MPLS Working Group
Internet Draft
Expiration Date: January 1999

D. Jamieson
B. Jamoussi
G. Wright
P. Beaubien
Nortel (Northern Telecom) Ltd.
August 1998

MPLS VPN Architecture

<draft-jamieson-mpls-vpn-00.txt>

Status of this Memo

Abstract

This Internet Draft defines an architectural model for building Virtual Private Networks (VPNs) in an MPLS domain. The proposed model takes advantage of both network layer peering and packet switching, and link layer circuits and per-stream switching. The model provides a set of simple mechanisms for controlling VPN membership, including registration, propagation, discovery, and dynamic creation of Label Switch Paths to provide connectivity.

The architectural constructs described in this document, when combined with the MPLS architecture [1], provide a flexible and scaleable basis for building VPNs.

Table of Contents

1. Introduction

Virtual Private Networks (VPNs) enable private restricted communications of distinct, closed networks over a common shared network infrastructure. Supporting VPNs with MPLS or other connectionless and connection-oriented layers requires three basic functions.

- Discovery of VPN members.

It is assumed that VPN members connect to a provider network and those members need to find out what other members there are in the VPN. Members may join and leave the service network and those changes need to be known by all remaining members. Mechanisms to support discovery include manual configuration, client-server approaches, and notification provided by the provider network

(i.e., auto-discovery). The discovery of membership in one VPN must not allow members of other VPNs to be discovered. That is, discovery within a VPN is kept separate from discovery in another VPN in the same provider network.

- Exchanging reachability and control traffic between VPN members.

  Members in the same VPN need to exchange reachability information about their network layer addresses. These addresses may be in a different space from the provider network and may in fact overlap with other VPN address spaces. Control traffic could include topology information specific to that VPN. As with the discovery mechanism, the exchange of reachability and control traffic must be kept separate between VPNs sharing the same provider network.

- Carrying data traffic between VPN members.

  This mechanism enables data traffic to be carried between users within a VPN. Data traffic from different VPNs is kept separate.

In [2] the discovery mechanism involves local configuration (VPNid) and then propagation in LDP, OSPF, or BGP. The reachability exchange is also accomplished by LDP, OSPF, or BGP. Topology information is not propagated between VPN member subnets over the MPLS network providing the VPN service. Data traffic is carried on LSPs which are created to connect all members of the same VPN.

This Internet-Draft proposes the use of OSPF, BGP-4, or TCP connections for the discovery mechanism. Reachability and control traffic (topology information) are exchanged over LSPs which are setup between members in the same VPN. Data traffic is carried on LSPs which are created to connect all members of the same VPN.

This internet draft is different from [2] and is proposed as an alternative.

In Section 2, an architectural overview of building VPNs in an MPLS domain is presented. Section 3 presents the details of the proposed architecture. Extending MPLS into the VPN member network is highlighted in Section 4. Section 5 summarizes the draft.

2. Architectural Overview

2.1 Building Blocks

The building blocks of the MPLS VPN architecture proposed in this draft are shown in Figure 1 and described in this section.

Private Network LSR (PNL):

The PNL is a device that runs standards based layer 3 (OSPF, BGP, RIP, static routes, etc.) protocols to distribute and calculate reachability information for the private network. It also runs an

LDP [3] process for the purpose of establishing Label Switched Paths (LSP) between itself and other members of the same VPN connected over the provider network. The PNL may be a physical device that resides in either the private or provider's premise. It could also be a logical device embedded in some other device, such as a Provider Edge LSR (PEL).

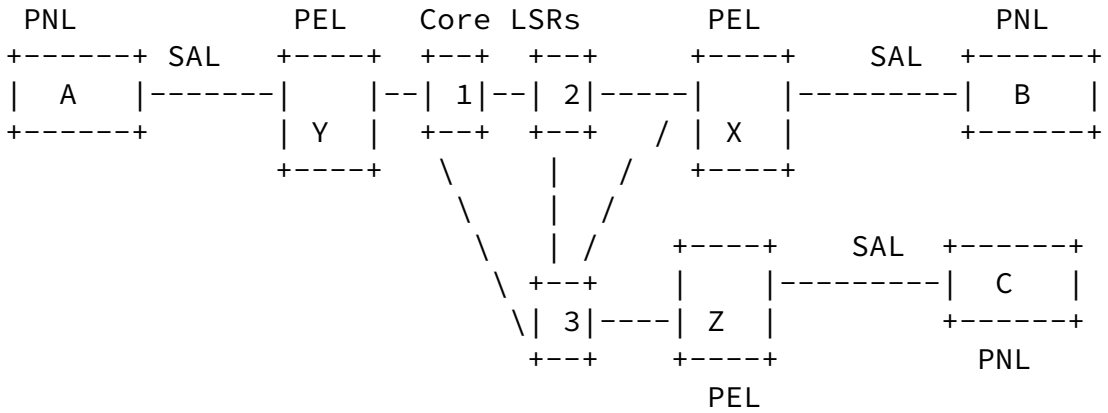

Figure 1. MPLS VPN Architecture

Provider Edge LSR (PEL):

The Provider Edge LSR (PEL) is an LSR in the provider domain. It has one or more Shared Access Links (SALs) connecting it to one or more PNLs. LDP peering is established over these SALs which is used to setup end to end (PNL to PNL) LSPs.

PELs dynamically discover other PELs supporting the same VPN and VPN subnets. LSPs are then established between those PELs to transport VPN traffic.

Core LSR:

Core LSRs provide transport across the provider network. They run a layer 3 protocol and MPLS. Core LSRs don't attach directly to PNLs.

Shared Access Link (SAL):

The SAL is a IP capable physical or logical link that connects the PNL to the PEL.

VPN Subnets:

A VPN subnet connects an IP subnet between 2 or more PNLs. A VPN subnet is uniquely identified within the provider network by a VPN

Id, an IP address and prefix.

VPN Subnet Interface (VSI):

The IP interface on a PNL for an VPN subnet. A SAL supports 1 or more VSIs.

The PNL device has a Shared Access Link (SAL) to a PEL. A VPN Subnet Interface (VSI) is established over the SAL. The VSI is viewed as a broadcast emulated LAN interface by the IP process running on the PNL. IP routing information can be exchanged between all PNLs of the same VPN subnet. The emulated LAN connectivity is achieved using a set of LSPs.

2.2 MPLS-VPN Architecture Summary

- The provider network provides LSPs that are used by PNLs of the same VPN subnet to exchange VPN routing information and to carry datagrams across the provider network.

- The exchange of routing information across provider network is

dynamic. This property eases network management and removes the need for static routing requiring operator intervention.

- No routing information is exchanged between PNLs and PELs. PNLs form peering relationships across the provider network. Eliminating the routing exchange between the PNL and the PEL provides several benefits:

  - Topology changes (route flapping) in the private network are transparent to the provider network. Routing engines in the LSRs inside the provider network are not affected by route flaps.

  - Topology changes in provider network are transparent to private network. When routes change in the provider network, new LSPs are created to re-route the VPN traffic without involving the PNLs.

  - Private routes are never mixed with provider routes. This eliminates possible address conflicts between VPNs.

- The provider network emulates a LAN for each VPN subnet. A particular PNL can send a unicast datagram to any other PNL in the same VPN subnet, or multicast a datagram to all other PNLs in the VPN subnet.

- The ELAN requires multicast capability. This functionality can be accomplished three ways: multipoint-to-multipoint LSPs, a set of

point-to-multipoint LSPs, or by PNL copy and send broadcast over existing unicast LSPs.

- Three types of LSPs are used to interconnect PNLs:

  - Multipoint-to-point LSP.

  Each PNL has a multipoint-to-point LSP directed to it. It is used by all other PNLs within the VPN subnet for unicast sends.

  - Multipoint-to-multipoint LSP (option 1).

  All PNLs are also interconnected using a bi-directional multipoint-to-multipoint LSP. It is used for sending multicast datagrams. There is one such LSP per VPN subnet.

- Point-to-multipoint LSP (option 2).

If multipoint-to-multipoint LSPs are not supported by the underlying infrastructure, then point-to-multipoint LSP going from each PNL to all other PNL in a VPN subnet is necessary.

- LSP scaling within a SAL.

N is defined as the number of PNLs in a VPN subnet. Each PNL therefore uses (assuming the single multipoint-to-point LSP model):

- 1 label for the incoming unicast datagram traffic from all other PNLs in the subnet,

- N-1 labels to send unicast datagrams to any other PNL in the subnet,

- 1 label to send and receive multicast traffic on the subnet using multicast option 1.

- N-1 labels to send and receive multicast traffic on the subnet using multicast option 2.

- MAC addresses are represented as labels. For a particular PNL, say PNL A, the MAC address of another PNL, say PNL B, is the label that must be used by PNL A to send unicast datagrams to PNL B. Because labels have local significance only, the MAC address used to reach a particular PNL is usually different for different senders.

- Layer 2 to layer 3 address mapping is achieved through one of 2 methods; propagating the information from the PEL to the PNL or a

modified ARP procedure

- When the PNL is an LSR in its own right, label stacking can be used to label-switch datagrams in that PNL (instead of doing layer-3 forwarding).

2.3 Emulated LAN Model

To provide maximum flexibility to the VPN members, the provider network appears as a Local Area Network (LAN) to the various VPN member sites as shown in Figure 2.

The MPLS architecture with architectural constructs described in this document provide for a flexible model to construct an emulated LAN in an efficient manner. There are several advantages to adopting an emulated LAN model as explained in this section:

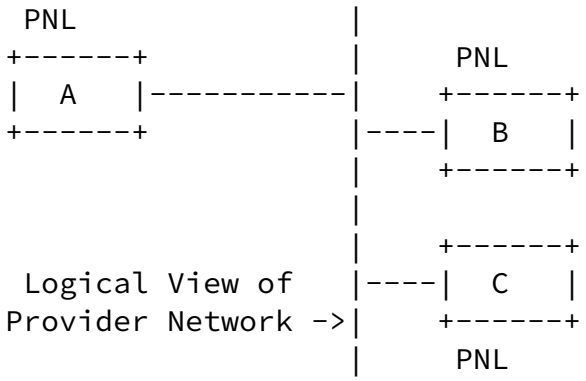

Figure 2. Emulated LAN in an MPLS Domain

- The emulated LAN model provides IP address space conservation. IP address pace conservation occurs two ways. First by eliminating the double addressing requirement for IP tunneling and by decreasing the subnet requirements for equivalent connectivity with current ATM or FR services

- The emulated LAN model simplifies the configuration of the VPN within the shared network. Adding or deleting a site from VPS only requires a change only on the interface being added or deleted.

2.4 Elements of a LAN model

Each node is identified by a MAC address. A MAC address is equivalent to a Label on a VSI port.

Each node on a LAN must be able to send a unicast packet to any other node on the LAN. This unicast traffic would include both control and

user traffic between any two given PNLs.

Each node on a LAN must be able to transmit a single packet onto the LAN and have it delivered to all other nodes on the LAN (multicast). These packets are sent to a multicast MAC address. Multicast traffic includes Hello packets, LSAs, ARP, etc.

2.5 Other models

This architecture does not rule out other models such as a star or point to point model. The details of other models are left for further study.

3. Architectural Details

This sections describes the following architectural components of the proposal:

- The provisioning of an SAL and the registration of VPN and VPN subnet information on a PEL

- The distribution of the VPN information across in the provider network

- The establishment of VPN subnet LSPs based on learned VPN subnet information

- The modeling of the VPN subnet LSPs into a LAN like broadcast media on the PNL

3.1 Registration of VPN and VPN subnet information on a PEL

The first step in adding a new site to a VPN subnet is to establish an SAL between the PEL and PNL. The SAL is the link over which LDP runs between the PEL and PNL. Only one SAL needs to be provisioned for all VPN subnets on a PNL, so if the SAL already existed this step can be skipped.

Once the SAL has been provisioned on the PEL, a VPN Identifier is assigned to the SAL. There is a one to one mapping between VPN Id and SAL. Again, if the SAL was already provisioned then the VPN Id will also have been provisioned.

The next step is to provision the VPN subnet information. This requires an IP address and prefix. The IP address and prefix are the same as the PNL's VSI to which this SAL is linked. The VPN Id together with the IP address and prefix define the VPN Subnet. The IP address itself defines an instance of the subnet. If the same PEL has

another SAL to another PNL that supports the same VPN Id and subnet then the IP address distinguishes between the two instances.

A protocol could be used to dynamically learn the IP address and prefix from the PNL. Because the learning of this information causes the consumption of resources in the provider network, appropriate control mechanism would have to be part of the protocol. The details of such a protocol are left for further study.

Once all of the VPN subnet information has been provisioned or learned on the PEL, LDP is triggered on the PEL to establish an LSP for the VPN subnet that goes from the PEL to the PNL. This LSP does not go any further at this point. It will be spliced onto a multipoint to point LSP later after other PELs supporting the same VPN subnet learn of the existence of this instance of the VPN subnet.

The successful establishment of this first LSP also signals to the PEL that the PNL has provisioned the associated VSI port and that port is enabled.

3.2 Distribution of VPN information

This section describes the distribution of the VPN subnet information within the provider network.

All PELs in the network, at least those that have links to the same VPN Subnet, must be made aware of the other PELs that support the same VPN Subnet. This is required to establish LSPs across the provider network for the VPN Subnet.

There are several ways to accomplish the distribution of the VPN information:

- Static provisioning
- OSPF opaque LSAs;
- TCP connections;
- BGP-4

Regardless of the distribution mechanism, the information that is distributed is the PEL provider IP address and a list of VPN records. Each VPN record is a VPN Id followed by a list of IP address/prefix pairs. This information is referred to as the VPN subnet information.

Other information that may be part of the VPN subnet information is a QOS value and a status flag. The status flag would indicate if the subnet is being added or withdrawn.

3.2.1 Static provisioning

Each PEL that has a connection to a VPN subnet can be provisioned with VPN subnet information from other PELs that have a connection to the same subnet.

3.2.2 OSPF Opaque LSAs Option

With opaque LSAs, the VPN subnet information is put into an opaque LSA and flooded throughout the OSPF AS. This information is delivered, reliably, to every other node via the normal LSA flooding mechanisms. The amount of information distributed in a single LSA (all, for a single VPN Id, for a single VPN subnet) is left for further study.

3.2.3 TCP connections/BGP Options

The TCP connection option allows for a TCP connection to be established between a PEL and all other PELs that support the same set of VPN subnets. The VPN information would be transmitted reliably across the TCP connections to the PEL peers. This option would require that the IP address of each PEL peer be provisioned, however, it provides an option that is independent of the layer 3 routing protocol(s) running in the provider network.

BGP-4, could also be modified to carry the VPN information. BGP-4 would require a new opaque update type in which it would carry the VPN information.

3.2.4 Withdrawal of VPN subnet information.

If an instance of a VPN subnet on a PEL is operationally or administratively disabled or deleted, the withdrawal of the VPN subnet information is distributed through the provider network using the same mechanism used to distribute new VPN subnet information. The format of a withdrawal message is left for further study. The withdrawal of an instance of VPN subnet information from a PEL will cause the removal of the LSPs that go to that VPN subnet instance on that PEL.

3.3 Establishment of VPN Subnet LSPs

VPN subnet LSPs are created when a PEL learns, via one of the distribution mechanism described in 3.2, that it has a VPN subnet in common with some other PEL in the provider network. Two types of LSPs are created; unicast LSPs and multicast LSPs.

3.3.1 Creation of Unicast LSPs

When a PEL receives new VPN information, it determines if any LSPs

need to be established.

First, the PEL determines if it has any VPNs in common with the new list. If so, it checks to see if it has any VPN subnets in common. If there are, LSPs are triggered for each of the IP addresses that are members of the subnets.

In Figure 3, the creation of LSPs is triggered when PEL X learns that PEL Y supports a common VPN subnet.

Using the example below, an LSP will be established from PNL B to PEL X. LDP then continues to establish the LSP from X to Y. At Y, the LSP is spliced onto the LSP that was created when the VPN subnet for PNL A was provisioned.

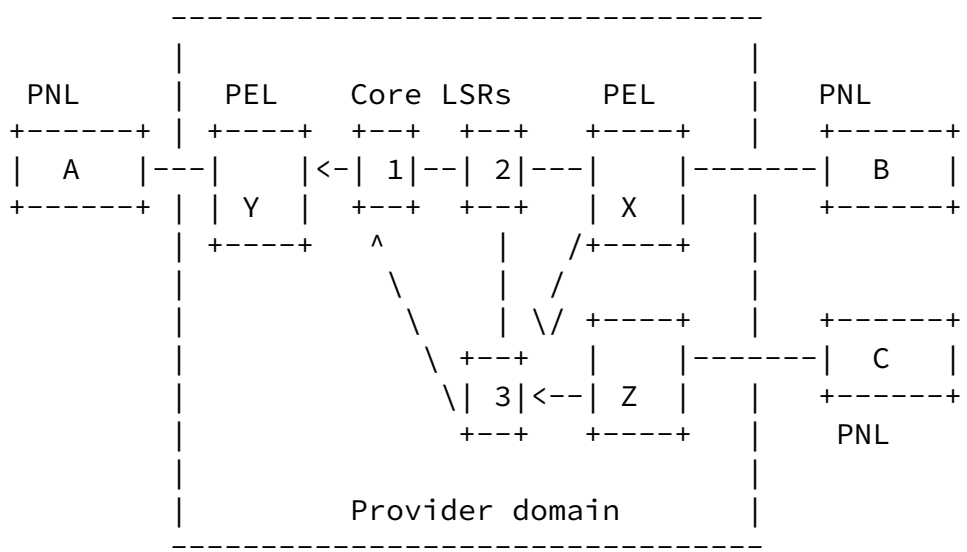

Figure 3. Unicast LSP Setup

Downstream label allocation is used from the PELs (leafs of the multi-point to point tree) to the PNL. Upstream on demand label allocation is used by the PEL (root of the mpt-to-pt tree) and its connected PNL.

The LSP that is created is a unidirectional LSP that carries data from PNL B to PNL A. Within the provider network, the LSP can be established along the best hop route or an explicitly provisioned route. If during the establishment of a best hop LSP, another LSP is encountered that goes to the same destination for the same VPN subnet, the LSPs can be merged. For example, when Z tries to establish an LSP to Y, an existing LSP to Y for the given VPN subnet will be encountered on core router 3. The LSP will be merged at that point.

3.5 Creation of Multicast LSPs

An emulated LAN must be able to multicast certain packets (Hellos, Routing Updates) across the LAN. This draft describes three options for providing this capability.

1> A single bi-directional multi-point to multi-point LSP

2> A set of unidirectional point to multi-point LSPs

3> No multicast LSP is established. VSI interface is responsible for copying and sending multicast packets on all outgoing unicast LSPs.

With option 1, when a PEL (e.g. X) learns of the existence of another PEL (e.g. Y) which supports a VPN subnet which it supports, the creation of both unicast and multicast LSPs are initiated. The multicast LSP is a bi-directional LSP that can follow either the next best hop route or an explicit route. If, during the creation of a next best hop multicast LSP, an existing multicast LSP is encountered for the same VPN, the LSP may be merged.

Even though a merge point is encountered during the creation of a multi-point to multi-point LSP, LDP must continue through to the

destination PNL in case the multicast LSP requires a new branch to reach the destination.

Option 2 is simply a less efficient version of option one, at least in terms of label consumption. In this case a point to multi-point LSP is established from each PNL to all other PNLs for the VPN subnet. Again, they are established at the same time as the unicast LSPs.

Option 3 is the least expensive in terms of label consumption and most expensive in terms of bandwidth and PNL/PEL resources. When the VSI media has a multicast packet to send it copies and sends the packet on each outgoing label for the VSI.

Changes required to LDP to support multicast LSPs is left for further study.

3.6 Layer 3 Modeling of VSI

For each VSI on a PNL there will be one multicast LSP, one incoming LSP and N-1 outgoing LSPs where N is the number of PNLs in the VPN subnet. The incoming label will be viewed by layer 3 as the MAC address for the interface. The outgoing labels will be viewed as

destination MAC addresses for all of the peer routers on the VSI. The multicast LSP will be viewed as the viewed as the multicast MAC address.

3.7 Layer 3 to Layer 2 address mapping

Two methods of mapping layer 3 to layer 2 addresses for a VSI interface are proposed. The first is the distribution of the layer 3 information learned on a PEL for a given VPN subnet into the PNL. This information is injected into the ARP table on the PNL. The second is a modified ARP protocol run between the PNLs on the VPN subnet.

When a PEL learns VPN information from other PELs, it learns the VSI IP addresses that belong to VPN subnets. The PEL then triggers LDP to establish an LSP form the PNL to the PEL to reach that peer IP address. Once the LSP is established, the mapping, IP address to label is known. This information is then propagated into the PNL

where it can be injected into the ARP table. It may be possible to use LDP on the PNL side to learn the mapping. The details of this mechanism are left for further study.

The other option is to use a modified ARP that runs across the VPN subnet. This would be similar to Inverse ARP in that when a new outgoing MAC label is enabled an ARP request is sent across that label. The receiver of the ARP request would put their own VSI IP address in the ARP response packet and send the packet.

The local significance of labels and multipoint to point LSPs provide an additional twist. The ARP response packet may need to be sent on the multicast path. An ARP request has the sender's IP address in the packet. If the receiver of an ARP request had already resolved the mapping of the sender's IP address to MAC label, the response can be sent on that unicast LSP, otherwise the response must be sent on the multicast LSP.

3.8 PNL Routing and Forwarding

Once the mapping for next hop IP address to MAC label is established, normal IP routing and forwarding can take place between the PNLs For each destination IP address that a PNL can send to, its forwarding table will contain an entry which contains the exit port, the next hop IP address to which the packet is to be sent and the MAC address/label for that next hop IP address.

4. Extending MPLS into the VPN Member's Network

The private network could run MPLS across the VPN by forming LDP

peers with other PNLs on the logical LAN and using a shim in the packet header to identify MPLS flows.

5. Summary

This internet draft presents a VPN architecture over MPLS networks. It addresses the three basic functions required to establish VPNs over MPLS. Using an emulated LAN model for connectivity across the provider network, simplifies the configuration and management coordination effort between the service provider and the VPN.

6. Security Considerations

One of the major functions of VPN is being able to provide both data privacy and addressing privacy for users [2]. The architecture proposed in this draft comes with built-in security which is robust under dynamic environment.

6.1 User Data Privacy and User Address Privacy

Both user data privacy and user address privacy are achieved by assigning different VPN identifier to different VPN and building a separate logical network for each VPN. These logical networks may share the same physical connections. But as far as users are concerned, they won't see each other at all. The exceptional case will be one user participate in multiple VPNs. But that would be a configuration issue.

6.2 Service Provider Security

Due to the emulated LAN model adopted in this architecture, each user won't see the service provider's network at all. i.e. the service provider's network is transparent to users. The latter case indicates that users can even have the same address space as the service provider's.

6.3 IP SEC

Since the original VPN IP addresses can be transported across the provider network IP SEC functionality is not impacted. One benefit provided by this mode is IP SEC can run in transport as opposed to tunnel mode reducing bandwidth consumption across the provider network.

7. Intellectual Property Considerations

Nortel may seek patent or other intellectual property protection for some of all of the technologies disclosed in this document. If any

standards arising from this document are or become protected by one or more patents assigned to Nortel, Nortel intends to disclose those patents and license them on reasonable and non-discriminatory terms.

8. Acknowledgment

The authors would like to acknowledge the valuable review and comments of Jerry Wu, Stephen Shew, Ian Duncan, and Scott Pegrum.

9. References

[1] Rosen et al, "Multiprotocol Label Switching Architecture", draft-ietf-mpls-arch-01.txt, March 1998.

[2] J. Heinanen, et. al, "VPN support with MPLS", <draft-heinanen-mpls-vpn-01.txt>, March 1998.

[3] Anderson, et. al., "Label Distribution Protocol", draft-mpls-ldp-00.txt, March 1998.

10. Authors' Addresses

Dwight Jamieson
Nortel (Northern Telecom), Ltd.
PO Box 3511 Station C
Ottawa ON K1Y 4H7
Canada

EMail: djamies@Nortel.ca

Bilel Jamoussi
Nortel (Northern Telecom), Ltd.
PO Box 3511 Station C
Ottawa ON K1Y 4H7
Canada

EMail: jamoussi@Nortel.ca

Gregory Wright
Nortel (Northern Telecom), Ltd.
PO Box 3511 Station C
Ottawa ON K1Y 4H7
Canada

EMail: gwright@Nortel.ca

Paul Beaubien
Nortel (Northern Telecom), Ltd.

PO Box 3511 Station C
Ottawa ON K1Y 4H7
Canada

EMail: beaubien@Nortel.ca