Internet Engineering Task Force (IETF)
Request for Comments: 6074
Category: Standards Track
ISSN: 2070-1721

E. Rosen
B. Davie
Cisco Systems, Inc.
V. Radoaca
Alcatel-Lucent
W. Luo

January 2011

Provisioning, Auto-Discovery, and Signaling in Layer 2 Virtual Private Networks (L2VPNs)

Abstract

Provider Provisioned Layer 2 Virtual Private Networks (L2VPNs) may have different "provisioning models", i.e., models for what information needs to be configured in what entities. Once configured, the provisioning information is distributed by a "discovery process". When the discovery process is complete, a signaling protocol is automatically invoked to set up the mesh of pseudowires (PWs) that form the (virtual) backbone of the L2VPN. This document specifies a number of L2VPN provisioning models, and further specifies the semantic structure of the endpoint identifiers required by each model. It discusses the distribution of these identifiers by the discovery process, especially when discovery is based on the Border Gateway Protocol (BGP). It then specifies how the endpoint identifiers are carried in the two signaling protocols that are used to set up PWs, the Label Distribution Protocol (LDP), and the Layer 2 Tunneling Protocol version 3 (L2TPv3).

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc6074.

Table of Contents

1. Introduction

[RFC4664] describes a number of different ways in which sets of pseudowires may be combined together into "Provider Provisioned Layer 2 VPNs" (L2 PPVPNs, or L2VPNs), resulting in a number of different kinds of L2VPN. Different kinds of L2VPN may have different "provisioning models", i.e., different models for what information needs to be configured in what entities. Once configured, the provisioning information is distributed by a "discovery process", and once the information is discovered, the signaling protocol is automatically invoked to set up the required pseudowires. The semantics of the endpoint identifiers that the signaling protocol uses for a particular type of L2VPN are determined by the provisioning model. That is, different kinds of L2VPN, with different provisioning models, require different kinds of endpoint identifiers. This document specifies a number of L2VPN provisioning models and specifies the semantic structure of the endpoint identifiers required for each provisioning model.

Either LDP (as specified in [RFC5036] and extended in [RFC4447]) or L2TP version 3 (as specified in [RFC3931] and extended in [RFC4667]) can be used as signaling protocols to set up and maintain PWs [RFC3985]. Any protocol that sets up connections must provide a way for each endpoint of the connection to identify the other; each PW signaling protocol thus provides a way to identify the PW endpoints. Since each signaling protocol needs to support all the different kinds of L2VPN and provisioning models, the signaling protocol must have a very general way of representing endpoint identifiers, and it is necessary to specify rules for encoding each particular kind of endpoint identifier into the relevant fields of each signaling protocol. This document specifies how to encode the endpoint identifiers of each provisioning model into the LDP and L2TPv3 signaling protocols.

We make free use of terminology from [RFC3985], [RFC4026], [RFC4664], and [RFC5659] -- in particular, the terms "Attachment Circuit", "pseudowire", "PE" (provider edge), "CE" (customer edge), and "multi-segment pseudowire".

Section 2 provides an overview of the relevant aspects of [RFC4447] and [RFC4667].

Section 3 details various provisioning models and relates them to the signaling process and to the discovery process. The way in which the signaling mechanisms can be integrated with BGP-based auto-discovery is covered in some detail.

Section 4 explains how the procedures for discovery and signaling can be applied in a multi-AS environment and outlines several options for the establishment of multi-AS L2VPNs.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119]

2. Signaling Protocol Framework

2.1. Endpoint Identification

Per [RFC4664], a pseudowire can be thought of as a relationship between a pair of "Forwarders". In simple instances of Virtual Private Wire Service (VPWS), a Forwarder binds a pseudowire to a single Attachment Circuit, such that frames received on the one are sent on the other, and vice versa. In Virtual Private LAN Service (VPLS), a Forwarder binds a set of pseudowires to a set of Attachment Circuits; when a frame is received from any member of that set, a MAC (Media Access Control) address table is consulted (and various 802.1d procedures executed) to determine the member or members of that set on which the frame is to be transmitted. In more complex scenarios, Forwarders may bind PWs to PWs, thereby "splicing" two PWs together; this is needed, e.g., to support distributed VPLS and some inter-AS scenarios.

In simple VPWS, where a Forwarder binds exactly one PW to exactly one Attachment Circuit, a Forwarder can be identified by identifying its Attachment Circuit. In simple VPLS, a Forwarder can be identified by identifying its PE device and its VPN.

To set up a PW between a pair of Forwarders, the signaling protocol must allow the Forwarder at one endpoint to identify the Forwarder at the other. In [RFC4447], the term "Attachment Identifier", or "AI", is used to refer to a quantity whose purpose is to identify a Forwarder. In [RFC4667], the term "Forwarder Identifier" is used for the same purpose. In the context of this document, "Attachment Identifier" and "Forwarder Identifier" are used interchangeably.

[RFC4447] specifies two Forwarding Equivalence Class (FEC) elements that can be used when setting up pseudowires, the PWid FEC element, and the Generalized ID FEC element. The PWid FEC element carries only one Forwarder identifier; it can be thus be used only when both forwarders have the same identifier, and when that identifier can be coded as a 32-bit quantity. The Generalized ID FEC element carries two Forwarder identifiers, one for each of the two Forwarders being

connected. Each identifier is known as an Attachment Identifier, and a signaling message carries both a "Source Attachment Identifier" (SAI) and a "Target Attachment Identifier" (TAI).

The Generalized ID FEC element also provides some additional structuring of the identifiers. It is assumed that the SAI and TAI will sometimes have a common part, called the "Attachment Group Identifier" (AGI), such that the SAI and TAI can each be thought of as the concatenation of the AGI with an "Attachment Individual Identifier" (AII). So the pair of identifiers is encoded into three fields: AGI, Source AII (SAII), and Target AII (TAII). The SAI is the concatenation of the AGI and the SAII, while the TAI is the concatenation of the AGI and the TAII.

Similarly, [RFC4667] allows using one or two Forwarder Identifiers to set up pseudowires. If only the target Forwarder Identifier is used in L2TP signaling messages, both the source and target Forwarders are assumed to have the same value. If both the source and target Forwarder Identifiers are carried in L2TP signaling messages, each Forwarder uses a locally significant identifier value.

The Forwarder Identifier in [RFC4667] is an equivalent term to Attachment Identifier in [RFC4447]. A Forwarder Identifier also consists of an Attachment Group Identifier and an Attachment Individual Identifier. Unlike the Generalized ID FEC element, the AGI and AII are carried in distinct L2TP Attribute-Value Pairs (AVPs). The AGI is encoded in the AGI AVP, and the SAII and TAII are encoded in the Local End ID AVP and the Remote End ID AVP, respectively. The source Forwarder Identifier is the concatenation of the AGI and SAII, while the target Forwarder Identifier is the concatenation of the AGI and TAII.

In applications that group sets of PWs into "Layer 2 Virtual Private Networks", the AGI can be thought of as a "VPN Identifier".

It should be noted that while different forwarders support different applications, the type of application (e.g., VPLS vs. VPWS) cannot necessarily be inferred from the forwarders' identifiers. A router receiving a signaling message with a particular TAI will have to be able to determine which of its local forwarders is identified by that TAI, and to determine the application provided by that forwarder. But other nodes may not be able to infer the application simply by inspection of the signaling messages.

In this document, some further structure of the AGI and AII is proposed for certain L2VPN applications. We note that [RFC4447] defines a TLV structure for AGI and AII fields. Thus, an operator who chooses to use the AII structure defined here could also make use

of different AGI or AII types if he also wanted to use a different structure for these identifiers for some other application. For example, the long prefix type of [RFC5003] could be used to enable the communication of administrative information, perhaps combined with information learned during auto-discovery.

2.2. Creating a Single Bidirectional Pseudowire

In any form of LDP-based signaling, each PW endpoint must initiate the creation of a unidirectional LSP. A PW is a pair of such LSPs. In most of the L2VPN provisioning models, the two endpoints of a given PW can simultaneously initiate the signaling for it. They must therefore have some way of determining when a given pair of LSPs are intended to be associated together as a single PW.

The way in which this association is done is different for the various different L2VPN services and provisioning models. The details appear in later sections.

L2TP signaling inherently establishes a bidirectional session that carries a PW between two PW endpoints. The two endpoints can also simultaneously initiate the signaling for a given PW. It is possible that two PWs can be established for a pair of Forwarders.

In order to avoid setting up duplicated pseudowires between two Forwarders, each PE must be able to independently detect such a pseudowire tie. The procedures of detecting a pseudowire tie are described in [RFC4667].

2.3. Attachment Identifiers and Forwarders

Every Forwarder in a PE must be associated with an Attachment Identifier (AI), either through configuration or through some algorithm. The Attachment Identifier must be unique in the context of the PE router in which the Forwarder resides. The combination <PE router, AI> must be globally unique.

As specified in [RFC4447], the Attachment Identifier may consist of an Attachment Group Identifier (AGI) plus an Attachment Individual Identifier (AII). In the context of this document, an AGI may be thought of as a VPN-ID, or some attribute that is shared by all the Attachment Circuits that are allowed to be connected.

It is sometimes helpful to consider a set of attachment circuits at a single PE to belong to a common "pool". For example, a set of attachment circuits that connect a single CE to a given PE may be considered a pool. The use of pools is described in detail in Section 3.3.

The details for how to construct the AGI and AII fields identifying the pseudowire endpoints in particular provisioning models are discussed later in this document.

We can now consider an LSP for one direction of a pseudowire to be identified by:

o <PE1, <AGI, AII1>, PE2, <AGI, AII2>>

and the LSP in the opposite direction of the pseudowire will be identified by:

o <PE2, <AGI, AII2>, PE1, <AGI, AII1>>

A pseudowire is a pair of such LSPs. In the case of using L2TP signaling, these refer to the two directions of an L2TP session.

When a signaling message is sent from PE1 to PE2, and PE1 needs to refer to an Attachment Identifier that has been configured on one of its own Attachment Circuits (or pools), the Attachment Identifier is called a "Source Attachment Identifier". If PE1 needs to refer to an Attachment Identifier that has been configured on one of PE2's Attachment Circuits (or pools), the Attachment Identifier is called a "Target Attachment Identifier". (So an SAI at one endpoint is a TAI at the remote endpoint, and vice versa.)

In the signaling protocol, we define encodings for the following three fields:

o Attachment Group Identifier (AGI)

o Source Attachment Individual Identifier (SAII)

o Target Attachment Individual Identifier (TAII)

If the AGI is non-null, then the SAI consists of the AGI together with the SAII, and the TAI consists of the TAII together with the AGI. If the AGI is null, then the SAII and TAII are the SAI and TAI, respectively.

The intention is that the PE that receives an LDP Label Mapping message or an L2TP Incoming Call Request (ICRQ) message containing a TAI will be able to map that TAI uniquely to one of its Attachment Circuits (or pools). The way in which a PE maps a TAI to an Attachment Circuit (or pool) should be a local matter (including the choice of whether to use some or all of the bytes in the TAI for the mapping). So as far as the signaling procedures are concerned, the TAI is really just an arbitrary string of bytes, a "cookie".

3. Applications

In this section, we specify the way in which the pseudowire signaling using the notion of source and target Forwarder is applied for a number of different applications. For some of the applications, we specify the way in which different provisioning models can be used. However, this is not meant to be an exhaustive list of the applications, or an exhaustive list of the provisioning models that can be applied to each application.

3.1. Individual Point-to-Point Pseudowires

The signaling specified in this document can be used to set up individually provisioned point-to-point pseudowires. In this application, each Forwarder binds a single PW to a single Attachment Circuit. Each PE must be provisioned with the necessary set of Attachment Circuits, and then certain parameters must be provisioned for each Attachment Circuit.

3.1.1. Provisioning Models

3.1.1.1. Double-Sided Provisioning

In this model, the Attachment Circuit must be provisioned with a local name, a remote PE address, and a remote name. During signaling, the local name is sent as the SAII, the remote name as the TAII, and the AGI is null. If two Attachment Circuits are to be connected by a PW, the local name of each must be the remote name of the other.

Note that if the local name and the remote name are the same, the PWid FEC element can be used instead of the Generalized ID FEC element in the LDP-based signaling.

With L2TP signaling, the local name is sent in Local End ID AVP, and the remote name in Remote End ID AVP. The AGI AVP is optional. If present, it contains a zero-length AGI value. If the local name and the remote name are the same, Local End ID AVP can be omitted from L2TP signaling messages.

3.1.1.2. Single-Sided Provisioning with Discovery

In this model, each Attachment Circuit must be provisioned with a local name. The local name consists of a VPN-ID (signaled as the AGI) and an Attachment Individual Identifier that is unique relative to the AGI. If two Attachment Circuits are to be connected by a PW, only one of them needs to be provisioned with a remote name (which of

course is the local name of the other Attachment Circuit). Neither needs to be provisioned with the address of the remote PE, but both must have the same VPN-ID.

As part of an auto-discovery procedure, each PE advertises its <VPN-id, local AII> pairs. Each PE compares its local <VPN-id, remote AII> pairs with the <VPN-id, local AII> pairs advertised by the other PEs. If PE1 has a local <VPN-id, remote AII> pair with value <V, fred>, and PE2 has a local <VPN-id, local AII> pair with value <V, fred>, PE1 will thus be able to discover that it needs to connect to PE2. When signaling, it will use "fred" as the TAII, and will use V as the AGI. PE1's local name for the Attachment Circuit is sent as the SAII.

The primary benefit of this provisioning model when compared to Double-Sided Provisioning is that it enables one to move an Attachment Circuit from one PE to another without having to reconfigure the remote endpoint. However, compared to the approach described in Section 3.3 below, it imposes a greater burden on the discovery mechanism, because each Attachment Circuit's name must be advertised individually (i.e., there is no aggregation of Attachment Circuit names in this simple scheme).

3.1.2. Signaling

The LDP-based signaling follows the procedures specified in [RFC4447]. That is, one PE (PE1) sends a Label Mapping message to another PE (PE2) to establish an LSP in one direction. If that message is processed successfully, and there is not yet an LSP for the pseudowire in the opposite (PE1->PE2) direction, then PE2 sends a Label Mapping message to PE1.

In addition to the procedures of [RFC4447], when a PE receives a Label Mapping message, and the TAI identifies a particular Attachment Circuit that is configured to be bound to a point-to-point PW, then the following checks must be made.

If the Attachment Circuit is already bound to a pseudowire (including the case where only one of the two LSPs currently exists), and the remote endpoint is not PE1, then PE2 sends a Label Release message to PE1, with a Status Code meaning "Attachment Circuit bound to different PE", and the processing of the Mapping message is complete.

If the Attachment Circuit is already bound to a pseudowire (including the case where only one of the two LSPs currently exists), but the AI at PE1 is different than that specified in the AGI/SAII fields of the Mapping message then PE2 sends a Label Release message to PE1, with a

Status Code meaning "Attachment Circuit bound to different remote Attachment Circuit", and the processing of the Mapping message is complete.

Similarly, with the L2TP-based signaling, when a PE receives an ICRQ message, and the TAI identifies a particular Attachment Circuit that is configured to be bound to a point-to-point PW, it performs the following checks.

If the Attachment Circuit is already bound to a pseudowire, and the remote endpoint is not PE1, then PE2 sends a Call Disconnect Notify (CDN) message to PE1, with a Status Code meaning "Attachment Circuit bound to different PE", and the processing of the ICRQ message is complete.

If the Attachment Circuit is already bound to a pseudowire, but the pseudowire is bound to a Forwarder on PE1 with the AI different than that specified in the SAI fields of the ICRQ message, then PE2 sends a CDN message to PE1, with a Status Code meaning "Attachment Circuit bound to different remote Attachment Circuit", and the processing of the ICRQ message is complete.

These errors could occur as the result of misconfigurations.

3.2. Virtual Private LAN Service

In the VPLS application [RFC4762], the Attachment Circuits can be thought of as LAN interfaces that attach to "virtual LAN switches", or, in the terminology of [RFC4664], "Virtual Switching Instances" (VSIs). Each Forwarder is a VSI that attaches to a number of PWs and a number of Attachment Circuits. The VPLS service requires that a single pseudowire be created between each pair of VSIs that are in the same VPLS. Each PE device may have multiple VSIs, where each VSI belongs to a different VPLS.

3.2.1. Provisioning

Each VPLS must have a globally unique identifier, which in [RFC4762] is referred to as the VPLS identifier (or VPLS-id). Every VSI must be configured with the VPLS-id of the VPLS to which it belongs.

Each VSI must also have a unique identifier, which we call a VSI-ID. This can be formed automatically by concatenating its VPLS-id with an IP address of its PE router. (Note that the PE address here is used only as a form of unique identifier; a service provider could choose to use some other numbering scheme if that was desired, as long as

each VSI is assigned an identifier that is unique within the VPLS instance. See Section 4.4 for a discussion of the assignment of identifiers in the case of multiple providers.)

3.2.2. Auto-Discovery

3.2.2.1. BGP-Based Auto-Discovery

This section specifies how BGP can be used to discover the information necessary to build VPLS instances.

When BGP-based auto-discovery is used for VPLS, the AFI/SAFI (Address Family Identifier / Subsequent Address Family Identifier) [RFC4760] will be:

- An AFI (25) for L2VPN. (This is the same for all L2VPN schemes.)
- A SAFI (65) specifically for an L2VPN service whose pseudowires are set up using the procedures described in the current document.

See Section 6 for further discussion of AFI/SAFI assignment.

In order to use BGP-based auto-discovery, there must be at least one globally unique identifier associated with a VPLS, and each such identifier must be encodable as an 8-byte Route Distinguisher (RD). Any method of assigning one or more unique identifiers to a VPLS and encoding each of them as an RD (using the encoding techniques of [RFC4364]) will do.

Each VSI needs to have a unique identifier that is encodable as a BGP Network Layer Reachability Information (NLRI). This is formed by prepending the RD (from the previous paragraph) to an IP address of the PE containing the VSI. Note that the role of this address is simply as a readily available unique identifier for the VSIs within a VPN; it does not need to be globally routable, but it must be unique within the VPLS instance. An alternate scheme to assign unique identifiers to each VSI within a VPLS instance (e.g., numbering the VSIs of a single VPN from 1 to n) could be used if desired.

When using the procedures described in this document, it is necessary to assign a single, globally unique VPLS-id to each VPLS instance [RFC4762]. This VPLS-id must be encodable as a BGP Extended Community [RFC4360]. As described in Section 6, two Extended Community subtypes are defined by this document for this purpose. The Extended Community MUST be transitive.

The first Extended Community subtype is a Two-octet AS Specific Extended Community. The second Extended Community subtype is an IPv4 Address Specific Extended Community. The encoding of such Communities is defined in [RFC4360]. These encodings ensure that a service provider can allocate a VPLS-id without risk of collision with another provider. However, note that coordination of VPLS-ids among providers is necessary for inter-provider L2VPNs, as described in Section 4.4.

Each VSI also needs to be associated with one or more Route Target (RT) Extended Communities. These control the distribution of the NLRI, and hence will control the formation of the overlay topology of pseudowires that constitutes a particular VPLS.

Auto-discovery proceeds by having each PE distribute, via BGP, the NLRI for each of its VSIs, with itself as the BGP next hop, and with the appropriate RT for each such NLRI. Typically, each PE would be a client of a small set of BGP route reflectors, which would redistribute this information to the other clients.

If a PE receives a BGP update from which any of the elements specified above is absent, the update should be ignored.

If a PE has a VSI with a particular RT, it can then import all the NLRIs that have that same RT, and from the BGP next hop attribute of these NLRI it will learn the IP addresses of the other PE routers which have VSIs with the same RT. The considerations in Section 4.3.3 of [RFC4364] on the use of route reflectors apply.

If a particular VPLS is meant to be a single fully connected LAN, all its VSIs will have the same RT, in which case the RT could be (though it need not be) an encoding of the VPN-id. A VSI can be placed in multiple VPLSes by assigning it multiple RTs.

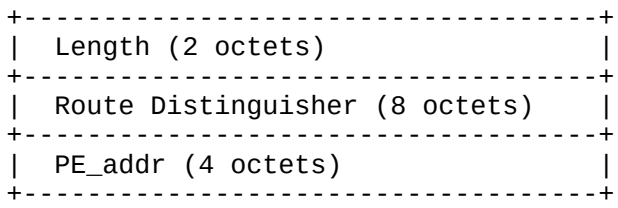
Note that hierarchical VPLS can be set up by assigning multiple RTs to some of the VSIs; the RT mechanism allows one to have complete control over the pseudowire overlay that constitutes the VPLS topology.

If Distributed VPLS (described in Section 3.5) is deployed, only the Network-facing PEs (N-PEs) participate in BGP-based auto-discovery. This means that an N-PE would need to advertise reachability to each of the VSIs that it supports, including those located in User-facing PEs (U-PEs) to which it is connected. To create a unique identifier for each such VSI, an IP address of each U-PE combined with the RD for the VPLS instance could be used.

In summary, the BGP advertisement for a particular VSI at a given PE will contain:

- an NLRI of AFI = L2VPN, SAFI = VPLS, encoded as RD:PE_addr
- a BGP next hop equal to the loopback address of the PE
- an Extended Community Attribute containing the VPLS-id
- an Extended Community Attribute containing one or more RTs.

See Section 6 for discussion of the AFI and SAFI values. The format for the NLRI encoding is:



Note that this advertisement is quite similar to the NLRI format defined in [RFC4761], the main difference being that [RFC4761] also includes a label block in the NLRI. Interoperability between the VPLS scheme defined here and that defined in [RFC4761] is beyond the scope of this document.

3.2.3. Signaling

It is necessary to create Attachment Identifiers that identify the VSIs. In the preceding section, a VSI-ID was encoded as RD:PE_addr, and the VPLS-id was carried in a BGP Extended Community. For signaling purposes, this information is encoded as follows. We encode the VPLS-id in the AGI field, and place the PE_addr (or, more precisely, the VSI-ID that was contained in the NLRI in BGP, minus the RD) in the TAII field. The combination of AGI and TAII is sufficient to fully specify the VSI to which this pseudowire is to be connected, in both single AS and inter-AS environments. The SAII MUST be set to the PE_addr of the sending PE (or, more precisely, the VSI-ID, without the RD, of the VSI associated with this VPLS in the sending PE) to enable signaling of the reverse half of the PW if needed.

The structure of the AGI and AII fields for the Generalized ID FEC in LDP is defined in [RFC4447]. The AGI field in this case consists of a Type of 1, a length field of value 8, and the 8 bytes of the

VPLS-id. The AIIs consist of a Type of 1, a length field of value 4, followed by the 4-byte PE address (or other 4-byte identifier). See Section 6 for discussion of the AGI and AII Type assignment.

The encoding of the AGI and AII in L2TP is specified in [RFC4667].

Note that it is not possible using this technique to set up more than one PW per pair of VSIs.

3.2.4. Pseudowires as VPLS Attachment Circuits

It is also possible using this technique to set up a PW that attaches at one endpoint to a VSI, but at the other endpoint only to an Attachment Circuit. There may be more than one PW terminating on a given VSI, which must somehow be distinguished, so each PW must have an SAII that is unique relative to the VSI-ID.

3.3. Colored Pools: Full Mesh of Point-to-Point Pseudowires

The "Colored Pools" model of operation provides an automated way to deliver VPWS. In this model, each PE may contain several pools of Attachment Circuits, each pool associated with a particular VPN. A PE may contain multiple pools per VPN, as each pool may correspond to a particular CE device. It may be desired to create one pseudowire between each pair of pools that are in the same VPN; the result would be to create a full mesh of CE-CE Virtual Circuits for each VPN.

3.3.1. Provisioning

Each pool is configured, and associated with:

- a set of Attachment Circuits;
- a "color", which can be thought of as a VPN-id of some sort;
- a relative pool identifier, which is unique relative to the color.

[Note: depending on the technology used for Attachment Circuits (ACs), it may or may not be necessary to provision these circuits as well. For example, if the ACs are frame relay circuits, there may be some separate provisioning system to set up such circuits. Alternatively, "provisioning" an AC may be as simple as allocating an unused VLAN ID on an interface and communicating the choice to the customer. These issues are independent of the procedures described in this document.]

The pool identifier and color, taken together, constitute a globally unique identifier for the pool. Thus, if there are n pools of a given color, their pool identifiers can be (though they do not need to be) the numbers 1-n.

The semantics are that a pseudowire will be created between every pair of pools that have the same color, where each such pseudowire will be bound to one Attachment Circuit from each of the two pools.

If each pool is a set of Attachment Circuits leading to a single CE device, then the Layer 2 connectivity among the CEs is controlled by the way the colors are assigned to the pools. To create a full mesh, the "color" would just be a VPN-id.

Optionally, a particular Attachment Circuit may be configured with the relative pool identifier of a remote pool. Then, that Attachment Circuit would be bound to a particular pseudowire only if that pseudowire's remote endpoint is the pool with that relative pool identifier. With this option, the same pairs of Attachment Circuits will always be bound via pseudowires.

3.3.2. Auto-Discovery

3.3.2.1. BGP-Based Auto-Discovery

This section specifies how BGP can be used to discover the information necessary to build VPWS instances.

When BGP-based auto-discovery is used for VPWS, the AFI/SAFI will be:

- An AFI specified by IANA for L2VPN. (This is the same for all L2VPN schemes.)
- A SAFI specified by IANA specifically for an L2VPN service whose pseudowires are set up using the procedures described in the current document.

See Section 6 for further discussion of AFI/SAFI assignment.

In order to use BGP-based auto-discovery, there must be one or more unique identifiers associated with a particular VPWS instance. Each identifier must be encodable as an RD (Route Distinguisher). The globally unique identifier of a pool must be encodable as NLRI; the pool identifier, which we define to be a 4-byte quantity, is appended to the RD to create the NLRI.

When using the procedures described in this document, it is necessary to assign a single, globally unique identifier to each VPWS instance.

This identifier must be encodable as a BGP Extended Community [RFC4360]. As described in Section 6, two Extended Community subtypes are defined by this document for this purpose. The Extended Community MUST be transitive.

The first Extended Community subtype is a Two-octet AS Specific Extended Community. The second Extended Community subtype is an IPv4 Address Specific Extended Community. The encoding of such Communities is defined in [RFC4360]. These encodings ensure that a service provider can allocate a VPWS identifier without risk of collision with another provider. However, note that co-ordination of VPWS identifiers among providers is necessary for inter-provider L2VPNs, as described in Section 4.4.

Each pool must also be associated with an RT (route target), which may also be an encoding of the color. If the desired topology is a full mesh of pseudowires, all pools may have the same RT. See Section 3.4 for a discussion of other topologies.

Auto-discovery proceeds by having each PE distribute, via BGP, the NLRI for each of its pools, with itself as the BGP next hop, and with the RT that encodes the pool's color. If a given PE has a pool with a particular color (RT), it must receive, via BGP, all NLRI with that same color (RT). Typically, each PE would be a client of a small set of BGP route reflectors, which would redistribute this information to the other clients.

If a PE receives a BGP update from which any of the elements specified above is absent, the update should be ignored.

If a PE has a pool with a particular color, it can then receive all the NLRI that have that same color, and from the BGP next hop attribute of these NLRI will learn the IP addresses of the other PE routers that have pools switches with the same color. It also learns the unique identifier of each such remote pool, as this is encoded in the NLRI. The remote pool's relative identifier can be extracted from the NLRI and used in the signaling, as specified below.

In summary, the BGP advertisement for a particular pool of attachment circuits at a given PE will contain:

- an NLRI of AFI = L2VPN, SAFI = VPLS, encoded as RD:pool_num;
- a BGP next hop equal to the loopback address of the PE;
- an Extended Community Attribute containing the VPWS identifier;
- an Extended Community Attribute containing one or more RTs.

See Section 6 for discussion of the AFI and SAFI values.

3.3.3. Signaling

The LDP-based signaling follows the procedures specified in [RFC4447]. That is, one PE (PE1) sends a Label Mapping message to another PE (PE2) to establish an LSP in one direction. The address of PE2 is the next-hop address learned via BGP as described above. If the message is processed successfully, and there is not yet an LSP for the pseudowire in the opposite (PE1->PE2) direction, then PE2 sends a Label Mapping message to PE1. Similarly, the L2TPv3-based signaling follows the procedures of [RFC4667]. Additional details on the use of these signaling protocols follow.

When a PE sends a Label Mapping message or an ICRQ message to set up a PW between two pools, it encodes the VPWS identifier (as distributed in the Extended Community Attribute by BGP) as the AGI, the local pool's relative identifier as the SAII, and the remote pool's relative identifier as the TAII.

The structure of the AGI and AII fields for the Generalized ID FEC in LDP is defined in [RFC4447]. The AGI field in this case consists of a Type of 1, a length field of value 8, and the 8 bytes of the VPWS identifier. The TAII consists of a Type of 1, a length field of value 4, followed by the 4-byte remote pool number. The SAII consists of a Type of 1, a length field of value 4, followed by the 4-byte local pool number. See Section 6 for discussion of the AGI and AII Type assignment. Note that the VPLS and VPWS procedures defined in this document can make use of the same AGI Type (1) and the same AII Type (1).

The encoding of the AGI and AII in L2TP is specified in [RFC4667].

When PE2 receives a Label Mapping message or an ICRQ message from PE1, and the TAI identifies a pool, and there is already a pseudowire connecting an Attachment Circuit in that pool to an Attachment Circuit at PE1, and the AI at PE1 of that pseudowire is the same as the SAI of the Label Mapping or ICRQ message, then PE2 sends a Label Release or CDN message to PE1, with a Status Code meaning "Attachment Circuit already bound to remote Attachment Circuit". This prevents the creation of multiple pseudowires between a given pair of pools.

Note that the signaling itself only identifies the remote pool to which the pseudowire is to lead, not the remote Attachment Circuit that is to be bound to the pseudowire. However, the remote PE may examine the SAII field to determine which Attachment Circuit should be bound to the pseudowire.

3.4. Colored Pools: Partial Mesh

The procedures for creating a partial mesh of pseudowires among a set of colored pools are substantially the same as those for creating a full mesh, with the following exceptions:

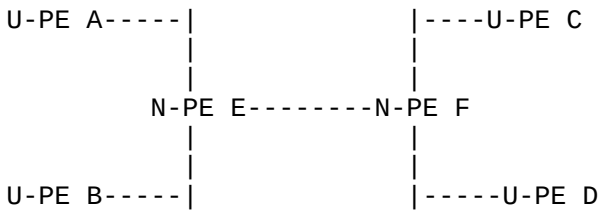
- Each pool is optionally configured with a set of "import RTs" and "export RTs";
- During BGP-based auto-discovery, the pool color is still encoded in the RD, but if the pool is configured with a set of "export RTs", these are encoded in the RTs of the BGP Update messages INSTEAD of the color;
- If a pool has a particular "import RT" value X, it will create a PW to every other pool that has X as one of its "export RTs". The signaling messages and procedures themselves are as in Section 3.3.3.

As a simple example, consider the task of building a hub-and-spoke topology with a single hub. One pool, the "hub" pool, is configured with an export RT of RT_hub and an import RT of RT_spoke. All other pools (the spokes) are configured with an export RT of RT_spoke and an import RT of RT_hub. Thus, the hub pool will connect to the spokes, and vice-versa, but the spoke pools will not connect to each other.

3.5. Distributed VPLS

In Distributed VPLS ([RFC4664]), the VPLS functionality of a PE router is divided among two systems: a U-PE and an N-PE. The U-PE sits between the user and the N-PE. VSI functionality (e.g., MAC address learning and bridging) is performed on the U-PE. A number of U-PEs attach to an N-PE. For each VPLS supported by a U-PE, the U-PE maintains a pseudowire to each of the other U-PEs in the same VPLS. However, the U-PEs do not maintain signaling control connections with each other. Rather, each U-PE has only a single signaling connection, to its N-PE. In essence, each U-PE-to-U-PE pseudowire is composed of three pseudowires spliced together: one from U-PE to N-PE, one from N-PE to N-PE, and one from N-PE to U-PE. In the terminology of [RFC5659], the N-PEs perform the pseudowire switching function to establish multi-segment PWs from U-PE to U-PE.

Consider, for example, the following topology:



where the four U-PEs are in a common VPLS. We now illustrate how PWs get spliced together in the above topology in order to establish the necessary PWs from U-PE A to the other U-PEs.

There are three PWs from A to E. Call these A-E/1, A-E/2, and A-E/3. In order to connect A properly to the other U-PEs, there must be two PWs from E to F (call these E-F/1 and E-F/2), one PW from E to B (E-B/1), one from F to C (F-C/1), and one from F to D (F-D/1).

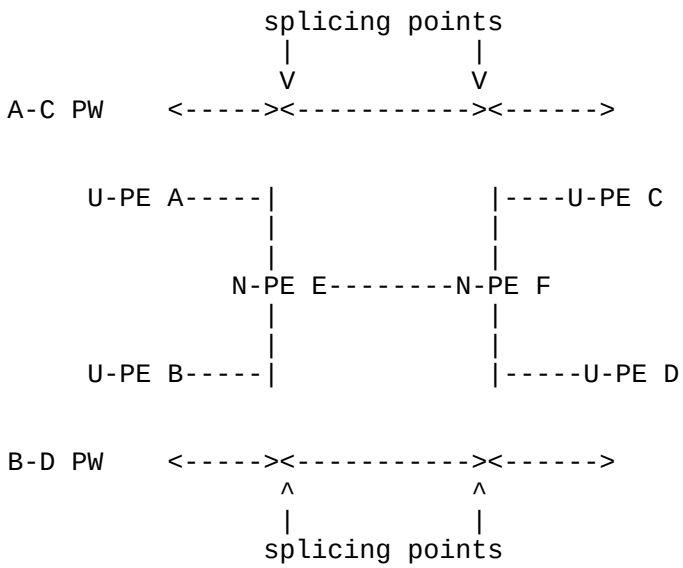
The N-PEs must then splice these pseudowires together to get the equivalent of what the non-distributed VPLS signaling mechanism would provide:

- PW from A to B: A-E/1 gets spliced to E-B/1.
- PW from A to C: A-E/2 gets spliced to E-F/1 gets spliced to F-C/1.
- PW from A to D: A-E/3 gets spliced to E-F/2 gets spliced to F-D/1.

It doesn’t matter which PWs get spliced together, as long as the result is one from A to each of B, C, and D.

Similarly, there are additional PWs that must get spliced together to properly interconnect U-PE B with U-PEs C and D, and to interconnect U-PE C with U-PE D.

The following figure illustrates the PWs from A to C and from B to D. For clarity of the figure, the other four PWs are not shown.



One can see that distributed VPLS does not reduce the number of pseudowires per U-PE, but it does reduce the number of control connections per U-PE. Whether this is worthwhile depends, of course, on what the bottleneck is.

3.5.1. Signaling

The signaling to support Distributed VPLS can be done with the mechanisms described in this document. However, the procedures for VPLS (Section 3.2.3) need some additional machinery to ensure that the appropriate number of PWs are established between the various N-PEs and U-PEs, and among the N-PEs.

At a given N-PE, the directly attached U-PEs in a given VPLS can be numbered from 1 to n. This number identifies the U-PE relative to a particular VPN-id and a particular N-PE. (That is, to uniquely identify the U-PE, the N-PE, the VPN-id, and the U-PE number must be known.)

As a result of configuration/discovery, each U-PE must be given a list of <j, IP address> pairs. Each element in this list tells the U-PE to set up j PWs to the specified IP address. When the U-PE signals to the N-PE, it sets the AGI to the proper-VPN-id, and sets the SAII to the PW number, and sets the TAII to null.

In the above example, U-PE A would be told <3, E>, telling it to set up 3 PWs to E. When signaling, A would set the AGI to the proper VPN-id, and would set the SAII to 1, 2, or 3, depending on which of the three PWs it is signaling.

As a result of configuration/discovery, each N-PE must be given the following information for each VPLS:

- A "Local" list: {<j, IP address>}, where each element tells it to set up j PWs to the locally attached U-PE at the specified address. The number of elements in this list will be n, the number of locally attached U-PEs in this VPLS. In the above example, E would be given the local list: {<3, A>, <3, B>}, telling it to set up 3 PWs to A and 3 to B.

- A local numbering, relative to the particular VPLS and the particular N-PE, of its U-PEs. In the above example, E could be told that U-PE A is 1, and U-PE B is 2.

- A "Remote" list: {<IP address, k>}, telling it to set up k PWs, for each U-PE, to the specified IP address. Each of these IP addresses identifies an N-PE, and k specifies the number of U-PEs at the N-PE that are in the VPLS. In the above example, E would be given the remote list: {<2, F>}. Since N-PE E has 2 U-PEs, this tells it to set up 4 PWs to N-PE F, 2 for each of its E's U-PEs.

The signaling of a PW from N-PE to U-PE is based on the local list and the local numbering of U-PEs. When signaling a particular PW from an N-PE to a U-PE, the AGI is set to the proper VPN-id, and SAII is set to null, and the TAII is set to the PW number (relative to that particular VPLS and U-PE). In the above example, when E signals to A, it would set the TAII to be 1, 2, or 3, respectively, for the 3 PWs it must set up to A. It would similarly signal 3 PWs to B.

The LSP signaled from U-PE to N-PE is associated with an LSP from N-PE to U-PE in the usual manner. A PW between a U-PE and an N-PE is known as a "U-PW".

The signaling of the appropriate set of PWs from N-PE to N-PE is based on the remote list. The PWs between the N-PEs can all be considered equivalent. As long as the correct total number of PWs are established, the N-PEs can splice these PWs to appropriate U-PWs. The signaling of the correct number of PWs from N-PE to N-PE is based on the remote list. The remote list specifies the number of PWs to set up, per local U-PE, to a particular remote N-PE.

When signaling a particular PW from an N-PE to an N-PE, the AGI is set to the appropriate VPN-id. The TAII identifies the remote N-PE, as in the non-distributed case, i.e., it contains an IP address of the remote N-PE. If there are n such PWs, they are distinguished by the setting of the SAII. In order to allow multiple different SAII values in a single VPLS, the sending N-PE needs to have as many VSI-IDs as it has U-PEs. As noted above in Section 3.2.2, this may be achieved by using an IP address of each attached U-PE, for example. A PW between two N-PEs is known as an "N-PW".

Each U-PW must be "spliced" to an N-PW. This is based on the remote list. If the remote list contains an element <i, F>, then i U-PWs from each local U-PE must be spliced to i N-PWs from the remote N-PE F. It does not matter which U-PWs are spliced to which N-PWs, as long as this constraint is met.

If an N-PE has more than one local U-PE for a given VPLS, it must also ensure that a U-PW from each such U-PE is spliced to a U-PW from each of the other U-PEs.

3.5.2. Provisioning and Discovery

Every N-PE must be provisioned with the set of VPLS instances it supports, a VPN-id for each one, and a list of local U-PEs for each such VPLS. As part of the discovery procedure, the N-PE advertises the number of U-PEs for each VPLS. See Section 3.2.2 for details.

Auto-discovery (e.g., BGP-based) can be used to discover all the other N-PEs in the VPLS, and for each, the number of U-PEs local to that N-PE. From this, one can compute the total number of U-PEs in the VPLS. This information is sufficient to enable one to compute the local list and the remote list for each N-PE.

3.5.3. Non-Distributed VPLS as a Sub-Case

A PE that is providing "non-distributed VPLS" (i.e., a PE that performs both the U-PE and N-PE functions) can interoperate with N-PE/U-PE pairs that are providing distributed VPLS. The "non-distributed PE" simply advertises, in the discovery procedure, that it has one local U-PE per VPLS. And of course, the non-distributed PE does no PW switching.

If every PE in a VPLS is providing non-distributed VPLS, and thus every PE is advertising itself as an N-PE with one local U-PE, the resultant signaling is exactly the same as that specified in Section 3.2.3 above.

3.5.4. Splicing and the Data Plane

Splicing two PWs together is quite straightforward in the MPLS data plane, as moving a packet from one PW directly to another is just a 'label replace' operation on the PW label. When a PW consists of two or more PWs spliced together, it is assumed that the data will go to the node where the splicing is being done, i.e., that the data path will pass through the nodes that participate in PW signaling.

Further details on splicing are discussed in [RFC6073].

4. Inter-AS Operation

The provisioning, auto-discovery, and signaling mechanisms described above can all be applied in an inter-AS environment. As in [RFC4364], there are a number of options for inter-AS operation.

4.1. Multihop EBGP Redistribution of L2VPN NLRIs

This option is most like option (c) in [RFC4364]. That is, we use multihop External BGP (EBGP) redistribution of L2VPN NLRIs between source and destination ASes, with EBGP redistribution of labeled IPv4 or IPv6 routes from AS to neighboring AS.

An Autonomous System Border Router (ASBR) must maintain labeled IPv4 /32 (or IPv6 /128) routes to the PE routers within its AS. It uses EBGP to distribute these routes to other ASes, and sets itself as the BGP next hop for these routes. ASBRs in any transit ASes will also have to use EBGP to pass along the labeled /32 (or /128) routes. This results in the creation of a set of label switched paths from all ingress PE routers to all egress PE routers. Now, PE routers in different ASes can establish multi-hop EBGP connections to each other and can exchange L2VPN NLRIs over those connections. Following such exchanges, a pair of PEs in different ASes could establish an LDP session to signal PWs between each other.

For VPLS, the BGP advertisement and PW signaling are exactly as described in Section 3.2. As a result of the multihop EBGP session that exists between source and destination AS, the PEs in one AS that have VSIs of a certain VPLS will discover the PEs in another AS that have VSIs of the same VPLS. These PEs will then be able to establish the appropriate PW signaling protocol session and establish the full mesh of VSI-VSI pseudowires to build the VPLS as described in Section 3.2.3.

For VPWS, the BGP advertisement and PW signaling are exactly as described in Section 3.3. As a result of the multihop EBGP session that exists between source and destination AS, the PEs in one AS that

have pools of a certain color (VPN) will discover PEs in another AS that have pools of the same color. These PEs will then be able to establish the appropriate PW signaling protocol session and establish the full mesh of pseudowires as described in Section 3.2.3. A partial mesh can similarly be established using the procedures of Section 3.4.

As in Layer 3 VPNs, building an L2VPN that spans the networks of more than one provider requires some co-ordination in the use of RTs and RDs. This subject is discussed in more detail in Section 4.4.

4.2. EBGP Redistribution of L2VPN NLRIs with Multi-Segment Pseudowires

A possible drawback of the approach of the previous section is that it creates PW signaling sessions among all the PEs of a given L2VPN (VPLS or VPWS). This means a potentially large number of LDP or L2TPv3 sessions will cross the AS boundary and that these sessions connect to many devices within an AS. In the case where the ASes belong to different providers, one might imagine that providers would like to have fewer signaling sessions crossing the AS boundary and that the entities that terminate the sessions could be restricted to a smaller set of devices. Furthermore, by forcing the LDP or L2TPv3 signaling sessions to terminate on a small set of ASBRs, a provider could use standard authentication procedures on a small set of inter-provider sessions. These concerns motivate the approach described here.

[RFC6073] describes an approach to "switching" packets from one pseudowire to another at a particular node. This approach allows an end-to-end, multi-segment pseudowire to be constructed out of several pseudowire segments, without maintaining an end-to-end control connection. We can use this approach to produce an inter-AS solution that more closely resembles option (b) in [RFC4364].

In this model, we use EBGP redistribution of L2VPN NLRI from AS to neighboring AS. First, the PE routers use Internal BGP (IBGP) to redistribute L2VPN NLRI either to an ASBR, or to a route reflector of which an ASBR is a client. The ASBR then uses EBGP to redistribute those L2VPN NLRI to an ASBR in another AS, which in turn distributes them to the PE routers in that AS, or perhaps to another ASBR which in turn distributes them, and so on.

In this case, a PE can learn the address of an ASBR through which it could reach another PE to which it wishes to establish a PW. That is, a local PE will receive a BGP advertisement containing L2VPN NLRI corresponding to an L2VPN instance in which the local PE has some attached members. The BGP next-hop for that L2VPN NLRI will be an ASBR of the local AS. Then, rather than building a control

connection all the way to the remote PE, it builds one only to the ASBR. A pseudowire segment can now be established from the PE to the ASBR. The ASBR in turn can establish a PW to the ASBR of the next AS, and splice that PW to the PW from the PE as described in Section 3.5.4 and [RFC6073]. Repeating the process at each ASBR leads to a sequence of PW segments that, when spliced together, connect the two PEs.

Note that in the approach just described, the local PE may never learn the IP address of the remote PE. It learns the L2VPN NLRI advertised by the remote PE, which need not contain the remote PE address, and it learns the IP address of the ASBR that is the BGP next hop for that NLRI.

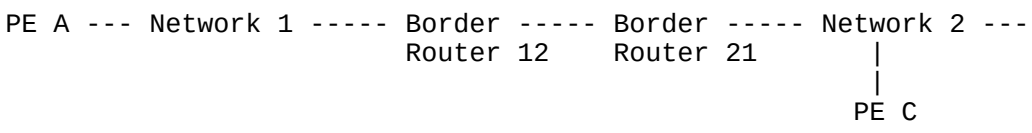
When this approach is used for VPLS, or for full-mesh VPWS, it leads to a full mesh of pseudowires among the PEs, just as in the previous section, but it does not require a full mesh of control connections (LDP or L2TPv3 sessions). Instead, the control connections within a single AS run among all the PEs of that AS and the ASBRs of the AS. A single control connection between the ASBRs of adjacent ASes can be used to support however many AS-to-AS pseudowire segments are needed.

Note that the procedures described here will result in the splicing points (PW Switching PEs (S-PEs) in the terminology of [RFC5659]) being co-located with the ASBRs. It is of course possible to have multiple ASBR-ASBR connections between a given pair of ASes. In this case, a given PE could choose among the available ASBRs based on a range of criteria, such as IGP metric, local configuration, etc., analogous to choosing an exit point in normal IP routing. The use of multiple ASBRs would lead to greater resiliency (at the timescale of BGP routing convergence) since a PE could select a new ASBR in the event of the failure of the one currently in use.

As in layer 3 VPNs, building an L2VPN that spans the networks of more than one provider requires some co-ordination in the use of RTs and RDs. This subject is discussed in more detail in Section 4.4.

4.3. Inter-Provider Application of Distributed VPLS Signaling

An alternative approach to inter-provider VPLS can be derived from the Distributed VPLS approach described above. Consider the following topology:



where A, B, and C are PEs in a common VPLS, but Networks 1 and 2 are networks of different service providers. Border Router 12 is Network 1's border router to network 2, and Border Router 21 is Network 2's border router to Network 1. We suppose further that the PEs are not "distributed", i.e, that each provides both the U-PE and N-PE functions.

In this topology, one needs two inter-provider pseudowires: A-B and A-C.

Suppose a service provider decides, for whatever reason, that it does not want each of its PEs to have a control connection to any PEs in the other network. Rather, it wants the inter-provider control connections to run only between the two border routers.

This can be achieved using the techniques of Section 3.5, where the PEs behave like U-PEs, and the BRs behave like N-PEs. In the example topology, PE A would behave like a U-PE that is locally attached to BR12; PEs B and C would be have like U-PEs that are locally attached to BR21; and the two BRs would behave like N-PEs.

As a result, the PW from A to B would consist of three segments: A-BR12, BR12-BR21, and BR21-B. The border routers would have to splice the corresponding segments together.

This requires the PEs within a VPLS to be numbered from 1-n (relative to that VPLS) within a given network.

4.4. RT and RD Assignment Considerations

We note that, in order for any of the inter-AS procedures described above to work correctly, the two ASes must use RTs and RDs consistently, just as in Layer 3 VPNs [RFC4364]. The structure of RTs and RDs is such that there is not a great risk of accidental collisions. The main challenge is that it is necessary for the operator of one AS to know what RT or RTs have been chosen in another AS for any VPN that has sites in both ASes. As in Layer 3 VPNs, there are many ways to make this work, but all require some co-operation among the providers. For example, provider A may tag all the NLRI for a given VPN with a single RT, say RT_A, and provider B can then configure the PEs that connect to sites of that VPN to import NLRI that contains that RT. Provider B can choose a different RT, RT_B, tag all NLRI for this VPN with that RT, and then provider A can import NLRI with that RT at the appropriate PEs. However, this does require both providers to communicate their choice of RTs for each VPN. Alternatively, both providers could agree to use a common RT for a given VPN. In any case, communication of RTs between the

providers is essential. As in Layer 3 VPNs, providers may configure RT filtering to ensure that only coordinated RT values are allowed across the AS boundary.

Note that a single VPN identifier (carried in a BGP Extended Community) is required for each VPLS or VPWS instance. The encoding rules for these identifiers [RFC4360] ensure that collisions do not occur with other providers. However, for a single VPLS or VPWS instance that spans the networks of two or more providers, one provider will need to allocate the identifier and communicate this choice to the other provider(s), who must use the same value for sites in the same VPLS or VPWS instance.

5. Security Considerations

This document describes a number of different L2VPN provisioning models, and specifies the endpoint identifiers that are required to support each of the provisioning models. It also specifies how those endpoint identifiers are mapped into fields of auto-discovery protocols and signaling protocols.

The security considerations related to the signaling protocols are discussed in the relevant protocol specifications ([RFC5036], [RFC4447], [RFC3931], and [RFC4667]).

The security considerations related to BGP-based auto-discovery, including inter-AS issues, are discussed in [RFC4364]. L2VPNs that use BGP-based auto-discovery may automate setup of security mechanisms as well. Specification of automated security mechanisms are outside the scope of this document, but are recommended as a future work item.

The security considerations related to the particular kind of L2VPN service being supported are discussed in [RFC4664], [RFC4665], and [RFC4762].

The way in which endpoint identifiers are mapped into protocol fields does not create any additional security issues.

6. IANA Considerations

IANA has assigned an AFI and a SAFI for L2VPN NLRI. Both the AFI and SAFI are the same as the values assigned for [RFC4761]. That is, the AFI is 25 (L2VPN) and the SAFI is 65 (already allocated for VPLS). The same AFI and SAFI are used for both VPLS and VPWS auto-discovery as described in this document.

[RFC4446] defines registries for "Attachment Group Identifier (AGI) Type" and "Attachment Individual Identifier (AII) Type". Type 1 in each registry has been assigned to the AGI and AII formats defined in this document.

IANA has assigned two new LDP status codes. IANA already maintains a registry of name "STATUS CODE NAME SPACE" defined by [RFC5036]. The following values have been assigned:

0x00000030 Attachment Circuit bound to different PE

0x0000002D Attachment Circuit bound to different remote Attachment Circuit

Two new L2TP Result Codes have been registered for the CDN message. IANA already maintains a registry of L2TP Result Code Values for the CDN message, defined by [RFC3438]. The following values have been assigned:

27: Attachment Circuit bound to different PE

28: Attachment Circuit bound to different remote Attachment Circuit

[RFC4360] defines a registry entitled "Two-octet AS Specific Extended Community". IANA has assigned a value in this registry from the "transitive" range (0x0000-0x00FF). The value is as follows:

- 0x000A Two-octet AS specific Layer 2 VPN Identifier

[RFC4360] defines a registry entitled "IPv4 Address Specific Extended Community". IANA has assigned a value in this registry from the "transitive" range (0x0100-0x01FF). The value is as follows:

- 0x010A Layer 2 VPN Identifier

7. BGP-AD and VPLS-BGP Interoperability

Both BGP-AD and VPLS-BGP [RFC4761] use the same AFI/SAFI. In order for both BGP-AD and VPLS-BGP to co-exist, the NLRI length must be used as a demultiplexer.

The BGP-AD NLRI has an NLRI length of 12 bytes, containing only an 8-byte RD and a 4-byte VSI-ID. VPLS-BGP [RFC4761] uses a 17-byte NLRI length. Therefore, implementations of BGP-AD must ignore NLRI that are greater than 12 bytes.

8. Acknowledgments

Thanks to Dan Tappan, Ted Qian, Ali Sajassi, Skip Booth, Luca Martini, Dave McDysan, Francois Le Faucheur, Russ Gardo, Keyur Patel, Sam Henderson, and Matthew Bocci for their comments, criticisms, and helpful suggestions.

Thanks to Tissa Senevirathne, Hamid Ould-Brahim, and Yakov Rekhter for discussing the auto-discovery issues.

Thanks to Vach Kompella for a continuing discussion of the proper semantics of the generalized identifiers.

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3438] Townsley, W., "Layer Two Tunneling Protocol (L2TP) Internet Assigned Numbers Authority (IANA) Considerations Update", BCP 68, RFC 3438, December 2002.

[RFC3931] Lau, J., Townsley, M., and I. Goyret, "Layer Two Tunneling Protocol - Version 3 (L2TPv3)", RFC 3931, March 2005.

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, February 2006.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[RFC4447] Martini, L., Rosen, E., El-Aawar, N., Smith, T., and G. Heron, "Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)", RFC 4447, April 2006.

[RFC4667] Luo, W., "Layer 2 Virtual Private Network (L2VPN) Extensions for Layer 2 Tunneling Protocol (L2TP)", RFC 4667, September 2006.

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, January 2007.

[RFC5036] Andersson, L., Minei, I., and B. Thomas, "LDP Specification", RFC 5036, October 2007.

[RFC6073] Martini, L., Metz, C., Nadeau, T., Bocci, M., and M. Aissaoui, "Segmented Pseudowire", RFC 6073, January 2011.

9.2. Informative References

[RFC3985] Bryant, S. and P. Pate, "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture", RFC 3985, March 2005.

[RFC4026] Andersson, L. and T. Madsen, "Provider Provisioned Virtual Private Network (VPN) Terminology", RFC 4026, March 2005.

[RFC4446] Martini, L., "IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)", BCP 116, RFC 4446, April 2006.

[RFC4664] Andersson, L. and E. Rosen, "Framework for Layer 2 Virtual Private Networks (L2VPNs)", RFC 4664, September 2006.

[RFC4665] Augustyn, W. and Y. Serbest, "Service Requirements for Layer 2 Provider-Provisioned Virtual Private Networks", RFC 4665, September 2006.

[RFC4761] Kompella, K. and Y. Rekhter, "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling", RFC 4761, January 2007.

[RFC4762] Lasserre, M. and V. Kompella, "Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling", RFC 4762, January 2007.

[RFC5003] Metz, C., Martini, L., Balus, F., and J. Sugimoto, "Attachment Individual Identifier (AII) Types for Aggregation", RFC 5003, September 2007.

[RFC5659] Bocci, M. and S. Bryant, "An Architecture for Multi-Segment Pseudowire Emulation Edge-to-Edge", RFC 5659, October 2009.

Authors' Addresses

Eric Rosen
Cisco Systems, Inc.
1414 Mass. Ave.
Boxborough, MA 01719
USA

EMail: erosen@cisco.com

Bruce Davie
Cisco Systems, Inc.
1414 Mass. Ave.
Boxborough, MA 01719
USA

EMail: bsd@cisco.com

Vasile Radoaca
Alcatel-Lucent
Think Park Tower 6F
2-1-1 Osaki, Tokyo, 141-6006
Japan

EMail: vasile.radoaca@alcatel-lucent.com

Wei Luo

EMail: luo@weiluo.net