and A2 with E.

In this case, PE1 must tell PE2 to associate A2 with E. If E has specified properties, PE1 needs to be configured with these, and must tell PE2 about them.

It is not completely obvious whether single endpoint configuration of Emulated VCs is really worthwhile. The provisioning system that configures the Emulated VCs will generally have to consult the configuration of both endpoint PEs to determine the availability of Attachment VCs, to set up QoS for Attachment VCs, etc. In any event, the signaling procedures of draft- martini-l2circuit-trans-mpls are easily extended to handle the case of signaling Emulated VCs that are configured at only one endpoint.

### 7.4. Single Endpoint Configuration of Attachment VCs

The need to configure the Attachment VCs on BOTH the CE and the PE could be eliminated by an appropriate LMI. Then an Attachment VC could be configured just on the PE, and the PE would use the LMI to inform the CE. The feasibility of this depends on the particular technology used for the Attachment VC, and whether such LMI procedures exist for that technology. If they do, they exist whether or not an L2VPN service of the sort envisioned here is being offered, so we need not consider this any further.

We have assumed that the L2VPN service will be a PVC rather than an SVC service. A model for supporting an SVC service is discussed in draft-ouldbrahim-bgpgmpls-ovpn. With the SVC model, the need to configure the Attachment VC on the PE could be eliminated. We do not further consider this here.

### 7.5. Zero-Configuration Attachment VCs

In the "Zero-Configuration of Attachment VCs" model, only the Emulated VC is configured (at one or both endpoints). The signaling of the Emulated VC doesn't specify a particular Attachment VC to associate it with, nor is a particular Attachment VC configured to be associated with the Emulated VC. Rather, the Emulated VC is simply associated with an interface at each end. When the Emulated VC is set up, each endpoint PE creates an Attachment VC on the specified interface, and some sort of LMI or signaling procedure is used to inform the CE of this.

Whether this is feasible depends on the particular technology used for the Attachment VCs. To the extent that an Attachment VC uses a scarce resource, this is not really feasible. For example, if the CE and the PE are connected via an ATM or Frame Relay switch, one could not automatically create an Attachment VC when the Emulated VC is setup. An Attachment VC in these technologies requires a switch cross-connect entry, and these scarce resources might not be available in the absence of an explicit provisioning process. As another example, if the Attachment VCs must have specific QoS properties, it might be necessary to do explicit provisioning to ensure that the necessary QoS characteristics can be met.

On the other hand, if an Attachment VC is nothing more than the value of some multiplexing field, with no particular QoS characteristics, and no use of switches between PE and CE, then it might be feasible to create the Attachment VCs automatically as a side-effect of

setting up an Emulated VC. The procedures for doing this would depend on the technology used for the Attachment VCs.





## 7.6. Auto-discovery of remote CEs

If an L2VPN is to have a hub and spoke topology, some further simplification of the configuration could be made, as one could provide some sort of auto-discovery of the hub CE. Then when a new spoke CE is attached to a PE, the PE would automatically determine which other PE attaches the corresponding hub CE.

Then when one adds a new spoke CE to the L2VPN, one wouldn't have to explicitly configure the attached PE with the information about how to reach the hub. However, a hub and spoke topology is already simple to configure, and this additional simplification is probably not worth the additional mechanism.

If an L2VPN is to have a fully meshed topology, simplification of the configuration could be made. If a PE is attached to a CE of a particular VPN, it could auto-discover all the other PEs that are attached to CEs of the same VPN, and could discover for each such PE the set of CEs in that VPN to which it is attached. This could be done using the auto-discovery techniques first described in RFC2547 and later extended and generalized in draft-ouldbrahim-bgpvpn-auto. Once a PE discovers the complete set of CEs in a given VPN, it can signal an Emulated VC from each of that VPN'S CEs to which it is attached to each other CE in that VPN. (This does presuppose that the Emulated VCs are of uniform characteristics. If each has some specific QoS property, for example, then this degree of auto-discovery would be impossible, and more explicit provisioning would be required.)

If the Attachment VC technology used by that VPN allows for Zero-Configuration Attachment VCs, a full mesh of point-to-point (CE-CEs) virtual circuits could be set up, with very little configuration. A PE would just need to be configured to know which VPN each of its CEs belongs to.

Whether this sort of auto-discovery is worthwhile is somewhat dubious, though, since it only really pays off if the L2VPN consists of a full mesh of point-to-point connections, and this is a very unusual topology. (Whereas a full mesh of L3 connectivity is the

common case, a full mesh of L2 connections is rather uncommon.) As discussed in draft-ouldbrahim-bgvpn- auto, additional configuration information ("colors") could be added to facilitate different topologies, but once one departs from either the hub and spoke or the full mesh topology, figuring out how to make the right topology auto-configure itself quickly becomes more difficult than explicitly provisioning it.

An auto-discovery scheme of this sort, though combined with a





particular signaling and encapsulation scheme, is detailed in [MPLSL2VPN].

## 8. References

[AUTO] "Using BGP as an Auto-Discovery Mechanism for Network-based VPNs", Ould-Brahim, et. al., draft-ouldbrahim-bgpvpn-auto-01.txt, 3/01.

[L2ENCAPS] "Encapsulation Methods for Transport of Layer 2 Frames Over MPLS", Martini, et. al., draft-martini-l2circuit-encap-mpls-01.txt, 2/01.

[L2SIG] "Transport of Layer 2 Frames Over MPLS", Martini, et. al., draft-martini-l2circuit-trans-mpls-05.txt, 2/01.

[MPLSL2VPN] "MPLS-based Layer 2 VPNs", Kompella, et. al., draft-kompella-mpls-l2vpn-02.txt, 11/00.

[RFC2547bis] "BGP/MPLS VPNs", Rosen et. al. draft-rosen-rfc2547bis-03.txt, 3/01.

## 9. Authors' Information

Eric C. Rosen
Cisco Systems, Inc.
250 Apollo Drive
Chelmsford, MA, 01824

E-mail: erosen@cisco.com

Clarence Filsfils
Cisco Systems, Inc.
Avenue Marcel Thiry, 77
B-1200 Brussels Belgium

E-mail: cfilsfil@cisco.com